remote sensing

Special Issue Reprint

Recent Advances in Hyperspectral Remote Sensing

Theories, Technologies and Applications

Edited by
Jiaxin Li and Lianru Gao

mdpi.com/journal/remotesensing

MDPI

Recent Advances in Hyperspectral Remote Sensing: Theories, Technologies and Applications

Recent Advances in Hyperspectral Remote Sensing: Theories, Technologies and Applications

Guest Editors

Jiaxin Li

Lianru Gao

Assisted by

Ke Zheng

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editors

Jiaxin Li
School of Computer Science and Technology
Chongqing University of Posts and Telecommunications
Chongqing
China

Lianru Gao
Aerospace Information Research Institute
Chinese Academy of Sciences
Beijing
China

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Remote Sensing* (ISSN 2072-4292), freely accessible at: https://www.mdpi.com/journal/remotesensing/special_issues/W49E1152UA.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-9403-1 (Hbk)
ISBN 978-3-7258-9404-8 (PDF)
https://doi.org/10.3390/books978-3-7258-9404-8

© 2026 by the authors. Articles in this reprint are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The reprint as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

About the Editors . . . . . **vii**

Tugcan Dundar
Superpixel-Level Joint-Sparse and Graph-Regularized Framework for Hyperspectral Image Classification
Reprinted from: *Remote Sens.* **2026**, *18*, 2699, https://doi.org/10.3390/rs18162699 . . . . . **1**

Jianfei Cao, Yihui Liu, Xinrong Duan, Xiaoli Liu, Xiukun Zhang, Qixin Shi and Xibo Xu
Enhancing the Estimation and Mapping of Soil Cadmium by Using Geospatial Information-Guided Machine Learning and Principal Component Spectra
Reprinted from: *Remote Sens.* **2026**, *18*, 2341, https://doi.org/10.3390/rs18142341 . . . . . **25**

Jie Shen, Yimeng Ma and Houqun Yang
A Hierarchical Semantic Consistency Constraint Framework for Hyperspectral and LiDAR Data Joint Classification
Reprinted from: *Remote Sens.* **2026**, *18*, 2058, https://doi.org/10.3390/rs18122058 . . . . . **45**

Jochem Verrelst, Bhagyashree Verma and Pablo Reyes-Muñoz
Scale-Aware Interpretation of Vegetation Traits and SIF-Based Dynamics in Earth Observation
Reprinted from: *Remote Sens.* **2026**, *18*, 1951, https://doi.org/10.3390/rs18121951 . . . . . **71**

Chang Zhong, Yu Zhao, Mengchun Pan, Qi Zhang, Xinxin Sui, Li Chen, et al.
Bathymetric Inversion of Tibetan Plateau Lakes Using Hyperspectral Imagery and ICESat-2 Data
Reprinted from: *Remote Sens.* **2026**, *18*, 1886, https://doi.org/10.3390/rs18121886 . . . . . **110**

Xibo Xu, Ying Wang, Xinrui Dai, Qi Shen, Quanyuan Wu, Zeqiang Wang and Jianfei Cao
Improving Satellite-Based Estimation and Mapping of Soil Lead by Using an Enhanced Spectral Feature Set and XGBoost Model
Reprinted from: *Remote Sens.* **2026**, *18*, 1446, https://doi.org/10.3390/rs18091446 . . . . . **139**

Cai Liu, Junting Qiu, Junchuan Yu, Yanbo Zhao, Yuanquan Xu, Xin Zhang, et al.
Applications of Airborne Hyperspectral Imagery in Rare Earth Element Exploration: A Case Study of the World-Class Bayan Obo Deposit, China
Reprinted from: *Remote Sens.* **2026**, *18*, 1110, https://doi.org/10.3390/rs18081110 . . . . . **163**

Nannan Hu, Zhongao Wang, Minghao Wang and Yuefeng Zhao
Double-Attention Context Interactive Network for Hyperspectral Image Classification
Reprinted from: *Remote Sens.* **2026**, *18*, 1059, https://doi.org/10.3390/rs18071059 . . . . . **176**

Andrea Taramelli, Sara Liburdi, Alessandra Nguyen Xuan, Simone Mancon, Serena Sapio and Emiliana Valentini
Retrieval of Multiple Variables from Hyperspectral Data: A PRISMA-Aligned Systematic Review of Classical Physics-Based Machine Learning and Hybrid Algorithms in Vegetation and Raw Materials Application Domains
Reprinted from: *Remote Sens.* **2026**, *18*, 798, https://doi.org/10.3390/rs18050798 . . . . . **201**

Yuannan Gui, Lu Xu, Dongping Ming, Yanfei Wei and Ming Huang
Hyperspectral Image Classification Using SIFANet: A Dual-Branch Structure Combining CNN and Transformer
Reprinted from: *Remote Sens.* **2026**, *18*, 398, https://doi.org/10.3390/rs18030398 . . . . . **239**

Kewen Qu, Xiaojuan Luo and Wenxing Bao
DNMF-AG: A Sparse Deep NMF Model with Adversarial Graph Regularization for Hyperspectral Unmixing
Reprinted from: *Remote Sens.* **2026**, *18*, 155, https://doi.org/10.3390/rs18010155 . . . . . **267**

Xiao-Yu Yu, Jian-Li Zhao, Sheng Fang, Tian-Heng Zhang, Liang Li and Xin-Yue Huang
Weighted Total Variation for Hyperspectral Image Denoising Based on Hyper-Laplacian Scale Mixture Distribution
Reprinted from: *Remote Sens.* **2026**, *18*, 135, https://doi.org/10.3390/rs18010135 . . . . . **296**

Jing Yang, Jianbin Zhao, Lu Chen, Haorui Ning and Ying Li
SATRNet: Self-Attention-Aided Deep Unfolding Tensor Representation Network for Robust Hyperspectral Anomaly Detection
Reprinted from: *Remote Sens.* **2025**, *17*, 3137, https://doi.org/10.3390/rs17183137 . . . . . **323**

Genrui Zhang, Wenzheng Wang, Yuqi Han, Chenwei Deng and Xingshi Luo
Reliable Task-Constrained Band Selection Method for Hyperspectral Anomaly Detection
Reprinted from: *Remote Sens.* **2025**, *17*, 3081, https://doi.org/10.3390/rs17173081 . . . . . **355**

About the Editors

Jiaxin Li

Jiaxin Li is currently a lecturer and doctorate at the School of Computer Science and Technology, Chongqing University of Posts and Telecommunications. His research interests include hyperspectral remote sensing, multimodal remote sensing fusion, and deep learning.

He has authored over 40 peer-reviewed papers with more than 3200 citations in Google Scholar, including 7 ESI highly cited papers. He received awards from the Science and Technology Research Program of Chongqing Municipal Education Commission, the Postdoctoral Fellowship Program (Grade B) of China Postdoctoral Science Foundation, and the Chongqing Postdoctoral Science Foundation. He is the 2026 Top reviewer of *TGRS* and has served as Associate Editor of *IEEE Internet of Things Journal*; Guest Editor of *JAG*, *JSTARS*, and *Remote Sensing*; and reviewer for over 30 journals.

Lianru Gao

Lianru Gao is currently a Professor with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, CAS. In the last ten years, he was the PI of ten scientific research projects at national and ministerial levels, including projects by the National Natural Science Foundation of China (2018–2020, 2022, 2025, and 2024–2028) and by the National Key Research and Development Program of China (2021–2025). He has authored more than 240 peer-reviewed articles, and there are more than 150 journal articles included in Science Citation Index (SCI). He has co-authored three academic books including *Hyperspectral Image Information Extraction*. He holds 30 national invention patents in China. His research focuses on hyperspectral image processing and information extraction. Dr. Gao was supported by the National Science Foundation for Distinguished Young Scholars of China in 2023, and was awarded the Outstanding Science and Technology Achievement Prize of the CAS in 2016 and the Second Prize of The State Scientific and Technological Progress Award in 2018. He received the 2021 Outstanding Paper Award at the IEEE Workshop on Hyperspectral Image Processing: Evolution in Remote Sensing (WHISPERS). He is an Associate Editor of *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING* (*TGRS*) and *IET Image Processing*. He is a fellow of the Institution of Engineering and Technology.

remote sensing

MDPI

Article

Superpixel-Level Joint-Sparse and Graph-Regularized Framework for Hyperspectral Image Classification

Tugcan Dundar

Department of Electrical and Electronics Engineering, Faculty of Engineering, Mersin University, Mersin 33343, Turkey; tugcandundar@mersin.edu.tr

Highlights

What are the main findings?

- A similarity-based weighting matrix derived from the relationship between superpixel pixels and training samples is incorporated into the $l_{2,1}$ norm regularization, leading to a more discriminative joint-sparse representation.
- Graph Laplacian regularization enhances spatial smoothness and reduces classification noise.

What are the implications of the main findings?

- The proposed framework provides an efficient spectral–spatial classification scheme for hyperspectral images, especially when labeled samples are restricted.
- The results indicate that incorporating the intrinsic manifold structure of hyperspectral data, along with capturing local spectral affinities, plays a key role in achieving accurate and robust spectral–spatial classification.

Abstract

Hyperspectral image classification (HSIC) remains challenging because high-dimensional spectral signatures must be interpreted together with spatially coherent land-cover structures, particularly when labeled samples are limited. This paper presents a superpixel-based spectral–spatial HSIC method called SJSGR, which combines joint-sparse representation with graph Laplacian regularization. The HSI is first partitioned into homogeneous superpixel regions so that neighbouring pixels with similar spectral characteristics can be represented jointly rather than independently. For each superpixel, a similarity-aware weighting matrix is constructed between the training dictionary and the superpixel samples, encouraging the coefficient matrix to select more label-consistent and representative training atoms. To further preserve local manifold structure, graph Laplacian regularization is incorporated into the optimization objective, enforcing smooth and coherent representation coefficients among neighboring pixels within each superpixel. The resulting unified formulation integrates spectral correlation, spatial consistency, and local geometric structure, and is solved by the alternating-direction method of multipliers (ADMM). Classification is then performed by assigning each superpixel to the class with the minimum reconstruction error. Experiments are conducted on three real-world HSI datasets called Indian Pines, Pavia University and Fanglu to compare the proposed framework with several sparse representation and graph-based HSIC methods. Experimental results on these datasets reveal the capability of the proposed method, obtaining overall accuracies of 98.12%, 98.04%, and 98.26% under 10%, 1% and 1% labeled samples, respectively. Besides obtaining nearly 1% higher overall accuracy than the compared methods under these low-training-sample distributions, the SJSGR also provided better classification performance even under much more limited numbers of training samples. The findings suggest that

superpixel-guided sparse representation with local manifold regularization is a promising direction for effective spectral–spatial HSIC.

Keywords: superpixel; sparse; graph; hyperspectral; classification

1. Introduction

Hyperspectral imaging sensors are capable of collecting reflected sunlight from land-cover materials over a wide range of the electromagnetic spectrum, spanning from visible to near-infrared regions in narrow intervals [1–3]. Such high spectral resolution allows sensors to obtain spectral signatures of objects. These signatures are employed to classify or detect samples in HSIs [4–7]. Early proposed HSIC algorithms relied solely on spectral information such as sparse representation classifier (SRC) [8], which finds a coefficient vector with few non-zero elements representing a given test sample by a set of training samples. Later, it was observed that the use of spatial information further improves the HSIC performance [9]. In this regard, several spectral–spatial HSIC methods [8,10–12] have been proposed in the literature. Among these, joint-SRC (JSRC) [8] has attracted significant attention owing to its simplicity and ease of extension. Chen et al. [8] utilized a fixed-size patch around a given test sample to incorporate the spatial information. The idea behind JSRC is to represent neighboring pixels sparsely and simultaneously over a common training dictionary, thereby exploiting their spectral correlation to improve HSIC accuracy.

To enhance the results of JSRC, many methods have been proposed in the literature. For example, Zhang et al. [13] introduced nonlocal weighted JSRC (NLW-JSRC) to measure the similarity between test samples and neighboring pixels by considering the structural similarity. Since the contributions of all neighboring pixels are assumed to be equal by JSRC, its performance is reduced in heterogeneous regions. In order to alleviate this issue, NLW-JSRC assigns distinct weights to neighboring pixels, yielding more informative sparse coefficients. Tu et al. [14] proposed correlation coefficient JSRC (CCJSR), which incorporates correlation coefficients between the test and training samples as a regularization term during the residual calculation. By simultaneously utilizing the representation residual and spectral correlation, CCJSR improves the discrimination between different classes compared with JSRC. A locally adaptive neighborhood selection scheme and multi-level JSRC were proposed in [15] to mitigate the limitations of JSRC using a fixed-scale patch and uniformly assigning equal weights to neighboring pixels. Chen et al. [16] extracted local matrix features and then employed kernel JSRC (KJSRC) [17] to better characterize the nonlinear structure of a given HSI. It is known that JSRC finds the sparse coefficients directly from the original spectral–spatial features, and also assumes a linear relationship among the samples. However, ref. [16] computes local matrix features to represent spectral–spatial information better, and then exploits kernel mapping to characterize the nonlinear structure of HSI. In order to make effective use of spatial information beyond a fixed-size patch, a superpixel-based discriminative sparse model (SBDSM) [18] was proposed. In [18], a superpixel is defined as a shape-adaptive region consisting of a set of neighboring samples with similar spectral characteristics. Additionally, it was observed that the time requirement of superpixel-based HSIC is much lower than that of fixed-size patch-based classification. Subsequently, several superpixel-based HSIC methods utilizing sparse representation have been developed. Dundar et al. [19] proposed multiscale superpixels and guided filter (MSSGF), which sparsely represents samples inside a given set of superpixel regions. Since the distributions of different homogeneous regions in HSI are not uniform, MSSGF constructs multiple superpixel maps to capture the spatial information better than JSRC. In this

https://doi.org/10.3390/rs18162699

way, both small and large homogeneous regions are considered. Zhang et al. [20] proposed an improved version of NLW-JSRC [13], called superpixel-based NLW-JSRC (SNLW-JSRC), which assigns distinct weight to each specific neighborhood inside a superpixel using a nonlocal weighting strategy. Compared to the NLW-JSRC, the spatial structure of HSI is better preserved by the SNLW-JSRC, which constrains the nonlocal search within the superpixel structures. The spectral similarity of samples and the presence of few potential outliers in each superpixel were modeled by a locality-constrained joint-sparse and weighted low-rank (LCJS-WLR) [21] classifier. This method encodes pixels within each superpixel using joint-sparse and weighted low-rank representations, while incorporating sparse noise regularization and a locality constraint on the joint-sparse term. To address a limitation of JSRC, where the intrinsic low-dimensional structure is not utilized, the LCJS-WLR employs low-rank regularization to extract shared spectral structure within each superpixel. By imposing a noise-regularization term, the effects of outliers inside each superpixel are suppressed. Additionally, the use of locality-constraint enables the classifier to select more discriminative training samples.

Although sparse representation-based approaches provide promising performance, they often fail to maintain the intrinsic manifold structure of the HSIs. More specifically, the interactions between neighbor pixels are not properly preserved, which may deteriorate representation coherence in homogeneous regions such as superpixels. To handle this issue, graph-based models have been widely introduced to characterize the underlying intrinsic geometric structure of HSIs. By defining a graph where nodes correspond to pixels and edges encode similarity, graph Laplacian regularization [22] can effectively encourage local smoothness. Recently, several graph Laplacian-regularized HSIC methods have been proposed in the literature. Liu et al. [23] proposed an approach that enforces structural consistency across bands by integrating graph Laplacian and sparse regularizations. Unlike JSRC, ref. [23] extracts band-level local image patches for pixels, and then finds the regarding sparse coefficients and learned dictionaries by imposing sparse representation and graph regularization terms. It was said that the use of graph regularization enforces the neighboring pixels to share similar sparse coefficients, thereby allowing the preservation of local geometric structures. However, since the graphs are constructed based on the randomly sampled training pixels, the lack of labeled samples may deteriorate the performance. Additionally, the graph weights are assigned binary (0/1) in accordance with the k-nearest-neighbor strategy, discarding the degree of similarity between neighboring pixels. Xue et al. [24] introduced sparse graph regularization (SGR) that constructs the graph Laplacian matrix based on the sparse coefficient matrix evaluated over the entire HSI, thereby ignoring the local neighborhood information. In this study, the graph construction process does not depend on the neighboring pixels. Instead, two pixels are assumed to be neighbors if they are represented by similar dictionary atoms. However, global graph construction may be negatively affected by outlier pixels, which may lead to incorrect neighborhood connections. As an enhancement over SGR, weighted SGR (WSGR) [25], was proposed to incorporate local neighborhood information within fixed-size patches through a Gaussian weighting scheme. In order to address a limitation of SGR, ref. [25] encodes a pixel as the weighted averages of its surrounding neighbors. In this way, it aimed to integrate the local neighborhood information. However, it was reported that the accuracy results are quite sensitive to the weighting method and patch size. Sellars et al. [26] presented the superpixel-based graph learning (SGL) method to effectively utilize spatial information in HSIs. They extracted several spectral–spatial features from the superpixels, and then these features were used to obtain the adaptive graph node weights between two adjacent superpixels. It was noticed that superpixel-guided graph construction better characterizes the geometric structure of HSI. In this study, instead of constructing the graphs

https://doi.org/10.3390/rs18162699

within superpixels, they are constructed between neighboring superpixels. Therefore, local information is captured using only a graph-based term. Additionally, the label of each superpixel is determined by a label propagation algorithm, which spreads information to neighboring superpixels. Therefore, superpixel segmentation faults or erroneous graph construction may directly affect the classification performance. Zhao et al. [27] introduced the sparse superpixel graph (SSG), where superpixels are employed as graph nodes instead of pixels, greatly reducing computational complexity while preserving both spatial structure and spectral information for more efficient HSIC. Different from [26], the graph construction stage is based on both global spectral and local spatial connections, resulting more reliable graph representations. In addition, instead of the adaptive graph weights in [26], this study [27] considers binary weights, which may lead to the loss of the degree or strength of similarity. Since a label-propagation strategy is employed to determine the class label of a superpixel, binary weighting can cause a weakly similar superpixel neighbor to enforce the same influence as a strongly similar one during label propagation. Yang et al. [28] introduced an approach that leverages data compactness, spectral variability and local manifold structure through a local low-rank framework and superpixel-based locality graph. Unlike [26,27], the graph connections are constructed between the pixels within each superpixel. Following the computation of the low-rank component, a pixel-wise model was employed for this task, leading to a loss of spatial context during final label assignment.

Recently, foundation models (FMs) for HSI analysis have been introduced in the literature [29–32]. Unlike conventional task-driven methods, FMs are deep learning models pre-trained on large-scale HSI datasets to learn generalized spectral–spatial representations that can be transferred to different HSI applications such as classification, target detection, semantic segmentation, and unmixing [33,34]. One of the earliest models proposed for hyperspectral foundations is HyperSIGMA [29], where the HSIs are pre-trained using a massive set of hyperspectral data to learn transferable spectral–spatial representations for a variety of HSI processing tasks. HyperSL [30] designed a spectral foundation model, which integrates wavelength-oriented positional encoding, to standardize spectral vectors to improve their generalization across different sensors. Another model proposed is SpectralEarth [31], which built a massive EnMAP-based pre-training dataset and showed that the use of self-supervised pre-training could increase the generalization capacity of hyperspectral FMs. HySens [32] proposed a sensor-agnostic FM that achieves knowledge transfer across different hyperspectral datasets by unifying different spectral bands within a common framework. While recent developments have shown that hyperspectral FMs can generalize well, they do need pre-training on large data sets as well as heavy computation. Hence, model-driven methods remain a preferred choice especially when there are limitations in training samples and computational complexity.

Motivated by the limitations of joint-sparsity-based methods and the merits of graph-based approaches in modeling local structures, we propose the SJSGR method for HSIC. In the first step, the HSI is divided into superpixels to construct local regions, and then the similarity between the training dictionary and HSI data is computed. This weighting matrix is calculated to better characterize the joint-sparse structure of each superpixel in the next stage. Secondly, the sparse structure and local manifold of the coefficient matrix in each superpixel are modeled using weighted joint-sparse and weighted graph Laplacian regularizations, respectively. Then, the designed optimization problem is solved by the alternating-direction method of multipliers (ADMM) [35], and the regularized coefficient matrix is obtained. Finally, the label of each superpixel is determined by computing the reconstruction error between the pixels within the superpixel and the training dictionary using the corresponding regularized coefficients.

https://doi.org/10.3390/rs18162699

The rest of the paper is structured as follows. Section 2 explains the proposed SJSGR in detail. The experimental setup and findings are reported in Section 3, and then the obtained results are discussed in Section 4. Finally, Section 5 summarizes the paper and suggests future research directions.

2. Proposed SJSGR Method

2.1. Notations

A hyperspectral datacube can be represented as $\mathcal{Y} \in \mathbb{R}^{r \times c \times b}$, consisting of b channels with a spatial dimension of $r \times c$. Let $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m] \in \mathbb{R}^{b \times m}$ be the matrix form of $\mathcal{Y}$, where $m = r \times c$ denotes the number of pixel vectors. Let $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_k] \in \mathbb{R}^{b \times k}$ be a training dictionary including k samples formed by stacking class-specific sub-dictionaries, i.e., $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_c]$, where c denotes the number of classes. Each sub-dictionary containing n_i samples can be given as $\mathbf{D}_i = [\mathbf{d}_{i_1}, \mathbf{d}_{i_2}, \cdots, \mathbf{d}_{i_{n_i}}]$, where $k = \sum_{i=1}^{c} n_i$. It is important to indicate that the columns of both $\mathbf{Y}$ and $\mathbf{D}$ are normalized to have unit-norm.

2.2. Formulation

The entire HSI can be represented as a linear combination of the dictionary, i.e., $\mathbf{Y} = \mathbf{DS} + \mathbf{E}$, where $\mathbf{S} \in \mathbb{R}^{k \times m}$ and $\mathbf{E} \in \mathbb{R}^{b \times m}$ denote the coefficient matrix and reconstruction error, respectively. It is known that the neighbor pixels of a HSI share similar spectral characteristics, which promotes joint-sparse structure of the coefficient matrix in local regions. In order to obtain these areas, we utilize the entropy-rate superpixel (ERS) algorithm [36] which provides homogeneous spatial regions called superpixels. Considering that the segmentation capacity of ERS is highly reliant on the pre-determined number of superpixels and that no standard formulation exists for this parameter [18,37–41], a simple formulation is introduced. The optimal number of superpixels is determined as $floor(m/(100 * \gamma^2))$ where $floor(\cdot)$ rounds a number down to the nearest integer less than or equal to that number, and γ is an adjustable parameter. This formulation allows the construction of suitable superpixel clusters across datasets with different spatial image sizes by tuning the γ parameter. After obtaining superpixels, the following optimization problem can be written to model the joint-sparse blocks as

$$\underset{\mathbf{S}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{Y} - \mathbf{DS}\|_F^2 + \lambda \sum_{g=1}^{l} \|\mathbf{S}_g\|_{2,1} \tag{1}$$

where $\|\cdot\|_F$ and $\|\cdot\|_{2,1}$ indicate the Frobenius norm and $l_{2,1}$ norm, respectively. Here, $\mathbf{S}_g \in \mathbb{R}^{k \times n_g}$ denotes the joint-sparse coefficient matrix of the gth superpixel in a total of l superpixels, and $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_g, \cdots, \mathbf{S}_l]$. The notation n_g represents the number of pixels in each superpixel. In order to improve the performance, we utilize a weighting matrix $\mathbf{W} \in \mathbb{R}^{k \times m}$ that holds the similarity of $\mathbf{Y}$ across $\mathbf{D}$, as

$$\begin{aligned} &\mathbf{W}_{t,j} = \sqrt{\|\mathbf{y}_j - \mathbf{d}_t\|_2^2 + \|p(\mathbf{y}_j) - p(\mathbf{d}_t)\|_2^2} \\ &\text{subject to } j = 1, \cdots, m \text{ and } t = 1, \cdots, k \end{aligned} \tag{2}$$

where $p(\cdot) \in \mathbb{R}^2$ represents the normalized pixel coordinates. It should be noted that the row and column indices are separately normalized by dividing each by its corresponding maximum index values. Each column of $\mathbf{W}$, encoding the similarity between a data sample $\mathbf{y}_j$ and the training dictionary, is then rescaled to $[0, 1]$. After incorporating $\mathbf{W}$, the problem becomes

$$\underset{\mathbf{S}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{Y} - \mathbf{DS}\|_F^2 + \lambda \sum_{g=1}^{l} \|\mathbf{W}_g \odot \mathbf{S}_g\|_{2,1} \tag{3}$$

https://doi.org/10.3390/rs18162699

Here, $\mathbf{W}_g \in \mathbb{R}^{k \times n_g}$ indicates the weight matrix of of the gth superpixel, and $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_g, \cdots, \mathbf{W}_l]$. The operator $\odot$ denotes element-wise multiplication. The idea behind this weighting approach is to penalize inappropriate training samples more. In order to exploit the spatial similarity between the pixels within each superpixel, the following graph Laplacian regularization term is inserted to (3) as

$$\underset{\mathbf{S}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{S}\|_F^2 + \sum_{g=1}^{l} \left(\lambda \|\mathbf{W}_g \odot \mathbf{S}_g\|_{2,1} + \tau \operatorname{Tr}\left(\mathbf{S}_g \mathbf{L}_g \mathbf{S}_g^{\mathrm{T}}\right)\right) \tag{4}$$

The expansion of graph regularization term is given by

$$\sum_{g=1}^{l} \operatorname{Tr}(\mathbf{S}_g \mathbf{L}_g \mathbf{S}_g^{\mathrm{T}}) = \frac{1}{2} \sum_{g=1}^{l} \sum_{(i,j) \in \mathcal{N}_g} \mathbf{Q}_g(i,j) \|\mathbf{s}_{g_i} - \mathbf{s}_{g_j}\|_2^2 \tag{5}$$

where $\mathbf{L}_g \in \mathbb{R}^{n_g \times n_g}$ denotes the graph Laplacian matrix of gth superpixel, $\mathcal{N}_g$ represents the set of neighboring pixels within each superpixel, and $\mathbf{s}_{g_i}$ and $\mathbf{s}_{g_j}$ are the coefficient vectors of these adjacent pixels. The integration of this regularization term enforces the joint-sparse coefficient matrix to have similar coefficients among neighboring pixels. Furthermore, $\mathbf{L}_g$ is defined as $\mathbf{L}_g = \mathbf{R}_g - \mathbf{Q}_g$, where $\mathbf{Q}_g \in \mathbb{R}^{n_g \times n_g}$ is the affinity matrix of gth superpixel, computed as

$$\underset{(i,j) \in \mathcal{N}_g}{\mathbf{Q}_g(i,j)} = \exp\left(-\frac{\|\mathbf{y}_{g_i} - \mathbf{y}_{g_j}\|_2^2}{2\sigma^2}\right) ; \; \substack{i=1,\cdots,n_g \\ j=1,\cdots,n_g} \tag{6}$$

where $\mathbf{y}_{g_i}$ and $\mathbf{y}_{g_j}$ are the neighboring pixels within that superpixel, and σ controls the scale of the Gaussian kernel, which is set as 0.5. The last term $\mathbf{R}_g \in \mathbb{R}^{n_g \times n_g}$ is a diagonal matrix evaluated by $\mathbf{R}_g(i,i) = \sum_{j=1}^{n_g} \mathbf{Q}_g(i,j)$.

In order to solve the proposed optimization problem in (4), we use ADMM after separating the problem into subproblems. The augmented Lagrangian form of (4) is given by

$$\begin{aligned}
&\mathcal{L}(\mathbf{S}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3) = \frac{1}{2}\|\mathbf{Y} - \mathbf{V}_1\|_F^2 \\
&+ \sum_{g=1}^{l} \left(\lambda \|\mathbf{W}_g \odot \mathbf{V}_{2_g}\|_{2,1} + \tau \operatorname{Tr}\left(\mathbf{V}_{3_g} \mathbf{L}_g \mathbf{V}_{3_g}^{\mathrm{T}}\right)\right) \\
&+ \frac{\mu}{2}\Big(\|\mathbf{D}\mathbf{S} - \mathbf{V}_1 - \mathbf{P}_1\|_F^2 + \|\mathbf{S} - \mathbf{V}_2 - \mathbf{P}_2\|_F^2 \\
&+ \|\mathbf{S} - \mathbf{V}_3 - \mathbf{P}_3\|_F^2\Big)
\end{aligned} \tag{7}$$

where $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$ are the variables to be updated alternately. The terms $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ are the Lagrange multipliers, and μ is a penalty parameter. The optimization stages of these variables are given by Algorithm 1 in detail. Suppose that the optimal coefficients of $\mathbf{S}$ are obtained after the convergence criterion is satisfied. Then, a class label is assigned to g^{th} superpixel based on the following evaluation

$$\underset{i=1,\cdots,c}{\mathbf{r}_g(i)} = \|\mathbf{Y}_g - \mathbf{D}_i \mathbf{S}_{g_i}\|_F \,/\, \|\mathbf{S}_{g_i}\|_F \tag{8}$$

where $\mathbf{r}_g \in \mathbb{R}^c$ is a residual vector of gth superpixel containing the class-based reconstruction errors, $\mathbf{Y}_g \in \mathbb{R}^{b \times n_g}$ denotes the pixels within that superpixel, and $\mathbf{S}_{g_i}$ refers to the coefficient matrix of superpixel belonging to the ith class, respectively. Finally, the label of this superpixel is determined as the class that yields the minimum reconstruction error in $\mathbf{r}_g$. The pipeline of the proposed method is illustrated in Figure 1.

https://doi.org/10.3390/rs18162699

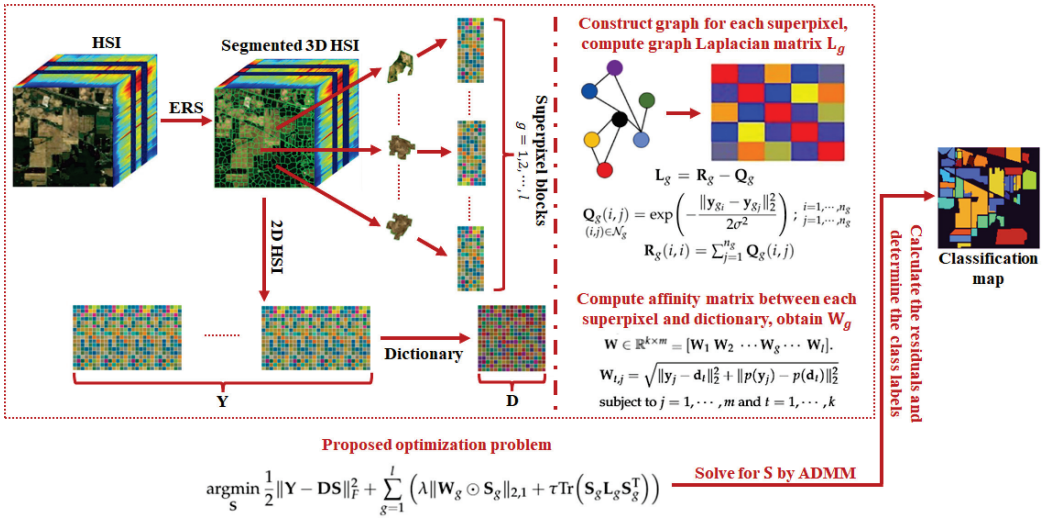


Figure 1. Flowchart of the proposed SJSGR method.

Algorithm 1 ADMM-based solution for the proposed SJSGR

Input: $\mathbf{Y}, \mathbf{D}, \mathbf{W}, \gamma, \lambda, \tau, \mu = 0.1, \epsilon = 0.1$

Initialization: $\mathbf{B} = (\mathbf{D}^{\mathrm{T}}\mathbf{D} + 2\mathbf{I})^{-1}, \mathbf{S}^{(0)} = \mathbf{B}\mathbf{D}^{\mathrm{T}}\mathbf{Y}, \mathbf{V}_1^{(0)} = \mathbf{D}\mathbf{S}, \mathbf{V}_2^{(0)} = \mathbf{V}_3^{(0)} = \mathbf{S}, \mathbf{P}_1^{(0)} = \mathbf{P}_2^{(0)} = \mathbf{P}_3^{(0)} = 0, u = 0$

for $g = 1$ to l **do**
  $\mathbf{L}_g = \mathbf{R}_g - \mathbf{Q}_g$
end for
repeat
  $\mathbf{S}^{(u+1)} = \mathbf{B}\left[\mathbf{D}^{\mathrm{T}}(\mathbf{V}_1^{(u)} + \mathbf{P}_1^{(u)}) + (\mathbf{V}_2^{(u)} + \mathbf{P}_2^{(u)}) + (\mathbf{V}_3^{(u)} + \mathbf{P}_3^{(u)})\right]$
  $\mathbf{V}_1^{(u+1)} = \frac{1}{1+\mu}\left(\mathbf{Y} + \mu(\mathbf{D}\mathbf{S}^{(u+1)} - \mathbf{P}_1^{(u)})\right)$
  for $g = 1$ to l **do**
    $\mathbf{V}_{2_g}^{(u+1)} = \text{vector_soft_row}(\mathbf{S}_g^{(u+1)} - \mathbf{P}_{2_g}^{(u)}, \lambda/\mu, \mathbf{W}_g)$
    $\mathbf{V}_{3_g}^{(u+1)} = \mu(\mathbf{S}_g^{(u+1)} - \mathbf{P}_{3_g}^{(u)})(2\tau\mathbf{L}_g + \mu\mathbf{I})^{-1}$
  end for
  $\mathbf{P}_1^{(u+1)} = \mathbf{P}_1^{(u)} - \mathbf{D}\mathbf{S}^{(u+1)} + \mathbf{V}_1^{(u+1)}$
  $\mathbf{P}_2^{(u+1)} = \mathbf{P}_2^{(u)} - \mathbf{S}^{(u+1)} + \mathbf{V}_2^{(u+1)}$
  $\mathbf{P}_3^{(u+1)} = \mathbf{P}_3^{(u)} - \mathbf{S}^{(u+1)} + \mathbf{V}_3^{(u+1)}$
  $\text{res_p} = \left\|\left[\mathbf{V}_1^{(u+1)}; \mathbf{V}_2^{(u+1)}; \mathbf{V}_3^{(u+1)}\right] - \left[\mathbf{D}\mathbf{S}^{(u+1)}; \mathbf{S}^{(u+1)}; \mathbf{S}^{(u+1)}\right]\right\|_F$
  $\text{res_d} = \left\|\left[\mathbf{V}_1^{(u+1)}; \mathbf{V}_2^{(u+1)}; \mathbf{V}_3^{(u+1)}\right] - \left[\mathbf{V}_1^{(u)}; \mathbf{V}_2^{(u)}; \mathbf{V}_3^{(u)}\right]\right\|_F$
  Update iteration: $u \leftarrow u + 1$
until $(\text{res_p} \le \epsilon \quad \text{OR} \quad \text{res_d} \le \epsilon)$

Output: The coefficient matrix $\mathbf{S}$.

Note: The function $\text{vector_soft_row}(\mathbf{A}, \psi, \mathbf{Z})$ provides weighted row-wise sparsity. Let $\mathbf{a}$ and $\mathbf{z}$ be the arbitrary row-vectors of $\mathbf{A}$ and $\mathbf{Z}$, respectively. Then, the function evaluates $\mathbf{a} \mapsto \frac{\max(\|\mathbf{a}\|_2 - \psi\|\mathbf{z}\|_2, 0)}{\max(\|\mathbf{a}\|_2 - \psi\|\mathbf{z}\|_2, 0) + \psi\|\mathbf{z}\|_2}\mathbf{a}$.

3. Experimental Results

3.1. Datasets

Experiments are conducted on three different hyperspectral datasets called Indian Pines (IP) [42], Pavia University (PU) [43] and Fanglu (FL) [44]. A summary of these datasets is presented in Table 1. It is seen that each dataset has distinct characteristics in terms of size, resolution, available classes and number of samples. Additionally, these are acquired by the different sensors, such as the Airborne Visible/Infrared Imaging Spectrometer

https://doi.org/10.3390/rs18162699

(AVIRIS), Reflective Optics System Imaging Spectrometer (ROSIS) and Aerosol Pushbroom Hyperspectral Imager (APHI) for the IP, PU and FL datasets, respectively. The use of datasets with different characteristics in HSIC is important because the accuracy results on these datasets demonstrate the generalization capability of the methods. More detailed information about the IP, PU and FL datasets including class names and available number of samples, are given in Table 2, Table 3 and Table 4, respectively. Also, corresponding ground-truth (GT) images of the IP, PU and FL datasets are presented in Figure 2.

Table 1. Comparison of the IP, PU and FL datasets.

Dataset	Sensor	Size	Spectral Res.	Spatial Res.	Num. of Classes	Num. of Samples
IP	AVIRIS	145 × 145 × 200	∼10 nm	20 m	16	10,366
PU	ROSIS	610 × 340 × 103	∼4 nm	1.3 m	9	42,776
FL	APHI	512 × 348 × 80	∼8 nm	2.25 m	10	53,734

Table 2. IP class distributions, with background colors correspond to the ground-truth class colors.

Class	Name	Samples
1	Alfalfa	54
2	Corn-notill	1434
3	Corn-mintill	834
4	Corn	234
5	Grass-pasture	497
6	Grass-trees	747
7	Grass-pasture-mowed	26
8	Hay-windrowed	489
9	Oats	20
10	Soybean-notill	968
11	Soybean-mintill	2468
12	Soybean-clean	614
13	Wheat	212
14	Woods	1294
15	Buildings-Grass-Trees-Drives	380
16	Stone-Steel-Towers	95

Table 3. PU class distributions, with background colors correspond to the ground-truth class colors.

Class	Name	Samples
1	Asphalt	6631
2	Meadows	18,649
3	Gravel	2099
4	Trees	3064
5	Painted metal sheets	1345
6	Bare soil	5029
7	Bitumen	1330
8	Self-Blocking Bricks	3682
9	Shadows	947

Table 4. FL class distributions, with background colors correspond to the ground-truth class colors.

Class	Name	Samples
1	Masson pine	5806
2	Bamboo forest	2318
3	Tea plant	28,428
4	Reed	214
5	Paddy	6809
6	Sweet potato	817
7	Caraway	429
8	Weed	1861
9	Water body	6141
10	Building/Road	911



(**a**) (**b**) (**c**)

Figure 2. GT images of (**a**) IP, (**b**) PU and (**c**) FL datasets, respectively.

3.2. Performance Metrics

The performance of the proposed SJSGR is compared with several methods in the literature such as JSRC [8], CCJSR [14], MSSGF [19], LCJS-WLR [21], SGR [24], SGL [26] and WSGR [25]. It should be noted that the source codes of methods JSRC, CCJSR, SGR, SGL and WSGR are publicly available, while methods MSSGF and LCJS-WLR are the authors' own methods. The codes for the other methods mentioned in the literature review of the sparse and graph-based studies, except for [28], were not publicly available, and therefore were not used in the experimental study. As for [28], because its methodology differs in scope from the proposed method and other methods compared, it was not included in the experimental study.

The performance of all methods are evaluated and compared in terms of class-based accuracy, overall accuracy (OA), average accuracy (AA) and Cohen's kappa coefficient (κ) metrics. Here, class accuracy represents the ratio of correctly classified samples to the total number of samples in a given class. AA denotes the sum of the class-based accuracies divided by the total number of classes. OA is the ratio of correctly classified samples to the total number of samples. The kappa coefficient is a statistical metric that assesses classification performance by normalizing the difference between OA and the accuracy expected by chance. All experiments are repeated ten times using randomly chosen training dictionaries, and then the results are averaged to minimize the bias.

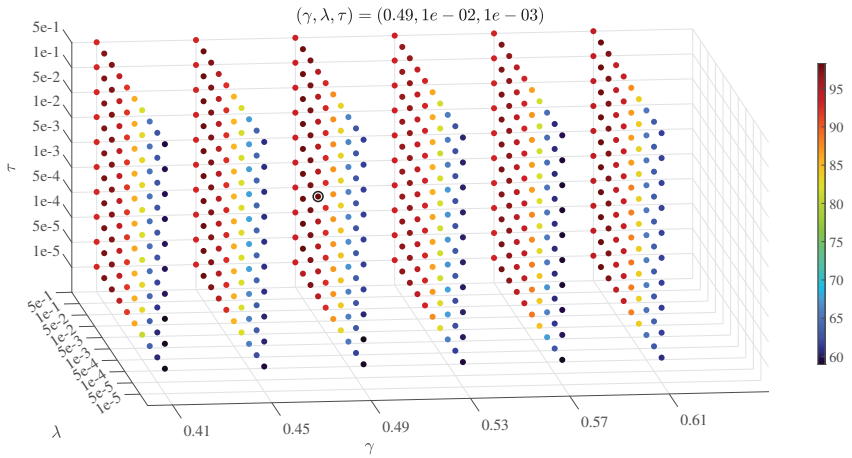
3.3. Parameter Tuning of SJSGR

The corresponding parameters of the other methods are fine-tuned for all datasets to perform a fair comparison. As for the SJSGR, it has several important parameters: γ for

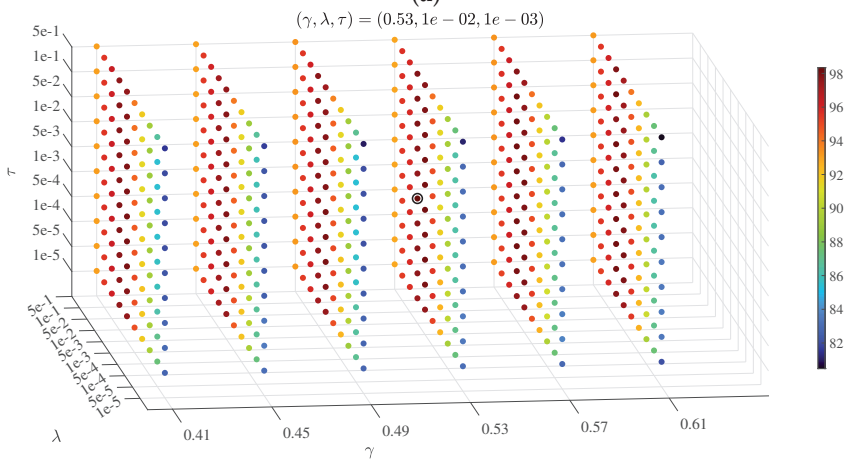
the number of superpixels, and λ and τ for the joint-sparse and graph regularization terms, respectively. To specify the optimal parameters for each dataset, we applied a grid search over the ranges of $\gamma = \{0.41, 0.45, 0.49, 0.53, 0.57, 0.61\}$ and $(\lambda, \tau) = \{10^{-5}, 5 \times 10^{-5}, 10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 10^{-2}, 5 \times 10^{-2}, 10^{-1}, 5 \times 10^{-1}\}$, and then we record the OA results. Figure 3 presents the observed results in terms of scatter plots. As can be seen from the colorbars indicated in the plots, high results were attained at many points. However, the highest OA value was obtained at only one intersection point for each dataset. In accordance with these observations given in Figure 3, γ was set to 0.49 for the IP, and 0.53 for the PU and FL datasets, respectively. For λ, it was adjusted to 10^{-2} for all datasets. Finally, the parameter τ was tuned as 10^{-3} for the IP and PU datasets, and 10^{-4} for the dataset FL, respectively. The optimal parameters of the datasets are given in Table 5. A closer examination of Figure 3 reveals that λ has the strongest influence on the OA results among all parameters. This effect is followed by the γ parameter. On the other hand, τ does not noticeably affect the OA results. However, it preserves their consistency. Since this parameter is responsible for graph regularization, its effect should be further analyzed through an ablation study.

Table 5. Optimal parameters of SJSGR across the datasets.

Parameter	IP	PU	FL
γ	0.49	0.53	0.53
λ	10^{-2}	10^{-2}	10^{-2}
τ	10^{-3}	10^{-3}	10^{-4}

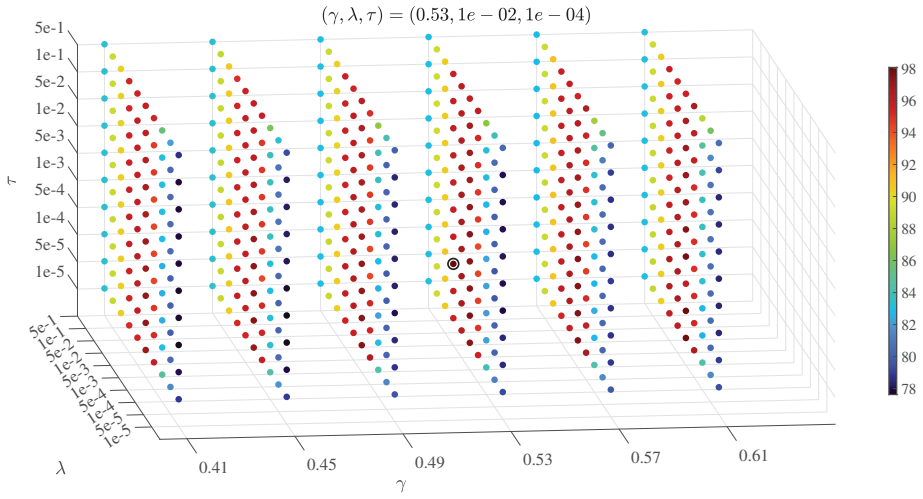


(**a**)



(**b**)

Figure 3. *Cont.*



(c)

Figure 3. Scatter plots showing the impact of parameter variations on the OA results across three datasets: (**a**) IP, (**b**) PU and (**c**) FL, respectively. The best results are marked with black circles for clarity.

3.4. Ablation Study

In order to assess the influence of the graph regularization term controlled by the τ parameter, we performed an ablation study by comparing the proposed SJSGR method with a modified version called SJS where the graph regularization term is discarded. Table 6 presents the results obtained on the datasets in terms of OA and AA metrics. It is seen that a slight increase is obtained in the OA and AA results for the IP dataset. As for the PU dataset, the OA and AA results are improved by approximately 1% and 4.5%, respectively. This improvement is due to the use of the graph regularization term. For the FL dataset, OA shows a slight enhancement, whereas AA improves by approximately 1%. Therefore, these ablation studies demonstrate that the graph regularization term improves the results slightly or moderately depending on the dataset, thus ensuring consistency.

Table 6. Ablation study of the proposed SJSGR across the datasets. Bold and underlined values indicate the best and second-best results, respectively.

Metric	SJS			SJSGR		
	IP	**PU**	**FL**	**IP**	**PU**	**FL**
OA	97.90	97.10	98.02	**98.12**	**98.04**	**98.26**
AA	98.17	91.89	93.92	**98.45**	**96.36**	**94.89**

3.5. Results

Here, both quantitative and qualitative results of the IP, PU and FL datasets are given, respectively.

3.5.1. Indian Pines

Table 7 presents the achieved accuracy results of the methods under 10% labeled training samples per class for the IP dataset. It is seen that JSRC provides the lowest accuracy results compared to the other methods. Incorporating the correlation coefficients between the training samples and local patches into the residual calculation, CCJSR outperforms JSRC by nearly 1% and 4% in terms of OA and AA. By considering local low-rank structure of each superpixel and imposing locality-preserving framework, LCJSWLR achieves a similar OA result compared to CCJSR, but obtains about 2% higher AA. The performance of LCJSWLR is followed by SGR, which provides nearly 1.2% higher OA, but attains a

lower AA than LCJSWLR. A weighted extension of SGR, called WSGR, provides slightly higher OA and approximately 7% better AA than SGR. Another graph regularization-based method called SGL has competitive accuracy results compared to WSGR. While SGL slightly outperforms WSGR in terms of OA, WSGR exceeds the AA result of SGL relatively. A similar condition is also observed among SGL and MSSGF. It is seen that MSSGF provides slightly higher OA than SGL, but SGL has nearly 2.5% better AA than MSSGF. Regarding the proposed SJSGR, it exceeds the best OA result by approximately 0.7% and the best AA result by about 0.5% in other methods. Therefore, the proposed method achieves the highest OA and AA results for the IP dataset.

Table 7. Accuracy results of the methods across the IP dataset. Bold and underlined values indicate the best and second-best results, respectively, while background colors correspond to the ground-truth class colors.

IP (10% LABELED TRAINING SAMPLES PER CLASS)									
Class	**Train/Test**	**JSRC**	**CCJSR**	**MSSGF**	**LCJSWLR**	**SGR**	**WSGR**	**SGL**	**SJSGR**
1	6/48	95.00	94.17	98.33	93.33	89.58	97.58	96.17	98.33
2	144/1290	95.61	94.54	97.52	92.99	95.76	95.18	94.43	96.65
3	84/750	95.09	94.64	97.25	92.85	94.83	93.21	96.32	96.48
4	24/210	92.48	94.67	96.86	96.76	93.81	100	95.52	97.76
5	50/447	92.57	94.81	96.51	93.69	95.75	95.53	98.17	98.52
6	75/672	95.12	98.39	98.66	97.65	98.78	97.59	99.43	99.32
7	3/23	69.57	60.87	96.52	83.48	93.91	100	100	100
8	49/440	99.50	99.82	100	98.41	100	100	99.77	99.77
9	2/18	8.89	40.00	64.44	38.89	10.00	100	100	100
10	97/871	91.02	94.28	92.95	90.40	92.38	94.75	94.23	95.65
11	247/2221	96.01	96.21	97.71	98.00	98.21	98.22	98.38	98.82
12	62/552	89.93	89.96	97.14	92.97	96.84	96.25	94.53	96.87
13	22/190	86.42	96.95	99.58	99.47	99.69	99.89	99.89	99.47
14	130/1164	98.61	99.18	99.78	100	99.90	99.48	99.91	100
15	38/342	93.80	93.80	96.67	98.60	98.16	100	98.88	98.72
16	10/85	75.53	92.71	97.65	97.18	99.76	99.29	93.18	98.82
OA (±std)		94.53	95.59	<u>97.46</u>	95.76	96.91	97.18	97.31	**98.12**
		±0.38	±0.70	±0.29	±0.40	±0.30	±0.52	±0.28	±0.25
AA (±std)		85.95	89.69	95.47	91.54	91.09	<u>97.94</u>	97.43	**98.45**
		±2.86	±1.72	±2.86	±2.55	±0.56	±0.36	±0.31	±0.29
κ (±std)		93.77	94.97	<u>97.14</u>	95.17	96.51	96.82	96.94	**97.83**
		±0.43	±0.80	±0.33	±0.46	±0.34	±0.59	±0.32	±0.27

In addition to the given quantitative findings regarding the IP dataset, we also present the qualitative observations for all methods. Figure 4 demonstrates the classification maps of the methods with OAs under the same train/test splitting percentages adjusted for this dataset. It is seen that the proposed SJSGR produces a classification map which is closer to the corresponding ground-truth (GT) map of the dataset. Compared to the other methods, SJSGR provides smoother intra-class local regions and sharper inter-class transitions. These visual representations also reveal the efficacy of the proposed method.

Apart from the aforementioned results, we also performed an experiment to observe the OA results of the methods under low-training-sample distributions such as 1%, 4% and 7% labeled samples per class. As seen from Figure 5, performances of the methods gradually increase depending on the training percentages. For 1% sample distribution, JSRC achieves the lowest accuracy, then followed by CCJSR and SGR, respectively. It can be said that MSSGF and LCJSWLR appear to yield average-level results among all methods.

https://doi.org/10.3390/rs18162699

While the highest OA is obtained by SJSGR, this achievement is closely followed by SGL and WSGR, respectively. For 4% sample tuning, JSRC again attains the lowest OA, which is traced by CCJSR and SGR, respectively. While the other methods except SJSGR provide close results between 92.5% and 95%, SJSGR obtains approximately 1% higher OA than SGL, which provides the closest result. Finally, for 7% sampling setting, it is seen that all methods attain OA results above 90%. Although the results of methods of this setting are close, the proposed SJSGR still appears to maintain the highest results.

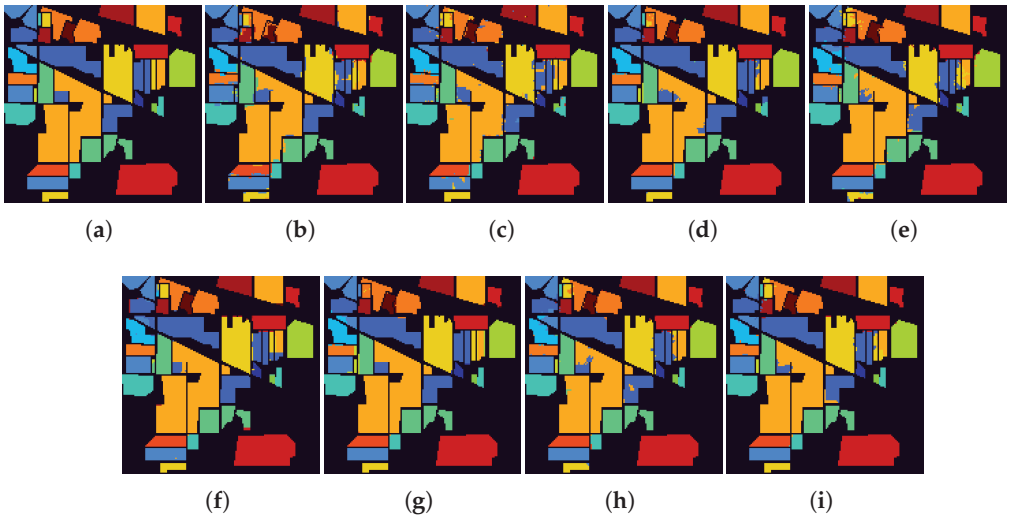


Figure 4. IP dataset: Classification maps of the methods with the following OAs; (**a**) GT, (**b**) JSRC (94.58%), (**c**) CCJSR (95.69%), (**d**) MSSGF (97.52%), (**e**) LCJSWLR (95.86%), (**f**) SGR (97.01%), (**g**) WSGR (97.25%), (**h**) SGL (97.39%) and (**i**) SJSGR (98.09%), respectively.

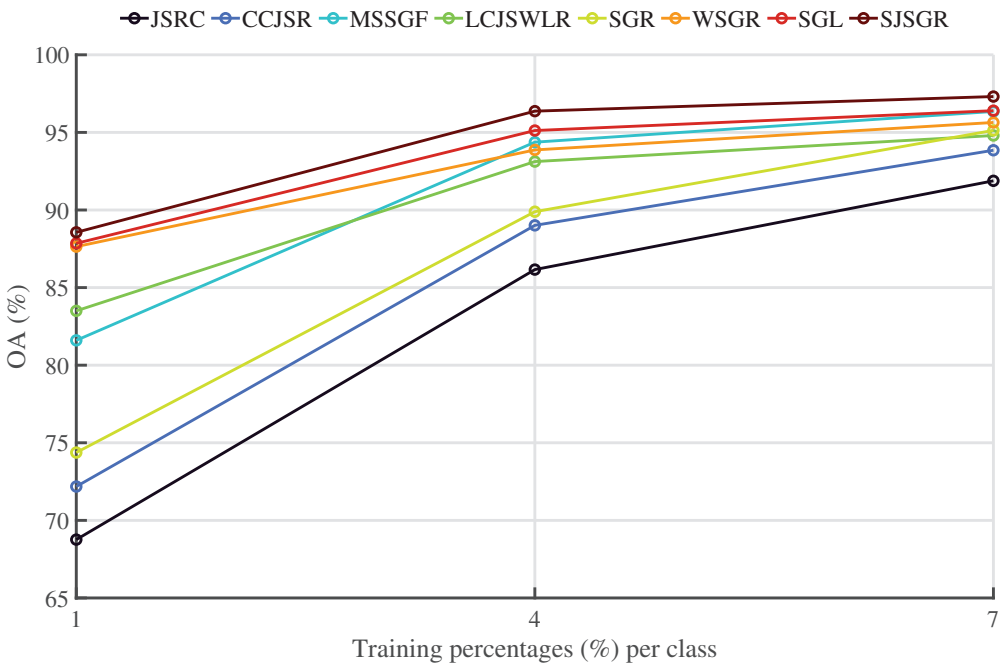


Figure 5. IP dataset: OA results of the methods with varying percentages of training samples per class.

3.5.2. Pavia University

Table 8 demonstrates the accuracy results of the methods under 1% labeled training samples per class for the PU dataset. Similar to the case of IP dataset, JSRC has the lowest OA and AA results compared to the others. The results of this method are followed by CCJSR, which provides nearly 1.2% and 7% higher OA and AA results, respectively.

Compared to CCJSR, SGR has about 0.5% and 1% higher OA and AA results. Results of this method are traced by MSSGF, which obtains nearly 4.8% and 1.5% higher OA and AA values. It can be seen that OA of MSSGF is followed by SGL, WSGR and LCJSWLR, providing almost 2.4%, 2.6% and 3% higher results, respectively. However, as for the AA result of MSSGF, it is almost 3.5%, 6% and 6.5% outperformed by LCJSWLR, SGL and WSGR, respectively. Compared to the methods analyzed so far, it is seen that the proposed SJSGR achieves the highest OA and AA results. For example, it outperforms the best OA result by approximately 1.2% and the best AA result by about 2.3% in other methods.

Table 8. Accuracy results of the methods across the PU dataset. Bold and underlined values indicate the best and second-best results, respectively, while background colors correspond to the ground-truth class colors.

PU (1% LABELED TRAINING SAMPLES PER CLASS)									
Class	**Train/Test**	**JSRC**	**CCJSR**	**MSSGF**	**LCJSWLR**	**SGR**	**WSGR**	**SGL**	**SJSGR**
1	67/6564	69.66	57.36	87.20	94.38	90.48	92.31	97.26	95.77
2	187/18,462	98.20	98.74	99.93	99.89	99.76	99.87	99.94	99.96
3	21/2078	83.56	80.35	90.18	99.63	84.80	93.93	78.65	99.77
4	31/3033	55.78	89.16	73.91	90.00	88.63	87.13	80.76	91.51
5	14/1331	91.10	99.53	99.68	99.68	100	100	98.39	99.25
6	51/4978	97.97	96.59	96.65	100	69.76	99.60	98.74	100
7	14/1316	89.38	61.23	99.71	98.56	97.17	100	99.24	99.64
8	37/3645	92.02	91.09	98.47	98.40	54.41	93.19	93.49	96.58
9	10/937	25.98	91.85	42.67	39.72	89.88	80.66	96.20	84.74
OA (±std)		87.37	88.53	93.82	<u>96.82</u>	89.07	96.47	96.25	**98.04**
		±0.24	±0.76	±0.62	±0.39	±1.48	±0.45	±0.47	±0.20
AA (±std)		78.18	85.10	87.60	91.14	86.10	<u>94.08</u>	93.63	**96.36**
		±1.23	±1.84	±0.84	±1.51	±1.86	±1.32	±0.90	±0.97
κ (±std)		83.24	84.81	91.71	<u>95.77</u>	85.27	95.32	95.01	**97.40**
		±0.33	±1.02	±0.85	±0.53	±1.98	±0.61	±0.63	±0.27

Figure 6 shows the classification maps of the methods with OAs using 1% labeled training samples per class for the PU dataset. It should be noted here that this dataset has both small and large local regions as seen from the corresponding GT image. Therefore, capturing and effectively utilizing spatial information from all these regions is not an easy task for the methods. However, it is clear that the proposed SJSGR generates a classification map which is more closer to the GT map compared to the others. When the classification maps are examined, it is seen that local regions, especially those with small areas, are often inaccurately classified by other methods, whereas SJSGR classifies these regions with high accuracy, except for class 9. Hence, it can be said that the proposed method has more uniform local regions and sharper boundaries across different classes for the PU dataset, compared to the other methods.

Besides both the quantitative and qualitative results above, Figure 7 shows the OA results of the methods under low-training-sample settings such as 0.1%, 0.4% and 0.7% labeled samples per class for the PU dataset. For the 0.1% sampling setting, it is seen that JSRC obtains the lowest OA, which is then followed by CCJSR, MSSGF, SGR and WSGR, respectively. Starting from SGL, the OA results are observed above 85%, where the proposed SJSGR provides the highest OA. For the 0.4% sample distribution, JSRC has again the lowest OA, which is followed by CCJSR, MSSGF and SGR, respectively. The results of WSGR, LCJSWLR, SGL and SJSGR are close to each other; however, the OA of SJSGR is closer to 95% compared to the others. When the amount of labeled training samples is set to 0.7%, it is seen that the OA results of the methods gradually increase. While the

https://doi.org/10.3390/rs18162699

OAs of JSRC, CCJSR and SGR remain below 90%, the result of MSSGF is slightly above that. At OAs of around 95%, WSGR, SGL and LCJSWLR obtain quite similar results. As for the proposed SJSGR, it attains an OA of approximately 97.5%, which is the highest result among all observations.

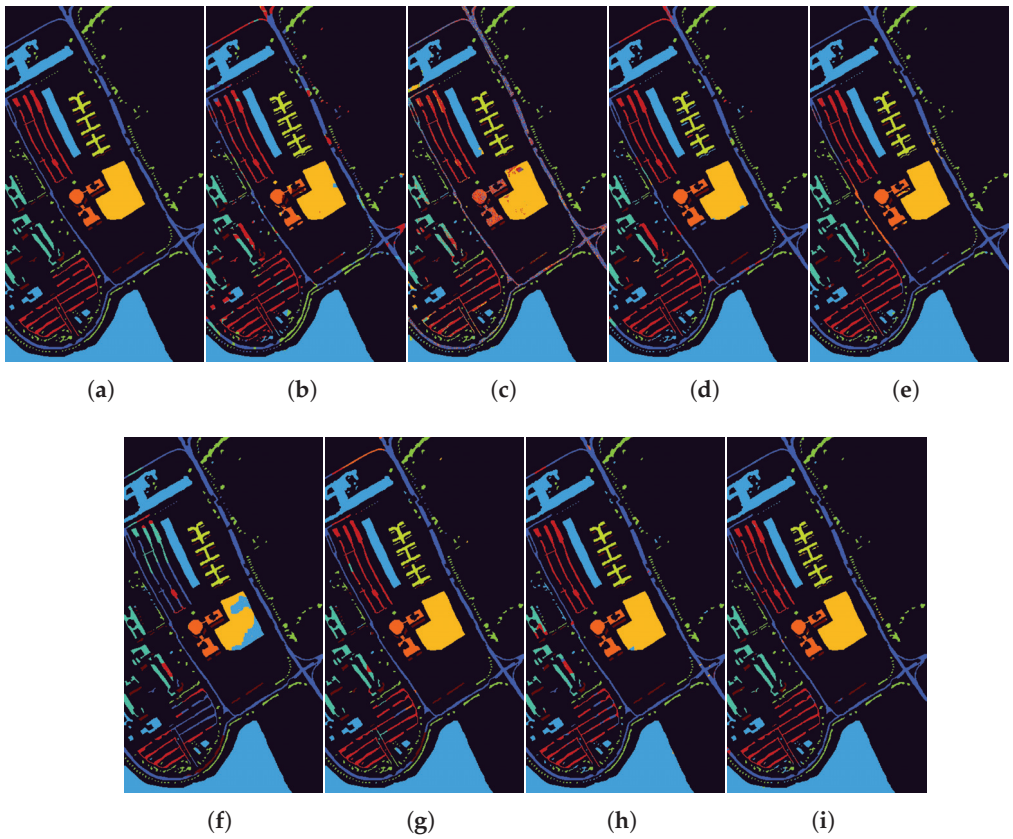


Figure 6. PU dataset—Classification maps of the methods with the following OAs: (**a**) GT, (**b**) JSRC (87.35%), (**c**) CCJSR (88.32%), (**d**) MSSGF (93.95%), (**e**) LCJSWLR (96.63%), (**f**) SGR (88.51%), (**g**) WSGR (96.68%), (**h**) SGL (96.43%) and (**i**) SJSGR (98.12%), respectively.

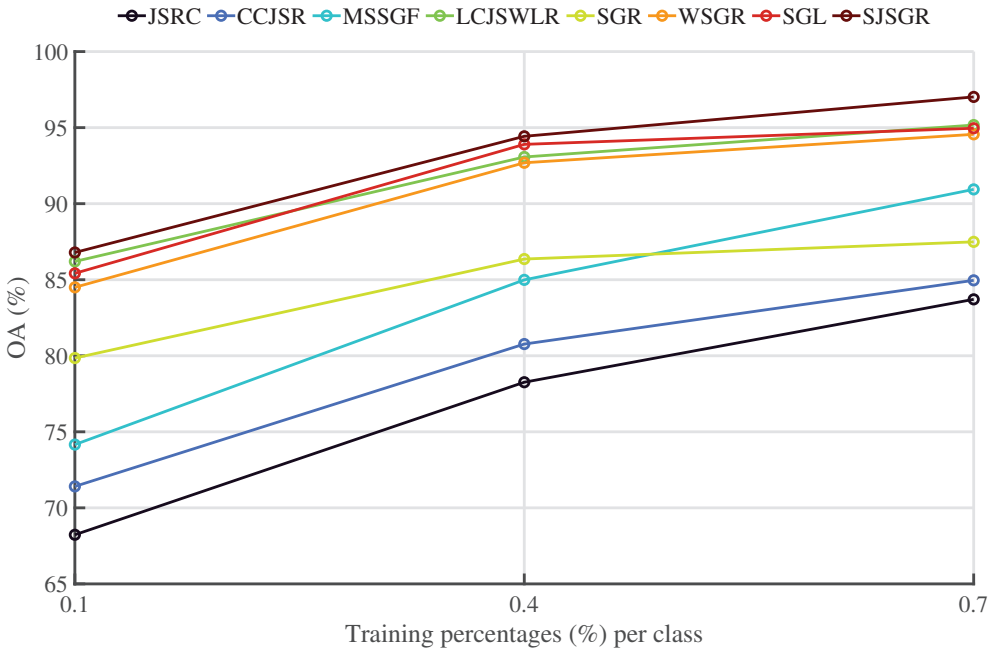


Figure 7. PU dataset: OA results of the methods with varying percentages of training samples per class.

https://doi.org/10.3390/rs18162699

3.5.3. Fanglu

Table 9 shows the accuracy results of the methods over the FL dataset using 1% labeled training samples per class. It can be seen that JSRC and SGR provide quite similar OA results, but JSRC attains nearly 0.7% higher AA than SGR. In terms of OA, WSGR follows the results of these two methods. As for MSSGF, it obtains a quite similar OA compared to WSGR. Following MSSGF, CCJSR provides approximately 0.4% higher OA, and then this improvement is followed by LCJSWLR and SGL, respectively. Additionaly, it can be said that LCJSWLR and SGL achieve almost identical OA results. As for the AA results of the methods, MSSGF and CCJSR outperform JSRC by about 8%. Following these methods, approximately 3% improvement is achieved by SGL. As for the AA of LCJSWLR, it is nearly 0.7% higher than the one of SGL. As seen from the AA of the proposed SJSGR, it outperforms the AA of LCJSWLR by about 2.5%. Finally, WSGR obtains the highest AA result for the FL dataset. It provides approximately 2% higher AA than the proposed SJSGR. Unlike the other two datasets, IP and PU, it can be seen that there is not much consistency between the OA and AA results of the methods in this dataset. The relatively inconsistent performance observed on this dataset may be attributed to its class imbalance, fewer spectral features and spectral similarity between classes, which are the challenging cases for the classifiers. However, when the results of the proposed SJSGR method are analyzed, it is clear that it provides the highest OA and the second highest AA results.

Table 9. Accuracy results of the methods across the FL dataset. Bold and underlined values indicate the best and second-best results, respectively, while background colors correspond to the ground-truth class colors.

FL (1% LABELED TRAINING SAMPLES PER CLASS)									
Class	**Train/Test**	**JSRC**	**CCJSR**	**MSSGF**	**LCJSWLR**	**SGR**	**WSGR**	**SGL**	**SJSGR**
1	59/5747	96.03	96.63	96.21	99.27	99.99	100	97.85	99.39
2	24/2294	90.35	90.61	83.12	95.19	59.78	94.87	96.15	96.20
3	285/28,143	99.44	99.46	99.98	99.16	97.76	94.57	99.12	99.53
4	3/211	91.75	88.72	100	91.28	2.09	100	100	98.67
5	69/6740	89.97	95.51	97.15	99.57	100	99.99	99.99	99.75
6	9/808	62.77	75.52	75.17	78.76	87.28	95.32	72.33	81.67
7	5/424	56.18	73.02	78.58	99.42	99.91	100	91.20	100
8	19/1842	72.73	77.99	66.54	70.62	52.33	87.34	82.46	81.10
9	62/6079	97.62	99.18	98.65	98.15	99.05	99.05	98.75	98.65
10	10/901	48.39	90.28	90.34	92.08	100	96.56	78.65	93.96
OA (±std)		94.55	96.73	96.37	97.43	94.54	96.23	<u>97.45</u>	**98.26**
		±0.31	±0.30	±0.44	±0.52	±0.75	±0.87	±0.65	±0.37
AA (±std)		80.52	88.69	88.58	92.35	79.82	**96.77**	91.65	<u>94.89</u>
		±1.79	±1.78	±1.38	±1.38	±1.49	±0.90	±0.93	±1.29
κ (±std)		91.77	95.09	94.54	96.33	92.02	94.53	<u>96.42</u>	**97.38**
		±0.49	±0.46	±0.69	±0.78	±1.09	±1.23	±0.96	±0.56

The visual representations of the methods under 1% labeled training samples per class are given in Figure 8 for the FL dataset. Similar to the case of PU, this dataset has also both small and large local areas as seen from the corresponding GT image. As seen from the OA results, the proposed SJSGR obtains a classification map which is more similar to the GT map compared to remaining ones. Examining the results in Table 9, it is clear that the reason for the higher AA result of WSGR compared to SJSGR is due to its effective classification performance in class 6. This effect is also clearly visible when analyzing the maps of methods WSGR and SJSGR, where relatively small regions of this class are

more correctly classified by WSGR. However, it can be said that SJSGR provides more homogeneous spatial regions except for this class.

Figure 9 gives the performance comparison of the methods in terms of OA by varying the percentages of labeled training samples per class such as 0.1%, 0.4% and 0.7% for the FL dataset. As seen for the 0.1% sampling percentage, JSRC yields the lowest OA among all methods. Following this, MSSGF, CCJSR, WSGR and SGR appear to provide average-level OA results, respectively. It can be seen that the OAs of these methods are approximately between 83.5% and 86%. As for the remaining ones, called SGL, LCJSWLR and SJSGR, the results are observed as above 90%, where SJSGR achieves the highest one at nearly 92%. For the 0.4% labeled samples, it is clear that the classification performances of all methods gradually enhance. While JSRC again obtains the lowest OA, the moderate ones called MSSGF, CCJSR, SGR and WSGR yield to provide the average-level OA results, respectively. Following these, LCJSWLR and SGL have quite close OA results at around below 96%. Then, SJSGR obtains the highest OA which is slightly higher than 96%. For the last percentage rate, 0.7%, all methods provide high OA results, and also the results become closer to each other. While JSRC has the lowest OA over again, SGR, WSGR, MSSGF and CCJSR follow these results, respectively. As for the others, LCJSWLR and SGL have quite similar OA results around 97%, while SJSGR attains the highest OA close to 98%.

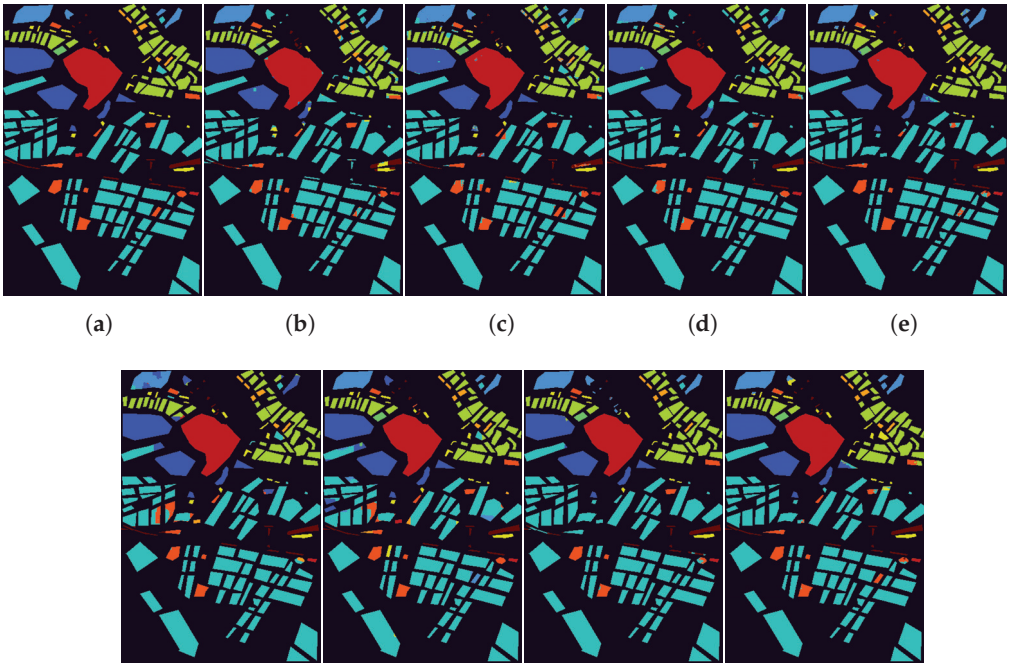


Figure 8. FL dataset: Classification maps of the methods annotated with OAs. From left to right: (**a**) GT, (**b**) JSRC (94.73%), (**c**) CCJSR (96.78%), (**d**) MSSGF (96.38%), (**e**) LCJSWLR (97.48%), (**f**) SGR (94.61%), (**g**) WSGR (96.32%), (**h**) SGL (97.51%) and (**i**) SJSGR (98.29%), respectively.

3.5.4. Computational Complexity and Time Analysis

The proposed SJSGR method has several steps to complete the whole classification process. Tables 10–12 present the steps, and also give the approximate computation times of each one for the IP, PU and FL datasets, respectively, using 10 samples per class. It is seen that a significant portion of the running time is caused by the ADMM iterations. This iteration process includes the steps between 5 and 15 given in Algorithm 1. The complexities of steps 5 and 6 are calculated as $\mathcal{O}(km(b+k))$ and $\mathcal{O}(kmb)$, respectively. As for steps 8 and 9, which need to be computed for each superpixel, the corresponding time complexities are given as $\mathcal{O}(kn_g)$ and $\mathcal{O}(kn_g^2)$, respectively. It should be noted that the term

https://doi.org/10.3390/rs18162699

$(2\tau\mathbf{L}_g + \mu\mathbf{I})^{-1}$ in step 9 is computed once during the graph construction stage, and then fixed over the course of the iterations. Due to the presence of l superpixels in total, these can be rewritten as $\mathcal{O}\left(k\sum_{g=1}^{l} n_g\right)$ and $\mathcal{O}\left(k\sum_{g=1}^{l} n_g^2\right)$, accordingly. Step 11 has a complexity of $\mathcal{O}(kmb)$, and steps 12 and 13 have $\mathcal{O}(km)$. Finally, the computational complexities of steps 14 and 15 require $\mathcal{O}(kmb)$ and $\mathcal{O}(bm + km)$, respectively. Therefore, the proposed SJSGR method has an overall complexity of $\mathcal{O}(km(b+k)) + \mathcal{O}(kmb) + \mathcal{O}\left(k\sum_{g=1}^{l} n_g\right) + \mathcal{O}\left(k\sum_{g=1}^{l} n_g^2\right) + \mathcal{O}(kmb) + \mathcal{O}(km) + \mathcal{O}(kmb) + \mathcal{O}(bm + km)$ for the per ADMM iteration, consuming the most of the total running time. Since $\sum_{g=1}^{l} n_g$ is equal to the total pixel count m, the term $\mathcal{O}\left(k\sum_{g=1}^{l} n_g\right)$ is reduced to $\mathcal{O}(km)$. By retaining only the dominant ones, the complexity of SJSGR per ADMM iteration can be given as $\mathcal{O}\left(km(b+k) + k\sum_{g=1}^{l} n_g^2\right)$.

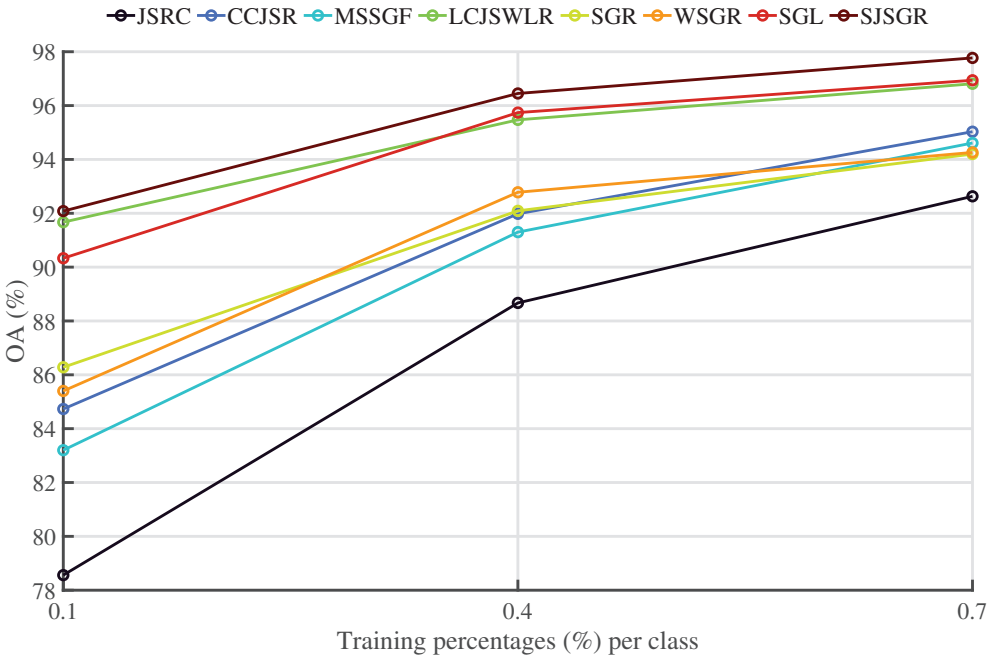


Figure 9. FL dataset: OA results of the methods with varying percentages of training samples per class.

Table 10. Computation time of each module of the SJSGR method for the IP dataset using 10 labeled samples per class.

Module	Processing Time (s)	Percentage (%)
Superpixel construction	0.06	0.43
Weighting matrix	0.84	6.02
Graph construction	0.34	2.44
ADMM initializations	0.01	0.07
ADMM iterations	12.58	90.11
Labeling	0.13	0.93
Total	13.96	100

The computational efficiency of the proposed and competing methods was evaluated by comparing their execution times. All experiments were conducted on the same hardware platform consisting of an Intel Core i7-12700K processor (3.60 GHz) and 128 GB of RAM to ensure a fair comparison. The execution times obtained with ten labeled samples per class are presented in Table 13 for all datasets. It is clearly observed that the LCJSWLR method requires the highest computational time across all datasets. This is mainly due to its complex optimization procedure and, more importantly, the low-rank minimization

performed within each superpixel, which significantly increases the computational cost. The JSRC method achieves the lowest computational time on the IP dataset, whereas the SGR method is faster on the other datasets. When the computational times of the methods that achieve classification accuracies closest to the proposed SJSGR method, namely SGL, WSGR, LCJSWLR, and MSSGF, are examined, the LCJSWLR is the slowest among them, as mentioned before. Compared with the SGL method, the SJSGR exhibits slightly lower computational times for all datasets. Regarding the remaining methods, WSGR and MSSGF are both several times faster than the SJSGR method across the datasets.

Table 11. Computation time of each module of the SJSGR method for the PU dataset using 10 labeled samples per class.

Module	Processing Time (s)	Percentage (%)
Superpixel construction	0.97	0.83
Weighting matrix	4.49	3.85
Graph construction	3.13	2.68
ADMM initializations	0.05	0.04
ADMM iterations	107.61	92.23
Labeling	0.43	0.37
Total	116.68	100

Table 12. Computation time of each module of the SJSGR method for the FL dataset using 10 labeled samples per class.

Module	Processing Time (s)	Percentage (%)
Superpixel construction	0.75	0.85
Weighting matrix	3.93	4.44
Graph construction	2.85	3.22
ADMM initializations	0.04	0.05
ADMM iterations	80.55	91.03
Labeling	0.37	0.41
Total	88.49	100

Table 13. Comparison of computational time (in seconds) of the methods using ten labeled samples per class across the datasets.

Dataset	JSRC	CCJSR	MSSGF	LCJSWLR	SGR	WSGR	SGL	SJSGR
IP	2.19	15.85	3.98	59.12	3.37	9.12	14.34	13.96
PU	25.14	58.48	39.46	522.95	17.58	26.51	122.39	116.68
FL	30.6	74.18	37.77	434.26	16.56	23.09	91.26	88.49

3.5.5. Comparison with the Deep Learning-Based Models

Although the proposed SJSGR method is based on a traditional machine learning classifer, its performance was also compared with the several deep learning-based methods such as SpectralFormer (https://github.com/danfenghong/IEEE_TGRS_SpectralFormer, accessed on 27 July 2026) [45], 3DSS−Mamba (https://github.com/IIP-Team/3DSS-Mamba, accessed on 27 July 2026) [46], HybridSN (https://github.com/gokriznastic/HybridSN, accessed on 28 July 2026) [47], SSEFN (https://github.com/liushuang963/SSEFN, accessed on 29 July 2026) [48] and DSFormer (https://github.com/YichuXu/DSFormer, accessed on 29 July 2026) [49]. In order to perform fair comparisons, the same training and test indices, utilized in the previous experiments, were adopted for each dataset. Specifically, 10% and 1% of the

samples from each class were employed for training in the IP and PU datasets, respectively, while the remaining samples used for testing. It should be noted that the FL dataset was not used for the comparison since the chosen DL-based models have no optimization parameters for this data. As for the IP and PU datasets, the training parameters of these models were tuned as the values provided in the related original articles. The results were recorded in terms of OA, AA and κ, respectively. Table 14 presents the results for both of the datasets. It is seen that the DSFormer achieves the highest OA and κ results, which are being closely followed by the SJSGR method for the IP dataset. As for the AA value, the SJSGR provides the best result, which is traced by the SSEFN model. Similarly, the DSFormer attains the highest metrics, which are followed by the SJSGR and SSEFN methods, respectively, for the PU dataset.

Table 14. Classification performance (%) of different methods on two hyperspectral datasets. Bold and underlined values indicate the best and second-best results, respectively.

Method	IP			PU		
	OA	**AA**	κ	**OA**	**AA**	κ
SpectralFormer	90.90	83.85	89.61	86.03	80.88	81.18
3DSS−Mamba	94.43	91.98	93.64	90.94	85.66	87.92
HybridSN	98.12	95.97	97.83	94.34	90.97	92.53
SSEFN	97.64	<u>97.91</u>	97.31	97.47	<u>96.78</u>	96.64
SJSGR	<u>98.14</u>	**98.45**	<u>97.88</u>	<u>98.04</u>	96.36	<u>97.40</u>
DSFormer	**98.27**	96.69	**98.02**	**98.55**	**97.52**	**98.08**

4. Discussion

The results obtained from the exhaustive experiments are discussed here. It should be noted that it is not possible to generalize the results of methods other than JSRC and the proposed SJSGR framework, because the performances of the methods vary depending on the datasets.

Throughout the experiments, it was observed that JSRC yields to provide lowest accuracy results. This method considers the spatial information of a HSI using a fixed-size patch constructed around a given test sample without assigning distinct weight to each neighbor. Such an assumption may cause misclassification of this test sample if it is around class borders. In such a case, the impact of pixels belonging to different classes may become more dominant. Apart from this, JSRC only applies a sparsity constraint which takes into account the spectral similarity of pixels, but ignores the local manifold structure of that patch. By incorporating the correlation coefficients between the training samples and local patches into the residual calculation, the CCJSR method outperforms JSRC across all datasets. This observation reveals the idea that such correlation may improve the results by inserting it into the optimization problem. As observed, MSSGF considerably outperforms CCJSR for the IP and FL datasets, but their corresponding results are quite competitive for the FL dataset. Since MSSGF utilizes multiscale superpixels, the spatial information is effectively employed in two ways. The first one is to obtain more homogeneous local regions through superpixels, and the second one is to capture local information in a broad range through multiscale superpixels. The use of these extensive spatial information with sparsity constraint enhances the results. The results indicate that LCJSWLR is superior to the MSSGF, except for the IP dataset. This method simultaneously employs weighted joint-sparse and low-rank regularizations in homogeneous regions obtained by single-scale superpixel segmentation. A unified utilization of such regularization terms enables the coefficient matrix to be more informative than the use of sparsity constraint only. Therefore, the results are improved compared to the MSSGF for the PU and FL datasets.

The reason for obtaining lower accuracy results for the IP dataset may be the ineffective use of spatial information.

Besides these sparsity-constraint-based methods, SGR is generally superior to JSRC and CCJSR, but is usually outperformed by LCJSWLR and MSSGF depending on the dataset. SGR computes the coefficient matrix of all pixels by introducing sparse-regularization only. In this stage, it does not consider the neighborhood information of these pixels. After that, it constructs a graph Laplacian matrix based on the coefficient matrix. By utilizing both coefficient and graph Laplacian matrices, a prediction function is evaluated. Due to ignoring neighborhood information and local manifold structure of the data, SGR yields limited performance. As an extension of SGR, the method WSGR provides higher accuracy results than SGR. Also, the results of WSGR become competitive with the ones of MSSGF and LCJSWLR. The reason for this observation could be the incorporation of spatial information in WSGR since it applies a weighting scheme to the central pixel and its neighborhoods in a fixed-size patch. In this way, each pixel of a HSI is transformed into the weighted one, and then the weighted pixels are processed using the same procedure as in SGR. As for the last graph regularization-based method, SGL, it provides competitive performance compared to the WSGR, MSSGF and LCJSWLR, which are the methods providing high accuracy results across the datasets. The satisfying results of SGL can be given as follows. It extracts additional descriptive features from each superpixel, and then these features are used to construct graphs between two linked superpixels. In this way, both local spatial information and global structure of the data are considered for the purpose of classification.

Based on the analyses conducted so far, the performances of methods vary depending on the dataset. For the proposed SJSGR, it was observed that the highest OA results are observed for all different training sample distributions across the datasets. Except for the FL dataset, the highest AA results are also obtained by SJSGR. The reasons behind this performance can be given as follows: It utilizes superpixels to capture the spatial information in a HSI, effectively. The proposed optimization problem incorporates the correlation between each superpixel-block and training dictionary. In this way, the potential of selecting more appropriate training samples is promoted. Due to the joint-sparse regularization term, the pixels within each superpixel are represented by the same training samples. Finally, the graph Laplacian-based regularization term enforces the coefficient matrix to have similar codes among the pixels of the superpixel. To sum up, unifying all these constraints in an optimization framework enhances the discriminative structure of the coefficient matrix.

Regarding the proposed SJSGR method, long execution times are observed, particularly on the PU and FL datasets. This is primarily attributed to the large number of generated superpixels resulting from the larger image sizes. Since the SJSGR method performs joint-sparse and graph regularization-based optimization within each superpixel, the computational cost increases as the number of superpixels grows. Although the WSGR and MSSGF methods require less computation time, their classification performance across all datasets remains inferior to that of the proposed method. This performance gap becomes even more pronounced under limited-training-sample conditions. The SGL method also achieves lower classification accuracies than the proposed method while requiring slightly more computation time. Overall, the experimental results demonstrate that although the proposed method requires longer execution times for datasets with large spatial dimensions, this additional computational cost is largely compensated by its higher classification accuracy results compared with the other methods.

Finally, the comparison with recent deep learning models shows that the SJSGR method provides competitive classification performance across the datasets. While several models attain slightly higher accuracies on certain datasets, the proposed method con-

sistently achieves comparable or higher results without depending on the deep feature learning. These observations demonstrate that the proposed framework maintains an effective alternative, especially when the labeled training samples are limited.

5. Conclusions

We propose a novel spectral–spatial HSIC method called SJSGR that efficiently captures both the local and geometric structure of each superpixel. Since neighboring pixels share similar spectral patterns, the representative coefficient matrix of these samples is expected to exhibit a joint-sparse structure. Considering this fact, we employ a weighted joint-sparse regularization to promote the selection of the most representative training samples by measuring the similarity between the samples within each superpixel and the dictionary. Besides being represented by common training samples, their representation coefficients are also expected to be similar due to high correlation among samples within each superpixel. To incorporate this information, we use a graph Laplacian regularization term which forces the neighboring samples to have similar coefficients. Experiments on three HSI datasets have shown that the SJSGR method is effective and comparable to state-of-the-art methods in the literature.

Although the proposed method demonstrates competitive classification performance, it has several limitations that should be acknowledged. First, the presence of noisy pixels in the HSI can prevent the formation of homogeneous superpixel regions. This will significantly affect the success of the method. Second, performing only single-scale segmentation may lead to a limited use of spatial information in the HSI, because the sizes of homogeneous regions within the image may vary. Finally, due to the ADMM-based optimization performed within each superpixel, increasing the number of superpixels negatively impacts the computational time of the method.

Future work may focus on using a superpixel segmentation algorithm with more precise local region modeling, which is expected to further improve the results. Additionally, a multiscale superpixel segmentation approach may be utilized to capture the spatial information more. Also, the computation time can be reduced by using faster optimization techniques.

Funding: This research received no external funding.

Data Availability Statement: The IP, PU and FL datasets presented in this study are publicly available at [https://purr.purdue.edu/publications/1947/1, accessed on 27 July 2026], [https://lesun.weebly.com/hyperspectral-data-set.html, accessed on 27 July 2026] and [https://www.geodoi.ac.cn/edoi.aspx?DOI=10.3974/geodb.2017.03.04.V1, accessed on 27 July 2026], respectively.

Conflicts of Interest: The author declares that there are no conflicts of interest.

References

1. Shaw, G.; Manolakis, D. Signal processing for hyperspectral image exploitation. *IEEE Signal Process. Mag.* **2002**, *19*, 12–16. [CrossRef]
2. Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral Remote Sensing Data Analysis and Future Challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36. [CrossRef]
3. Ghamisi, P.; Yokoya, N.; Li, J.; Liao, W.; Liu, S.; Plaza, J.; Rasti, B.; Plaza, A. Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 37–78. [CrossRef]
4. Manolakis, D.; Shaw, G. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Process. Mag.* **2002**, *19*, 29–43. [CrossRef]
5. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [CrossRef]
6. Benediktsson, J.; Palmason, J.; Sveinsson, J. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 480–491. [CrossRef]

7. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6232–6251. [CrossRef]
8. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral Image Classification Using Dictionary-Based Sparse Representation. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3973–3985. [CrossRef]
9. Fauvel, M.; Tarabalka, Y.; Benediktsson, J.A.; Chanussot, J.; Tilton, J.C. Advances in Spectral-Spatial Classification of Hyperspectral Images. *Proc. IEEE* **2013**, *101*, 652–675. [CrossRef]
10. Yu, H.; Gao, L.; Liao, W.; Zhang, B.; Zhuang, L.; Song, M.; Chanussot, J. Global Spatial and Local Spectral Similarity-Based Manifold Learning Group Sparse Representation for Hyperspectral Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3043–3056. [CrossRef]
11. Tu, B.; Zhou, C.; Liao, X.; Zhang, G.; Peng, Y. Spectral–Spatial Hyperspectral Classification via Structural-Kernel Collaborative Representation. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 861–865. [CrossRef]
12. Dundar, T.; Ince, T. Guided Filter of Random Patches Network and Relaxed-Collaborative-Representation-Based Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 6544–6560. [CrossRef]
13. Zhang, H.; Li, J.; Huang, Y.; Zhang, L. A Nonlocal Weighted Joint Sparse Representation Classification Method for Hyperspectral Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2056–2065. [CrossRef]
14. Tu, B.; Zhang, X.; Kang, X.; Zhang, G.; Wang, J.; Wu, J. Hyperspectral Image Classification via Fusing Correlation Coefficient and Joint Sparse Representation. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 340–344. [CrossRef]
15. Gao, Q.; Lim, S.; Jia, X. Improved Joint Sparse Models for Hyperspectral Image Classification Based on a Novel Neighbour Selection Strategy. *Remote Sens.* **2018**, *10*, 905. [CrossRef]
16. Chen, X.; Chen, N.; Peng, J.; Sun, W. Local Matrix Feature-Based Kernel Joint Sparse Representation for Hyperspectral Image Classification. *Remote Sens.* **2022**, *14*, 4363. [CrossRef]
17. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral Image Classification via Kernel Sparse Representation. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 217–231. [CrossRef]
18. Fang, L.; Li, S.; Kang, X.; Benediktsson, J.A. Spectral–Spatial Classification of Hyperspectral Images with a Superpixel-Based Discriminative Sparse Model. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4186–4201. [CrossRef]
19. Dundar, T.; Ince, T. Sparse Representation-Based Hyperspectral Image Classification Using Multiscale Superpixels and Guided Filter. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 246–250. [CrossRef]
20. Zhang, A.; Pan, Z.; Fu, H.; Sun, G.; Rong, J.; Ren, J.; Jia, X.; Yao, Y. Superpixel Nonlocal Weighting Joint Sparse Representation for Hyperspectral Image Classification. *Remote Sens.* **2022**, *14*, 2125. [CrossRef]
21. Dundar, T.; Ince, T. Locality Constraint Joint-Sparse and Weighted Low-Rank Based Hyperspectral Image Classification. In *2023 10th International Conference on Recent Advances in Air and Space Technologies (RAST)*; IEEE: New York, NY, USA, 2023; pp. 1–6.
22. Ince, T. Superpixel-Based Graph Laplacian Regularization for Sparse Hyperspectral Unmixing. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5501305. [CrossRef]
23. Liu, C.; Zhou, J.; Liang, J.; Qian, Y.; Li, H.; Gao, Y. Exploring Structural Consistency in Graph Regularized Joint Spectral-Spatial Sparse Coding for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1151–1164. [CrossRef]
24. Xue, Z.; Du, P.; Li, J.; Su, H. Sparse graph regularization for robust crop mapping using hyperspectral remotely sensed imagery with very few in situ data. *ISPRS J. Photogramm. Remote Sens.* **2017**, *124*, 1–15. [CrossRef]
25. Xue, Z.; Yang, S.; Zhang, L. Weighted Sparse Graph Regularization for Spectral–Spatial Classification of Hyperspectral Images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 1630–1634. [CrossRef]
26. Sellars, P.; Aviles-Rivero, A.I.; Schönlieb, C.B. Superpixel Contracted Graph-Based Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 4180–4193. [CrossRef]
27. Zhao, Y.; Yan, F. Hyperspectral Image Classification Based on Sparse Superpixel Graph. *Remote Sens.* **2021**, *13*, 3592. [CrossRef]
28. Yang, S.; Zhang, Y.; Jia, Y.; Zhang, W. Local Low-Rank Approximation with Superpixel-Guided Locality Preserving Graph for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 7741–7754. [CrossRef]
29. Wang, D.; Hu, M.; Jin, Y.; Miao, Y.; Yang, J.; Xu, Y.; Qin, X.; Ma, J.; Sun, L.; Li, C.; et al. HyperSIGMA: Hyperspectral Intelligence Comprehension Foundation Model. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 6427–6444. [CrossRef] [PubMed]
30. Kong, W.; Liu, B.; Bi, X.; Yu, C.; Li, X.; Chen, Y. HyperSL: A Spectral Foundation Model for Hyperspectral Image Interpretation. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5513119. [CrossRef]
31. Braham, N.A.A.; Albrecht, C.M.; Mairal, J.; Chanussot, J.; Wang, Y.; Zhu, X.X. SpectralEarth: Training Hyperspectral Foundation Models at Scale. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 16780–16797. [CrossRef]
32. Zhao, X.; Xiong, Z.; Zhu, X.X. HySens: Sensor-Agnostic Foundation Models for Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2026**, *64*, 5513915. [CrossRef]
33. Lu, S.; Guo, J.; Zimmer-Dauphinee, J.R.; Nieusma, J.M.; Wang, X.; VanValkenburgh, P.; Wernke, S.A.; Huo, Y. Vision Foundation Models in Remote Sensing: A survey. *IEEE Geosci. Remote Sens. Mag.* **2025**, *13*, 190–215. [CrossRef]

34. Vaddi, R.; Phaneendra Kumar Lakshmi Narasimha, B.; Mitra, S.; Mitra, S.; Bruzzone, L.; Roy, S.K. From Traditional to Foundation Models: A Survey for Land Use and Land Cover Hyperspectral Image Classification. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2025**, *15*, e70049. [CrossRef]
35. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122. [CrossRef]
36. Liu, M.; Tuzel, O.; Ramalingam, S.; Chellappa, R. Entropy rate superpixel segmentation. In *CVPR 2011*; IEEE: New York, NY, USA, 2011; pp. 2097–2104. [CrossRef]
37. Tu, B.; Kuang, W.; Zhao, G.; Fei, H. Hyperspectral Image Classification via Superpixel Spectral Metrics Representation. *IEEE Signal Process. Lett.* **2018**, *25*, 1520–1524. [CrossRef]
38. Xie, F.; Gao, Q.; Jin, C.; Zhao, F. Hyperspectral Image Classification Based on Superpixel Pooling Convolutional Neural Network with Transfer Learning. *Remote Sens.* **2021**, *13*, 930. [CrossRef]
39. Wang, N.; Zeng, X.; Duan, Y.; Deng, B.; Mo, Y.; Xie, Z.; Duan, P. Multi-Scale Superpixel-Guided Structural Profiles for Hyperspectral Image Classification. *Sensors* **2022**, *22*, 8502. [CrossRef] [PubMed]
40. Liao, N.; Gong, J.; Li, W.; Li, C.; Zhang, C.; Guo, B. SMALE: Hyperspectral Image Classification via Superpixels and Manifold Learning. *Remote Sens.* **2024**, *16*, 3442. [CrossRef]
41. Zhang, Y.; Duan, P.; Liang, L.; Kang, X.; Li, J.; Plaza, A. PFS3F: Probabilistic Fusion of Superpixel-Wise and Semantic-Aware Structural Features for Hyperspectral Image Classification. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 8723–8737. [CrossRef]
42. Baumgardner, M.F.; Biehl, L.L.; Landgrebe, D.A. *220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3*; Purdue University: West Lafayette, IN, USA, 2015.
43. Dell'Acqua, F.; Gamba, P.; Casella, V.; Zucca, F.; Benediktsson, J.A.; Wilkinson, G.; Galli, A.; Malinverni, E.S.; Jones, G.; Greenhill, D.; et al. HySenS data exploitation for urban land cover analysis. *Ann. Geophys.* **2006**, *49*, 311–318. [CrossRef]
44. Zhang, X.; Zhang, B.; Sun, Y. Hyperspectral Remote Sensing Dataset for Tea Farm. *Digit. J. Glob. Change Data Repos.* **2017**, *4*. [CrossRef]
45. Hong, D.; Han, Z.; Yao, J.; Gao, L.; Zhang, B.; Plaza, A.; Chanussot, J. SpectralFormer: Rethinking Hyperspectral Image Classification with Transformers. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5518615. [CrossRef]
46. He, Y.; Tu, B.; Liu, B.; Li, J.; Plaza, A. 3DSS-Mamba: 3D-Spectral-Spatial Mamba for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5534216. [CrossRef]
47. Roy, S.K.; Krishna, G.; Dubey, S.R.; Chaudhuri, B.B. HybridSN: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 277–281. [CrossRef]
48. Liu, S.; Fu, C.; Duan, Y.; Wang, X.; Luo, F. Spatial–Spectral Enhancement and Fusion Network for Hyperspectral Image Classification with Few Labeled Samples. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5502414. [CrossRef]
49. Xu, Y.; Wang, D.; Zhang, L.; Zhang, L. Dual selective fusion transformer network for hyperspectral image classification. *Neural Netw.* **2025**, *187*, 107311. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

remote sensing

MDPI

Article

Enhancing the Estimation and Mapping of Soil Cadmium by Using Geospatial Information-Guided Machine Learning and Principal Component Spectra

Jianfei Cao [1,2], Yihui Liu [3], Xinrong Duan [4], Xiaoli Liu [5], Xiukun Zhang [1], Qixin Shi [1] and Xibo Xu [1,6,*]

[1] College of Geography and Environment, Shandong Normal University, Jinan 250300, China; caojianfei@sdnu.edu.cn (J.C.)
[2] Satellite Application Center for Ecology and Environment, Ministry of Ecology and Environment, Beijing 100094, China
[3] Urban and Rural Planning Research Center, Linyi Natural Resources and Planning Bureau, Linyi 276000, China; liuyihuichn@whu.edu.cn
[4] Shandong Institute of Land Surveying and Mapping, Jinan 250102, China
[5] Shandong Provincial Territorial Spatial Ecological Restoration Center, Jinan 250014, China
[6] Key Laboratory of Environmental Change and Natural Disaster of Ministry of Education, Beijing Normal University, Beijing 100875, China
* Correspondence: xuxibo@stu.sdnu.edu.cn; Tel.: +86-077-2944-3069

Highlights

What are the main findings?

- An improved spectral estimation and mapping framework for soil cadmium (Cd) was developed using geospatial information-guided machine learning (GIGS) and principal component spectra.
- Compared to commonly used models (e.g., XGBoost, RF, SVR), the GIGS model achieved a maximum performance improvement of approximately 27.87% and a minimum improvement of approximately 16.42% (with reference to the R^2 value).

What are the implications of the main finding?

- GIGS-based spectral estimation framework provides an accurate and reliable tool for proximal or satellite estimation of soil Cd, thereby supporting cost-effective soil pollution management and environmental sustainability worldwide.

Abstract

Integrating machine learning with spectral data provides an effective and cost-efficient scheme for estimating cadmium (Cd) contents in soils compared with labor-intensive laboratory analyses. However, conventional machine learning–based spectral estimation methods often yield unsatisfactory performance because they rely on an unrealistic assumption that the functional relationships across geographically distinct subregions are homogeneous, thereby reducing the accuracy and stability of soil Cd estimation. To address this issue, a geospatial information-guided XGBoost (GIGS) model was developed, in which a spatial weighting module was incorporated into the XGBoost framework to account for spatial heterogeneity in functional relationships among sampling locations. The spatial heterogeneity of each soil spectral sample was quantified and assigned a corresponding spatial weight, which was subsequently incorporated during model calibration. The optimal principal component spectra (PCS) derived from spectral data were used as model inputs, with measured soil Cd content as the dependent variable, thereby establishing a robust spectral estimation model for soil Cd. Results indicated that the GIGS model exhibited satisfactory performance in the spectral estimation of soil Cd content, with R^2, RMSE, and

RPIQ values of 0.78, 0.04, and 2.02, respectively. Compared to commonly used spectral estimation models (e.g., XGBoost, random forest, support vector regression), the GIGS model achieved a maximum performance improvement of approximately 27.87% and a minimum improvement of approximately 16.42% (with reference to the R^2 value). Five PCS of soil spectral data were extracted as predictors for the model. PCS-1 was found to be closely associated with iron oxides, while PCS-2 primarily reflects the spectral characteristics of clay minerals. The spectral bands at 600 nm and 815 nm contribute most strongly to PCS-3, which is linked to soil organic matter and indirectly reflects the soil Cd status. PCS-4 and PCS-5 represent mixed spectral information derived from materials associated with soil Cd. An integrated framework combining the GIGS model and PCS data developed in this study provides an accurate and reliable tool for spectral estimation and mapping of soil Cd, thereby supporting cost-effective soil management and environmental sustainability worldwide.

Keywords: soil potentially toxic elements; spectral estimation; digital mapping; spatial heterogeneity; machine learning

1. Introduction

Soil is a vital component of the Earth's land ecosystem and provides the foundation for human life and production activities [1]. Approximately 95% of the food humans consume comes from soil, while only 5% comes from aquatic sources and other sources [2]. With intensive industrial and agricultural activities, large amounts of pollutants are released into the soil environment. Potentially toxic elements (PTEs) are naturally occurring chemical substances in soils. When combined with large external inputs, they have become some of the most significant pollutants [3–5]. Hou et al. [6] reported that PTE concentrations exceed risk thresholds on about 14% to 17% of arable land worldwide. According to the Food and Agriculture Organization of the United Nations (FAO), by 2050, nearly 90% of global soil resources may be at risk of degradation due to soil erosion, improper fertilizer and pesticide use, and industrial pollution [7].

Cadmium (Cd) is one of the most common PTEs found in soils, with about 19% of soils exceeding the soil quality standard according to a nationwide soil survey in China [8,9]. Cd in soil comes from both natural and anthropogenic sources. Cd is naturally present in bedrock, which acts as the parent material for soil formation [10]. During the processes of bedrock weathering and soil development, Cd gradually moves into the soil through various transformation mechanisms. Anthropogenic sources of Cd in soil mainly include agricultural, domestic, and industrial activities [11]. Facilities and products produced by human activities often contain higher levels of Cd, which can greatly increase the possibility of Cd accumulation in the soil environment [12]. Cd is non-essential and harmful to the human body; exposure can easily cause cancer, genetic mutations, and other health hazards [13]. The accumulation of Cd in soil is considered a significant threat to ecosystems, food safety, and public health [14,15]. Therefore, establishing an accurate and efficient method to detect Cd content in soil would support health risk assessments and pollution control efforts [16].

Proximal or satellite spectra have shown great potential for rapid, efficient, and non-destructive monitoring of soil Cd. The traditional approach for monitoring soil Cd involves soil sampling combined with laboratory chemical analysis [3]. This method is simple and requires low technical expertise; however, it is costly, time-consuming, and inefficient, making it unsuitable for large-scale or repeated monitoring of soil Cd status [17].

Laboratory analyses, such as atomic absorption spectroscopy and inductively coupled plasma mass spectrometry, also generate experimental waste that may cause environmental pollution if not properly managed [18]. In contrast, proximal or satellite spectral sensors can detect soil properties and subtle variations using hundreds of continuous, narrow spectral bands, making it an attractive method for analyzing soil components [3,19]. When combined with data-driven machine learning algorithms, spectral models can effectively extract soil property information and accurately estimate the soil Cd [20,21].

Establishing a functional relationship between the soil Cd contents and spectral data using data-driven machine learning methods is a key step in spectral estimation modeling of soil Cd [22,23]. Data-driven machine learning methods are effective for identifying nonlinear relationships within spectral data and establishing spectral estimation models for soil Cd [24–26]. Compared to traditional spectral estimation models based on empirical-based linear relationships, data-driven machine learning models offer higher accuracy and reliability in the spectral estimation of Cd contents in the soil [27]. The spectral features for the soil Cd are selected from spectral data as independent variables, while the measured Cd contents in soil are used as the dependent variable. The spectral estimation model of soil Cd was constructed after model training, parameter optimization, and validation.

The key assumption for the development of data-driven spectral estimation models of soil Cd is that the soil environment within a target geographic region is homogeneous. Specifically, this assumption presumes that the relationship coefficients between soil Cd and spectral data are consistent across different locations within the target geographical area, thereby allowing data-driven spectral estimation models to effectively identify functional relationships within the spectral data [28–30].

Although such an assumption simplifies the identification of empirical-based relationships from the spectral data and shortens the training time of the model, ignoring the spatial heterogeneity may lead to a spectral estimation model of soil Cd with poor accuracy and low generalization. Under the combined influence of human activities (such as agricultural and industrial activities) and natural environmental factors (winds and rainfall), soil ecosystems within a region exhibit significant spatial heterogeneity. This soil heterogeneity is reflected in differences in soil structure, composition, and habitat types; the influence of soil Cd contamination on spectral data varies across different geographic locations [30,31].

However, the assumption that spectral relationships remain unchanged and that geographical regions are homogeneous in spectral modeling contradicts the actual conditions of regional soil environments [29,32]. As a result, data-driven spectral estimation models often underestimate high soil Cd contamination, which reduces the models' accuracy and reliability.

Therefore, this study proposes a geospatial information-guided machine learning (GIGS) framework to account for spatial heterogeneity in functional relationships among sampling locations, and to enhance the spectral estimation of soil Cd. Firstly, a spatial weighting module was incorporated into the XGBoost model to analyze the spatial heterogeneity of soil spectral samples and assign a spatial weight to each sample during model calibration. This approach eliminates the influence of spatial heterogeneity on the fitting of empirical relationships in spectral modeling. Furthermore, the GIGS model is able to integrate results from both local and global modeling modes, which contributes to further improving the spectral estimation of soil Cd.

This study aims to: (1) develop a GIGS model for estimating soil Cd content; (2) derive principal component spectral (PCS) data to characterize soil Cd; and (3) evaluate the performance of the GIGS model.

2. Materials and Methods

2.1. Study Areas

The study area was a sewage-irrigated region located on the coastal plain north of Longkou City, Shandong Province, China (Figure 1). Its geographical coordinates are (37°34′35″N~37°44′12″N, 120°20′56″E~120°40′47″E), which extends for totally 527.85 km^2. It belongs to the temperate monsoon climate; the monsoon is more obvious, the four seasons are distinct, the average annual temperature is about 12 °C, and the annual precipitation is about 600 mm~700 mm. The soil types in the study area cover tidal soil, brown soil, and mortar soil, with pH ranging from 6.5 to 7.5 [33]. The study area is located in the Longkou Coal mine area, which is rich in coal resources. Coal mining waste water and coal gangue waste residue caused a serious threat to the soil environment [3]. The residue and waste liquid produced by metal smelters and thermal power plants also aggravated the enrichment of heavy metals in the soil [34,35].

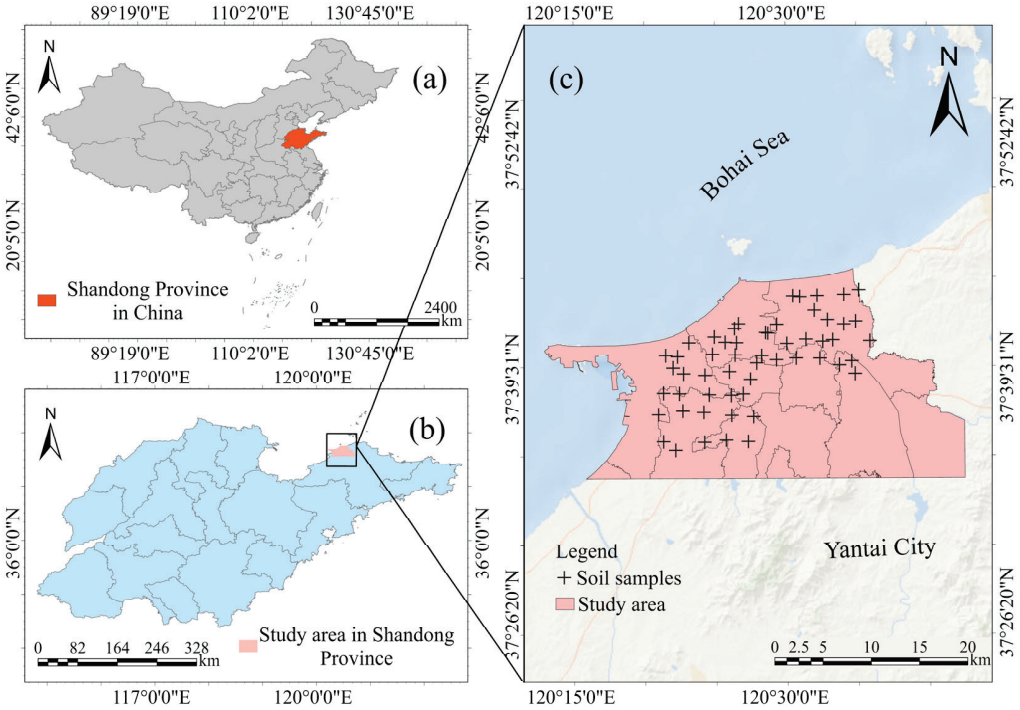


Figure 1. Geographical Locations of the study area and soil samplings: (**a**) Shandong Province in China; (**b**) Study area in Shandong Province; (**c**) Details of the study area and soil sampling.

2.2. Data Collection

2.2.1. Soil Sample Collection

After considering the soil types, land use, and sewage irrigation conditions in the study area, the sampling locations (n = 67) were predefined on a digital map using ArcGIS 10.8 software (Esri Inc., Redlands, CA, USA) (Figure 1). At each sampling site, the five-point sampling method was used to collect approximately 1 kg of soil from a depth of 0 to 20 cm. Soil samples were collected from five sites using a shovel. The collected samples were thoroughly mixed to form a composite sample, which was then placed in a clean self-sealing bag. The latitude, longitude, and elevation of each sampling point were determined using a handheld GPS. Additionally, the environmental conditions such as land use and temperature of the sampling locations were recorded with a smart camera. The field survey and sample collection were completed in June 2022.

https://doi.org/10.3390/rs18142341

2.2.2. Soil Spectral Measurement and Chemical Analysis

Soil pretreatment is crucial for the laboratory chemical analysis of soil samples. In the laboratory, the soils were air-dried under natural conditions, and gravel, as well as animal and plant residues, were removed. The soils were then ground and passed through a 100-mesh nylon sieve. Afterward, the soil samples were divided into two parts: one for chemical analysis and the other for spectral measurement.

The ASD FieldSpec-4 portable spectrometer (ASD Inc., Boulder, CO, USA) was used to measure the spectral reflectance of soil samples. The ASD FieldSpec-4 covers a spectral range of 350–2500 nm, with spectral resolutions of 1.4 nm (350–1100 nm) and 2 nm (1000–2500 nm). Soil spectral measurements were conducted in an adjustable light-controlled darkroom. Two 50 W halogen lamps, positioned 30 cm above the sample, were used as the light source, with a lighting angle of 15° and a probe distance of 20 cm.

Before measurement, the spectrometer was preheated for 30 min, and whiteboard calibration was performed to minimize the impact of instrument errors on data acquisition. Whiteboard calibration was repeated several times during the measurement process to reduce noise effects. To avoid random errors, each soil sample was measured five times, and the spectral ranges with a low signal-to-noise ratio were removed [3]. Abnormal spectra were identified, and the average of the remaining spectra was taken as the spectral data for the soil sample.

The inductively coupled plasma mass spectrometer (ICP-MS) method was used to determine the Cd contents in the soil samples [36]. Blank samples and standard soil samples (GBW 07449) were incorporated into the laboratory experiments for soil sample analysis [37]. Ten percent of the samples were randomly selected as parallel samples, with the relative standard deviation (RSD) of their measurements being less than 5%. The recovery rates of the standard samples ranged from 95% to 105%, demonstrating the reliability of the laboratory analysis for Cd in soil samples.

2.3. Principal Component Spectral Data

The direct spectral signatures of soil Cd are difficult to capture due to its low concentration and environmental interference. Therefore, spectrally active substances associated with Cd, such as soil organic matter, iron oxides, and clay minerals, are commonly used as proxies for soil Cd content. In this study, principal component transformation was applied to hyperspectral data for dimensionality reduction and feature extraction. The resulting principal component spectral (PCS) data, which represent spectral information related to Cd-associated materials, were used as input variables for spectral modeling, thereby reducing model convergence time and improving model robustness.

2.3.1. First-Order Derivative

Spectral transformation aimed to highlight the spectral signals of soil Cd and eliminate the noise information generated by environmental factors and instrument errors [38,39]. In this study, the Savitzky–Golay (SG) and First-order derivative (FOD) were used to transform the soil spectra data [40,41]. The SG transformation is a technique used to reduce random noise in soil spectral data. The SG transformation applies a *k*th-order polynomial fitting within a specified window width to the data. The expression for the SG transformation is as follows:

$$R_i = \frac{h_n}{H} \sum_{n=-w}^{+w} R_{k+n} \quad (1)$$

where R_i represents the ith spectral band, w is the filter window width, H is the normalization coefficient, and h_{n} is the smoothing coefficient, which is obtained by least-squares polynomial fitting.

https://doi.org/10.3390/rs18142341

First-order derivative (FOD) transformation is an effective technique to enhance features in soil spectra, including inflection points of the spectral curve and positions of reflectance extrema [42]. These features are associated with the Cd content in the soil. The FOD can be expressed as:

$$\mathrm{FOD}(\lambda_i) = \frac{R(\lambda_{i+1}) - \mathrm{R}(\lambda_{i-1})}{2\Delta\lambda} \tag{2}$$

where $R(\lambda_{\mathrm{i+2}})$, $R(\lambda_{\mathrm{i+1}})$, and $R(\lambda_{\mathrm{i-2}})$ indicate the reflectance values at the spectral bands of the (i + 2) th, i th, and ($i-2$) th, respectively. $\Delta\lambda$ represents the wavelength interval between the i th spectral band and the (i + 1) th spectral band.

2.3.2. Principal Component Transformation

Principal component transformation is a dimensionality reduction method that can extract the effective information of soil spectral data through information compression by using the orthogonal transformation [43,44]. Soil spectral data included a large number of spectral bands, and collinearity and redundancy existed. Principal component analysis can effectively reduce the data dimension, extract key information in the soil spectral data, and support the GIGS model construction. Principal component transformation was expressed as follows:

$$Z = \frac{X - \overline{X}}{d} \tag{3}$$

where X is the spectral reflectance, $\overline{X}$ is the average of the spectral reflectance, and d represents the standard deviation.

$$\mathrm{cov}(x_i, y_j) = \frac{\sum_{k=1}^{n} (x_{ki} - x_i)(x_{kj} - x_j)}{n - 1} \tag{4}$$

where k indicates the k-th soil sample, i and j denote the i-th and j-th spectral bands, respectively, X_{ki} is the reflectance of the i-th band of the k-th soil sample, and n is the number of soil samples.

$$\mathrm{c} = \frac{\lambda_k}{\sum_{i=1}^{p} \lambda_i} \tag{5}$$

where c represents the cumulative variance contribution rate of the principal component spectrum, λ represents the eigenvector, and i is the number of eigenvectors. Principal component calculations were performed using the Python software (Version 3.14).

2.4. GIGS Model

The GIGS model is an improved spectral estimation model for soil Cd that was developed using the XGBoost modeling framework (Version 2.1.0) and offers advantages in soil properties estimation in geographic environments with significant spatial heterogeneity [45,46]. The GIGS model integrates the results of two sub-models under global and local modes, and incorporates a spatial weighting module, thereby improving the robustness of soil Cd spectral estimation [47,48]. The framework for the GIGS model is illustrated in Figure 2.

We assume that a soil spectral dataset consisting of n samples was collected, where y_i denotes the target variable and x_i represents the explanatory variable at location i. The GIGS model under the global mode can be expressed as follows:

$$\hat{\mathrm{y}}_{\mathrm{g}} = \sum_{K=1}^{K} \mathrm{f}_k(\mathrm{x}_i), \mathrm{f}_k \in F \tag{6}$$

https://doi.org/10.3390/rs18142341

where $\hat{y}_g$ indicates the estimated value at location i, K is the number of decision trees, and f_k is a function belonging to the function space F; F represents the set of all possible decision trees.

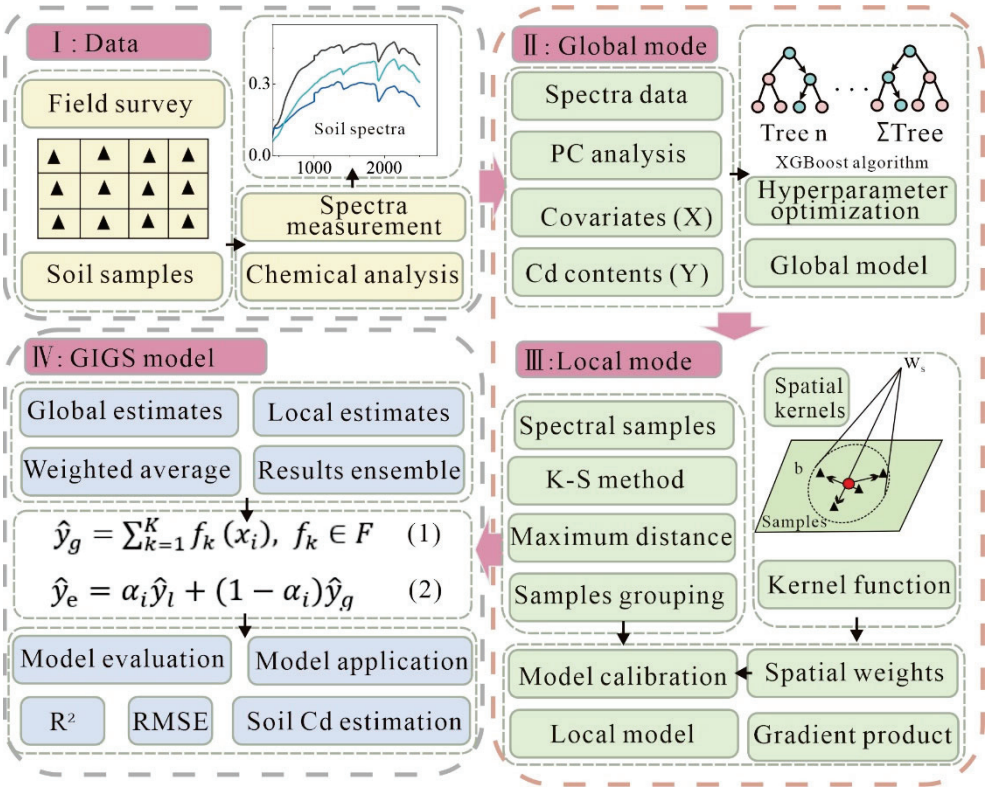


Figure 2. Flowchart for the construction of the geospatial information-guided spectral (GIGS) model.

The GIGS model in the local mode consists of three steps. First, the distances between all spectral samples are calculated using the Kennard–Stone algorithm, and two sample subsets are partitioned based on the principle of maximizing spectral distance [49]. Subsequently, a neighborhood is constructed within each sample subset. The Gaussian function is defined as the kernel function, and spatial kernels are used to calculate the spatial weights (ws) for each sample subset [47].

$$ws_{ij} = \exp\left(-\frac{d_{ij}^2}{b^2}\right) \tag{7}$$

where d_{ij} represents the Euclidean distance between the i th and j th samples with different geographic locations, and b is the threshold defined for the distance.

Spatial weights function in the GIGS model are used to assign the spatial weight for the spectral samples. By adopting a gradient product approach, the conventional weights obtained from the XGBoost algorithm are replaced with spatial weights during the model calibration [50]. The spatial weights are multiplied by the first- and second-order gradients, assigning a weight to the spectral samples. In addition, the GIGS model can also be calibrated by the soil spectral samples without the spatial weights.

In the end, the estimation results from the local mode and global mode in the GIGS model were integrated by using a weighted averaging approach, the ensemble model as shown in the following equation:

$$\widehat{y}_e = \begin{bmatrix} \widehat{y}_g & \alpha_i = 0 \\ \alpha_i \widehat{y}_l + (1-\alpha_i)\widehat{y}_g & \alpha_i \in (0,1) \\ \widehat{y}_l & \alpha_i = 1 \end{bmatrix} \quad (8)$$

where $\widehat{y}_e$ indicates the estimated value of soil Cd from the GIGS model at location i. α is a hyperparameter of the GIGS model, related to the spatial heterogeneity of the soil spectral samples, and is used to guide the model. The α value ranges from 0 to 1 and is optimized by the residuals of the global ($\widehat{y}_g$) and local sub-models ($\widehat{y}_l$). The estimation of the ensemble model is the weighted average of the prediction results from the local and global sub-models at location i. The case where α equals zero is a special scenario in which the estimation at location i is computed solely by the global mode of the GIGS model, without considering the local mode.

During the calibration of the GIGS model, the hyperparameters in the objective function, loss function, and regularization term for each decision tree are optimized by grid search and cross-validation, thereby improving the accuracy of the GIGS model and avoiding overfitting [46,50,51]. The GIGS model was implemented using Python software (Version 3.14).

2.5. Model Calibration and Evaluation

Model performance was evaluated using three indices: coefficient of determination (R^2), root mean square error (RMSE), and ratio of performance to quartiles (RPIQ) [52,53]. A well-performing model is characterized by lower RMSE values and higher R^2 and RPIQ values [54]. The collected soil samples (n = 67) were divided into the calibration set (n = 50) and the validation set (n = 17) based on the Kennard–Stone algorithm. The calibration set was used to train the GIGS model, while the validation set was used for independent evaluation of model performance and did not participate in model calibration.

In addition, several commonly used machine learning algorithms were applied to construct spectral estimation models of soil Cd, including extreme gradient boosting (XGBoost), random forest (RF), and support vector regression (SVR) [55,56]. Five-fold cross-validation and grid search were used to tune the hyperparameters to achieve the best performance of the above models. The accuracy of these algorithms was evaluated to provide a better assessment of the GIGS model's performance.

3. Result and Analysis

3.1. Descriptive Statistics Characteristics of the Cd Contents in the Soil Samples

The descriptive statistical characteristics of Cd contents in soils of sewage irrigation areas are shown in Figure 3. The soil Cd contents range from 0.14 to 0.87 mg kg^{-1}, with a mean of 0.33 mg kg^{-1} and a standard deviation of 0.12 mg kg^{-1}. The coefficient of variation is 35.65%. This coefficient, which is a dimensionless measure of standard deviation, reflects the variability of Cd concentrations among soil samples [10,57]. The coefficient of variation is less than 36% and higher than 15%, indicating that Cd exhibits moderate spatial heterogeneity in the study area.

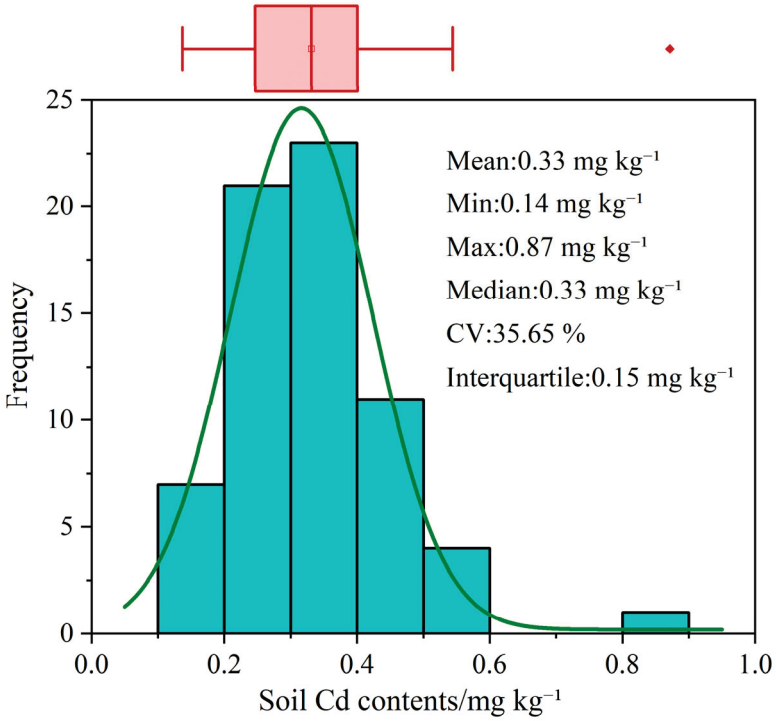


Figure 3. Descriptive statistical characteristics of the Cd contents in the soil samples in the sewage irrigation areas, China/(mg kg^{-1}).

According to the "Agricultural Land Pollution Control Standard in China" (GB 15618-2018) [58,59], the threshold for Cd pollution is 0.30 mg kg^{-1}. The mean Cd content (0.331 mg kg^{-1}) in the sewage irrigation areas exceeds this threshold, indicating a notable level of Cd pollution in the study area. One of the main sources of soil Cd is the untreated wastewater, exhaust gas, and solid waste discharged from industrial and coal mining activities. These pollutants are transported to rivers and farmlands through hydrological processes and atmospheric transport, leading to soil Cd pollution. A comprehensive analysis of the statistical characteristics of soil Cd not only provides a basis for the construction and optimization of the GIGS model, but also supports agricultural practices and precise environmental management.

3.2. Soil Spectral Analysis

The spectral reflectance of soil samples with different Cd contents shows clear differences in the Vis-NIR region (Figure 4a). Overall, the spectral curves (in the wavelength range of 400–2500 nm) of all soil samples increase, but the rate of increase differs across spectral ranges. Between 500–1000 nm, reflectance rises rapidly, while from 1000–2500 nm, it increases more slowly. Distinct absorption features around 1400 nm and 1900 nm are mainly related to soil moisture. The absorption feature near 2200 nm, as noted by [60,61], is mainly associated with the clay content in the soil. Figure 4b shows the average spectral curves with the FOD transformation. After FOD transformation, the spectral features of soils become more pronounced, local characteristics are enhanced, and small spectral changes are amplified. In addition, Figure 4b also reveals a negative correlation between soil spectral reflectance and Cd content. The spectral curves of soils with different Cd contents are more clearly separated, laying a solid foundation for building accurate spectral models to estimate soil Cd content.

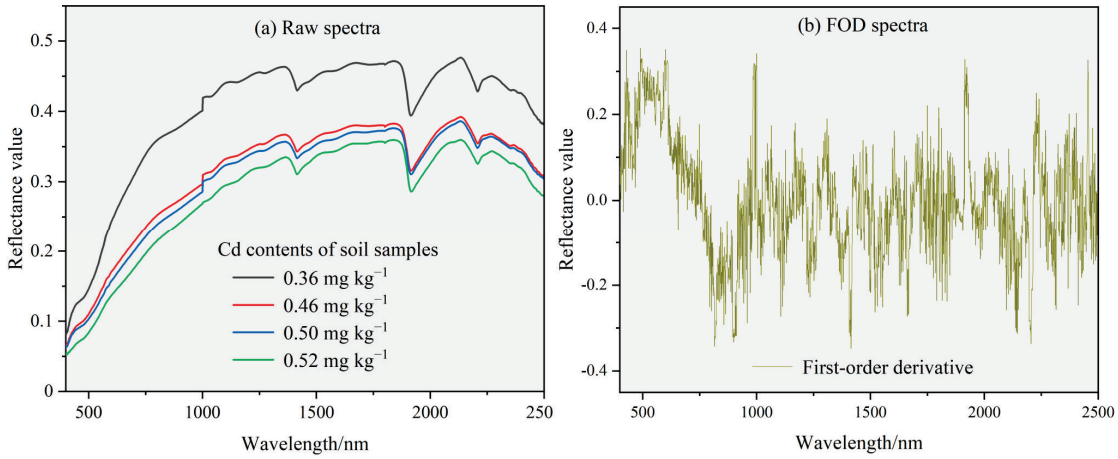


Figure 4. Spectral curves for soil with different Cd contents and the first-order derivative (FOD) spectra data.

3.3. Selecting the PCS from the Spectral Data as Covariates

PCS analysis was performed on the soil spectral data to extract the key information and remove redundant spectral data. The first ten principal components and their eigenvalues are presented in Figure 5. PCS-1 explained 45.85% of the total variance in the spectral data, while PCS-2 and PCS-3 accounted for 23.36% and 10.56% of the total variance, respectively. In addition, to improve the computational efficiency of the spectral model and prevent overfitting, the five PCS were selected as covariates for spectral modeling.

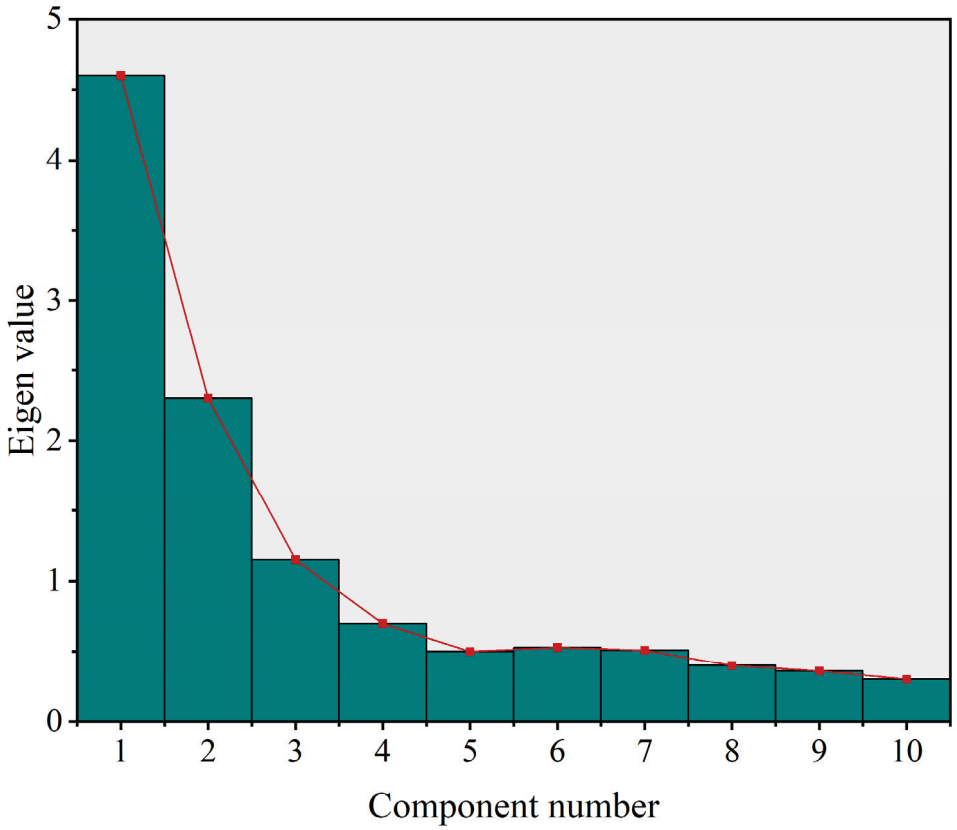


Figure 5. Key information and the principal component spectra (PCS) extracted from the soil spectra data.

3.4. Model Comparison and Evaluation

Five PCS were used as input variables, and the measured Cd content in the soil samples was taken as the dependent variable. Soil Cd spectral estimation models were

developed using GIGS, XGBoost, RF, and SVR algorithms. The results of these models are presented in Figure 6. Detailed accuracy metrics can be found in Table S1. XGBoost, RF, and SVR models exhibited moderate performance, with RPIQ values of 1.35, 1.01, and 1.15, and R^2 values of 0.67, 0.61, and 0.63, respectively. In contrast, the GIGS model achieved satisfactory performance, with the precision indices of R^2, RMSE, and RPIQ of 0.78, 0.04, and 2.02, respectively. Compared to the other models, the performance of the GIGS model improved by 16.42%, 27.87%, and 23.81% (with reference to the R^2 value), respectively.

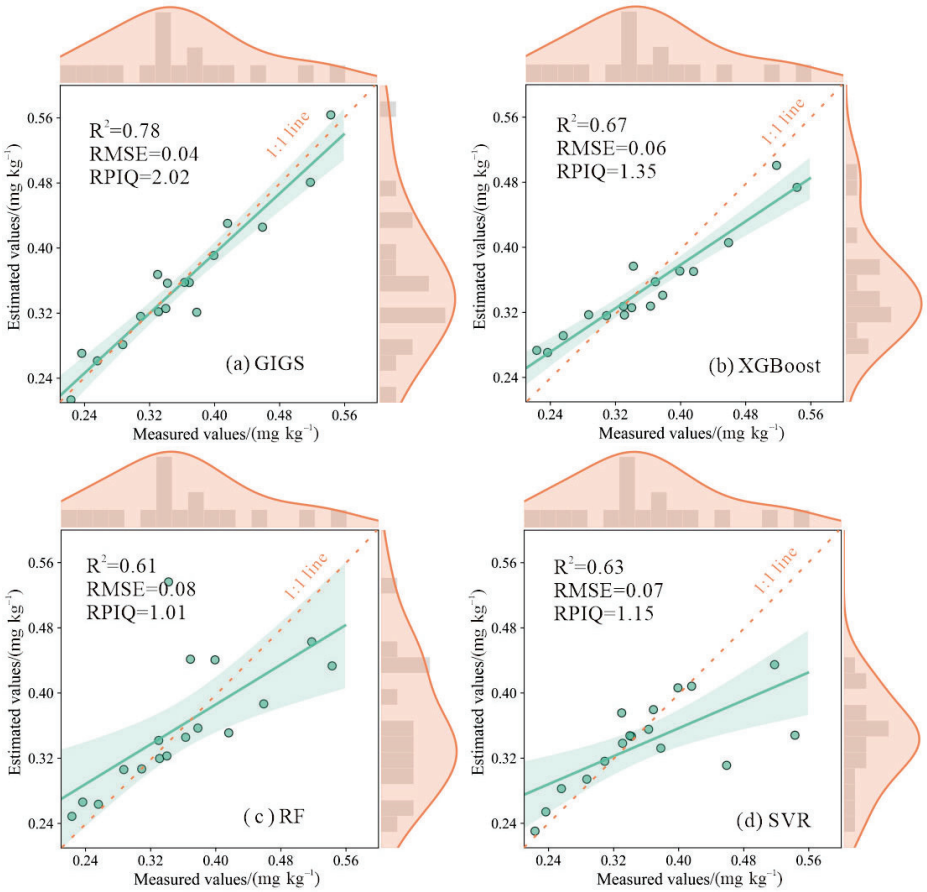


Figure 6. Scatter plots of measured versus estimated Cd contents in soils, and the accuracy indices for the geospatial information-guided spectral (GIGS), extreme gradient boosting (XGBoost), random forest (RF), and support vector regression (SVR) models.

4. Discussion

4.1. Mechanism Exploration of the Principal Component Spectral Data of Soil Cd

In this study, a spatial weighting module was added to the GIGS model to analyze the spatial heterogeneity of soil spectral samples. Each sample was given a spatial weight during model calibration. The estimation function captured by the GIGS model effectively explains the spatial heterogeneity of the spectral samples. Moreover, the GIGS model combines both local and global models for the spectral estimation of soil Cd, which improves the accuracy of spectroscopic estimation in sewage irrigation areas.

The information from spectrally active bands related to soil Cd was effectively interpreted by the extracted PCS. The loading values of these bands on the PCS are shown in Table 1. Scores of spectral samples on the principal components in the projection space are indicated in Figure 7. The absolute loading values at 429 nm, 490 nm, and 1000 nm were significant on PCS-1, indicating that PCS-1 captures most of the information for these spectral bands. The 490 nm and 1000 nm bands are absorption peaks that have been reported to be closely related to soil iron oxides [62]. Therefore, PCS-1 can be regarded

as an indirect spectral indicator of soil Cd associated with iron oxides [63]. The loading values at 900 nm, 1415 nm, 1915 nm, and 2205 nm were higher than those of other spectrally active bands on PCS-2, and the spectral information contained in these bands showed the strongest correlation with PCS-2. The bands at 1415 nm and 1915 nm correspond to the combined spectral features of moisture-related (-OH) stretching vibrations and Al-OH bending vibrations in soil clay minerals [22,60,64]. The 2205 nm band is an absorption peak of soil clay minerals, and the peak near 900 nm is caused by Fe^{3+} in the soil, indicating that PCS-2 mainly reflects the information of soil Fe^{3+}, moisture, and clay minerals [65]. The spectral information of PCS-3 was mainly derived from the bands at 600 nm and 815 nm. The 600 nm band is a typical reflection peak for soil organic matter, and the region near 815 nm is a sensitive band for organic matter response [66,67]. Thus, the indirect spectral indicators for soil Cd in PCS-3 represent the spectral information of soil organic matter [31,68]. Overall, the extracted PCS represent the comprehensive spectral information of soil Cd and indirectly characterize Cd in sewage-irrigated soils through indicators such as organic matter, clay, and iron oxides.

Table 1. Loading values of important spectral bands on the top three principal component spectra (PCS).

Spectrally Active Bands	PCS-1	PCS-2	PCS-3
429 nm	−0.711	−0.448	0.361
490 nm	0.758	−0.492	0.201
600 nm	0.661	−0.361	0.792
815 nm	0.410	0.448	0.867
900 nm	0.431	0.766	0.418
1000 nm	−0.937	−0.041	−0.010
1415 nm	0.494	0.778	−0.140
1915 nm	0.341	−0.843	−0.137
2205 nm	−0.074	0.812	0.203
2455 nm	0.694	0.700	−0.103

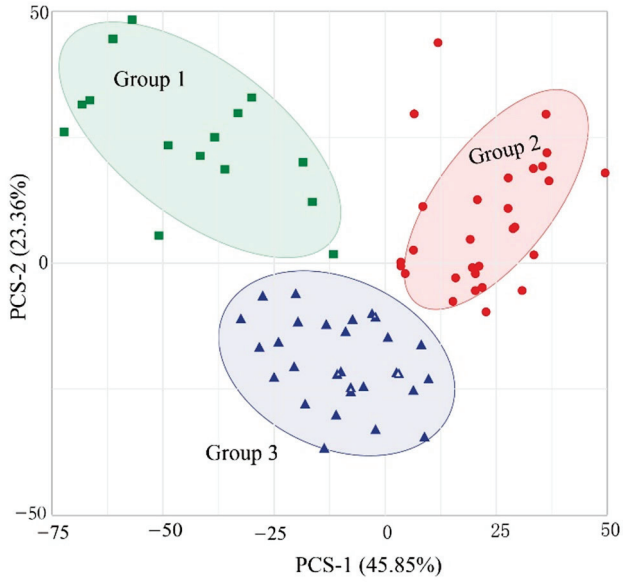


Figure 7. Scores of spectral samples on the principal components in the projection space.

4.2. Performance and Error Analysis of the GIGS Model

The sample set division is important for calibrating the GIGS model using both local and global modeling approaches. Moosavi et al. [69] and Shi et al. [70] found that local

models can perform better than global models in estimating soil properties from spectra, because they take into account the spatial heterogeneity of soil spectral samples. Choosing a suitable sample set is key to achieving the best performance of the local model in the GIGS model [48,71,72]. In this study, the Kennard–Stone algorithm was used to divide the calibration sample set into two equally sized subsets. Although this was the best choice for the current study, if the total number of samples, spatial heterogeneity, or soil properties change, the way the sample set is divided should be reassessed.

The GIGS model can recognize the spatial heterogeneity in soil spectral samples, but it does not model the spatial autocorrelation of samples within specific geographic environments. Due to spatial heterogeneity, the functional relationship and those weight coefficients between spectral data and soil Cd content vary across different locations [29,30,32]. The GIGS model addresses this negative effect from the spatial heterogeneity by using spatial weighting in the modeling process. Furthermore, the GIGS model did not achieve perfect accuracy (with R^2 and RPIQ values of 0.78 and 2.02, respectively). This shows that the spatial relationships or dependencies among soil spectral samples within a given region are not fully represented, which impacts the performance of the model. At present, the soil environment in sewage irrigation areas is complex and highly variable [73]. The GIGS model offers a way to reduce the negative influence of spatial heterogeneity on soil spectral estimation.

Both conventional machine learning models and the GIGS-based spectral estimation model were run for 50 iterations, with the resulting standard deviations of soil Cd estimates serving as indicators of model accuracy and uncertainty. The distribution of standard deviations across all soil samples is illustrated in Figure 8. Among the four models evaluated, GIGS (Figure 8a) achieved the lowest mean standard deviation (0.045), indicating the most stable and consistent estimation performance. XGBoost (Figure 8b) and RF (Figure 8c) ranked second and third, with mean standard deviations of 0.059 and 0.065, respectively, while SVR (Figure 8d) exhibited the largest variability, with a mean standard deviation of 0.071. To further investigate spatial performance, the study area was partitioned into four subregions (E, W, N, and S) based on differences in soil environmental conditions and anthropogenic disturbance. Subregion E displayed the highest coefficient of variation in soil Cd content (44.72%), substantially exceeding those of subregions W (36.13%), N (35.57%), and S (33.17%), thus reflecting significant spatial heterogeneity. GIGS maintains a relatively low standard deviation (0.055) in subregion E, while all three conventional machine learning models exhibit values exceeding 0.06. These findings highlight the superior performance of the GIGS model for spectral estimation of soil Cd, particularly in areas characterized by strong spatial heterogeneity.

4.3. Transferability and Limitations Analysis

To evaluate the transferability and robustness of the GIGS model, two case study areas, Chanqing County (CQC), China, and Thessaly and Central Greece (TCG), Greece (see Figure 9), were selected for validation. The CQC is located in eastern China, with geographic coordinates ranging from 36°14′ to 36°41′N and 116°30′ to 117°4′E, covering an area of approximately 1180 km^2. It belongs to a warm temperate continental monsoon climate zone. The region lies on the alluvial plain of the lower Yellow River and is characterized by three main soil types: brown soil, loam, and alluvial soil. The TCG is situated in the central plain of Greece, covering an area of about 1315 km^2. Its geographic coordinates range from 39°23′ to 39°45′N and 22°10′ to 22°47′E. The region has a typical Mediterranean climate. The main soil types include brown soil, terra rossa, and vertisols. The dominant crops in this area are wheat, lentils, and cotton.

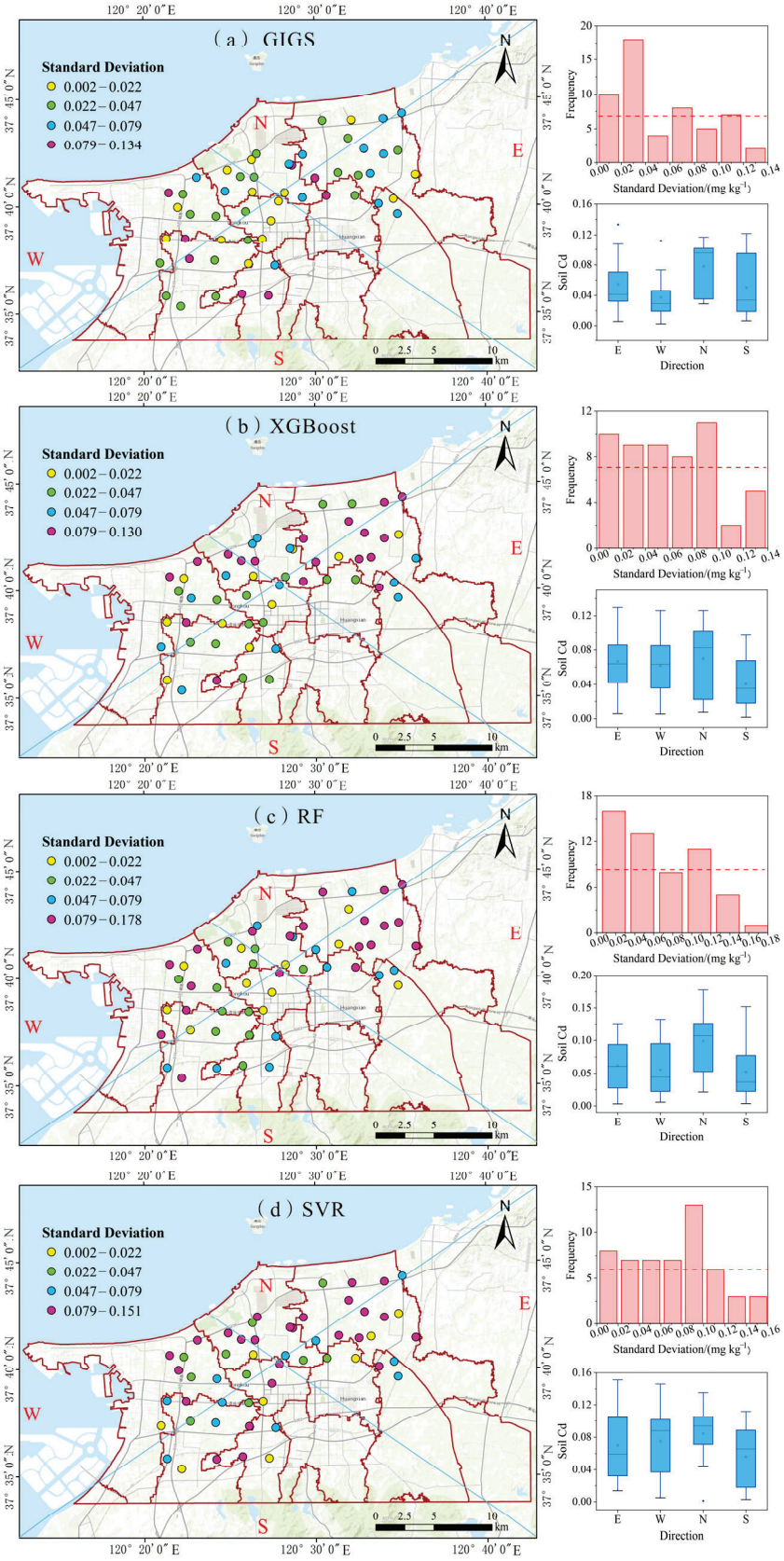


Figure 8. Standard deviation of soil Cd spectral estimation using the GIGS model in a sewage irrigation area of China: The solid blue lines represent the boundaries of the four subregions within the study area; the solid red lines indicate the boundaries of the towns within the study area; and the dashed red lines represent the mean frequency of standard deviations across different ranges.

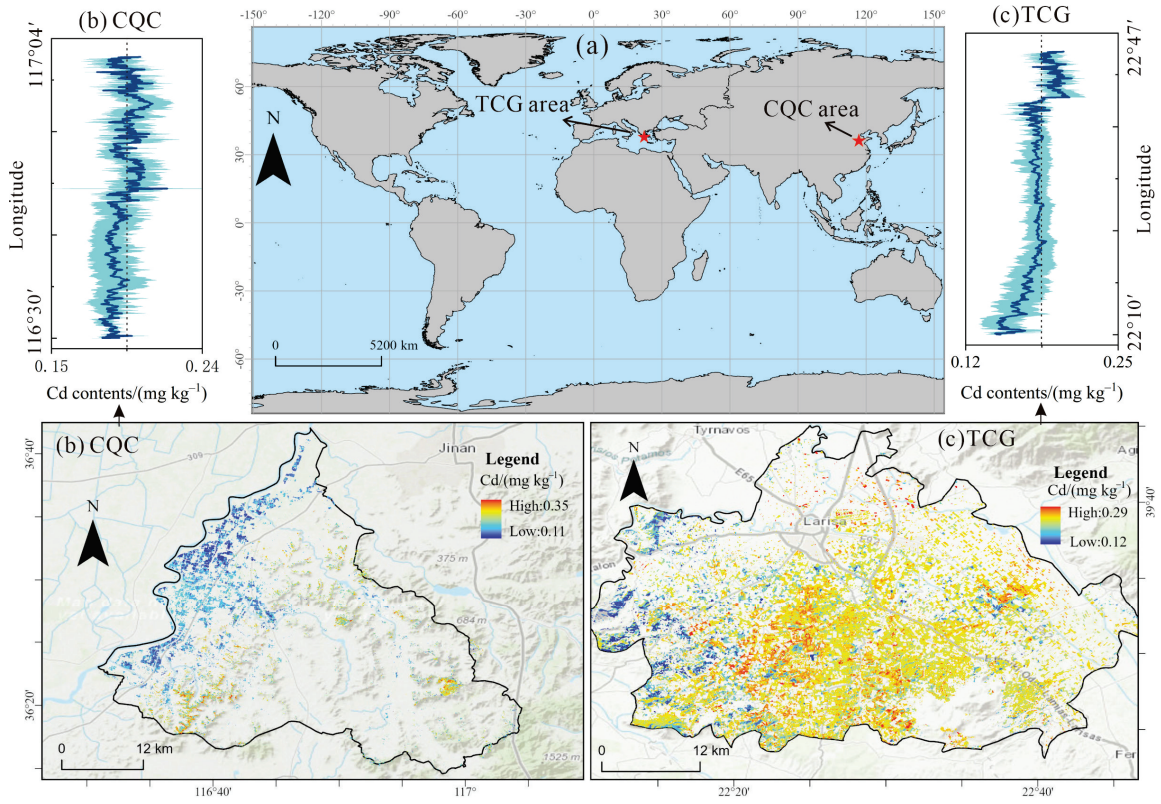


Figure 9. Transferability of the GIGS model in two case study areas, Chanqing County (CQC), China, and Thessaly and Central Greece (TCG), Greece: (**a**) The geographic locations of the CQC and TCG areas in the world; (**b**) Estimated map of soil Cd in the CQC area; (**c**) Estimated map of soil Cd in the TCG area.

The soil samples and corresponding spectral data for both the CQC and TCG areas were obtained from publicly available datasets [53,74]. A total of 304 soil samples were collected in the CQC area. Near-synchronous Landsat 8 OLI imagery (22 April 2020) was acquired from the Google Earth Engine platform to derive PCS data corresponding to the soil samples. Detailed information about the dataset can be found in Xu et al. [53]. Similarly, soil samples for the TCG area were obtained from Tóth et al. [74], along with near-synchronous Landsat 8 OLI imagery (26 July 2013). The preprocessing of soil samples and spectral data is described in Section 2, which supports a comprehensive and reliable evaluation of the GIGS model framework.

The estimated soil Cd and its spatial distribution are shown in Figure 9. The GIGS model achieved acceptable accuracy in both the CQC area and TCG area for spectral estimation of soil Cd. The R^2 and RPIQ values were 0.59 and 1.00 for the CQC area, and 0.64 and 1.17 for the TCG area. As shown in Figure 9b, areas with high soil Cd content are mainly distributed in the eastern part of the CQC area, which may be related to the parent material of the soil [10,16]. Figure 9b also indicates that hotspots of soil Cd contamination in the TCG area are mainly located in the eastern region, which are associated with agricultural activities and parent material [35,74]. These results provide evidence for the reliability and transferability of the GIGS model.

The GIGS model is combined with Sentinel-2 MSI imagery to generate a distribution map of soil Cd in the study area (Figure 1). The Cd content in the soils in the northern part of the study area is higher than that in the southern part, which is related to industrial activities and coal mining (as indicated in Figure S1). The mapping results show that soil Cd contents range from 0.12 mg kg^{-1} to 0.79 mg kg^{-1}. This range is basically consistent with the statistical results from ground-based soil samplings (0.14 mg kg^{-1} to 0.87 mg kg^{-1}).

These findings confirm that the GIGS model framework is effective for proximal estimation of soil Cd and for satellite-based mapping.

Selecting strong spectral features as input variables can improve the performance of the GIGS model. Advanced supervised search methods such as the genetic algorithm and successive projection algorithm have been used to select spectral features from the numerous bands in soil spectral data [75–77]. The selected spectral features are closely related to soil components, which can greatly improve the accuracy of soil spectral estimation models. Although the PCS analysis method used in this study helps extract key spectral information and reduce data redundancy and collinearity, further research is needed to achieve more precise spectral feature extraction and to develop robust spectral feature sets.

Future research should focus on developing new theoretical frameworks for the spectral estimation and monitoring of soil Cd. The GIGS method proposed in this study improves estimation accuracy by accounting for the spatial heterogeneity of soil spectral samples and reducing spatial non-stationarity between spectral data and Cd content. This study represents a meaningful improvement over existing theoretical frameworks for soil Cd spectral estimation [20,78]. Building on this work, new theoretical frameworks should be developed in the context of rapid advances in AI technologies to enable the analysis of spatiotemporal dynamics, driving factors, and future trends of soil Cd [79].

We acknowledge that the GIGS approach has several limitations that should be carefully considered and addressed as a priority. First, the XGBoost model shows stable and strong performance across different application scenarios [50,78]. Considering the balance between efficiency and cost, the GIGS approach was developed based on an XGBoost model framework with a spatial weighting module. In future work, the GIGS approach should be further improved by incorporating advanced AI frameworks, such as transformer-based deep learning and pretrained foundation models [79,80]. In addition, although representative soil samples were collected for calibrating the GIGS model, and both independent validation and cross-validation were used to improve model reliability, limitations such as small sample size, limited validation samples, and the risk of overfitting cannot be avoided and should be addressed in future studies. Finally, the spectral estimation and digital mapping of soil Cd in this study focused on topsoil (0–20 cm). Future work can integrate multi-source remote sensing data and soil process models to obtain PTEs information in deeper soil layers [81].

5. Conclusions

Integrating the PCS data with a GIGS model developed in the study provided a framework for efficiently estimating and mapping soil Cd content and achieved satisfactory performance in a case study. A spatial weighting module from the GIGS model was used to analyze the spatial heterogeneity of each soil spectral sample. Each sample was assigned a spatial weight based on its geographical location, which helped reduce the negative effects of spatial heterogeneity on spectral modeling. The optimal principal component spectra (PCS) derived from spectral data were used as model inputs, with measured soil Cd content as the dependent variable, thereby establishing a robust spectral estimation model for soil Cd. Results indicated that the GIGS model exhibited satisfactory performance in the spectral estimation of soil Cd content, with R^2, RMSE, and RPIQ values of 0.78, 0.04, and 2.02, respectively. Compared to commonly used spectral estimation models (e.g., XGBoost, RF, SVR), the GIGS model showed an improvement of about 27.87% at the highest and about 16.42% at the lowest (based on R^2). Five PCS of soil spectral data were extracted as predictors for the model. These predictors were closely associated with clay minerals, organic matter, and iron oxides in the soil, thereby indirectly reflecting the state of Cd content. This study provides an efficient spectral estimation and mapping framework for

soil Cd content, which can strongly support soil quality monitoring and environmental sustainability around the world.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/rs18142341/s1: Table S1: Accuracy indices for the GIGS, XGBoost, RF, and SVR models; Figure S1: Map of soil Cd distribution in the study area.

Author Contributions: Conceptualization, J.C. and X.Z.; methodology, X.D. and X.Z.; software, Y.L.; validation, X.X., Q.S. and J.C.; formal analysis, Q.S.; investigation, X.X.; resources, X.L. and Y.L.; data curation, X.L.; writing—original draft preparation, J.C.; writing—review and editing, X.X.; visualization, Q.S.; supervision, X.X.; project administration, J.C.; funding acquisition, J.C., Y.L. and X.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of Shandong Province, China (ZR2025QC1023; ZR2024QD268), National Natural Science Foundation of China (42501074), and Key Laboratory of Environmental Change and Natural Disaster of the Ministry of Education, Beijing Normal University (2023-KF-12).

Data Availability Statement: The datasets used in this study can be obtained from the corresponding author upon reasonable request.

Acknowledgments: We gratefully acknowledge the support of the State Key Laboratory of Earth Surface Processes and Disaster Risk Reduction at Beijing Normal University.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chen, M.; Zhang, X.; Jiang, P.; Liu, J.; You, S.; Lv, Y. Advances in heavy metals detoxification, tolerance, accumulation mechanisms, and properties enhancement of Leersia hexandra Swartz. *J. Plant Interact.* **2022**, *17*, 766–778. [CrossRef]
2. Spanner, J.; Napolitano, G. *Healthy Soils Are the Basis for Healthy Food Production*; FAO: Rome, Italy, 2015; pp. 1–4.
3. Wang, Y.; Zhang, X.; Sun, W.; Wang, J.; Ding, S.; Liu, S. Effects of hyperspectral data with different spectral resolutions on the estimation of soil heavy metal content: From ground-based and airborne data to satellite-simulated data. *Sci. Total Environ.* **2022**, *838*, 156129. [CrossRef] [PubMed]
4. Guo, Z.; Zhang, Y.; Xu, R.; Xie, H.; Xiao, X.; Peng, C. Contamination vertical distribution and key factors identification of metal(loid)s in site soil from an abandoned Pb/Zn smelter using machine learning. *Sci. Total Environ.* **2023**, *856*, 159264. [CrossRef] [PubMed]
5. Yin, G.; Chen, X.; Zhu, H.; Chen, Z.; Su, C.; He, Z.; Qiu, J.; Wang, T. A novel interpolation method to predict soil heavy metals based on a genetic algorithm and neural network model. *Sci. Total Environ.* **2022**, *825*, 153948. [CrossRef] [PubMed]
6. Hou, D.; Jia, X.; Wang, L.; McGrath, S.P.; Zhu, Y.-G.; Hu, Q.; Zhao, F.-J.; Bank, M.S.; O'Connor, D.; Nriagu, J. Global soil pollution by toxic metals threatens agriculture and human health. *Science* **2025**, *388*, 316–321. [CrossRef] [PubMed]
7. Jeffrey, M.; Achurch, H. Save our soil to save the planet. In *Global Soil Security*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 389–395.
8. Suhani, I.; Sahab, S.; Srivastava, V.; Singh, R.P. Impact of cadmium pollution on food safety and human health. *Curr. Opin. Toxicol.* **2021**, *27*, 1–7. [CrossRef]
9. Ministry of Ecology and Environment of the People's Republic of China. *National Soil Contamination Survey Report*; Ministry of Ecology and Environment of the People's Republic of China: Beijing, China, 2014.
10. Lv, J.; Liu, Y.; Zhang, Z.; Dai, J. Factorial kriging and stepwise regression approach to identify environmental factors influencing spatial multi-scale variability of heavy metals in soils. *J. Hazard. Mater.* **2013**, *261*, 387–397. [CrossRef] [PubMed]
11. Meharg, A.A.; Meharg, C. The pedosphere as a sink, source, and record of anthropogenic and natural arsenic atmospheric deposition. *Environ. Sci. Technol.* **2021**, *55*, 7757–7769. [CrossRef] [PubMed]
12. Nriagu, J.O.; Pacyna, J.M. Quantitative assessment of worldwide contamination of air, water and soils by trace metals. *Nature* **1988**, *333*, 134–139. [CrossRef] [PubMed]
13. Wang, X.; Liu, X.; Yang, X.; Cheng, T.; Yang, Z. Environmental risks for application of iron and steel slags in soils in China: A review. *Pedosphere* **2021**, *31*, 28–42. [CrossRef]
14. Liu, S.; Zhao, H.; Wu, K.; Zhang, Z.; Hou, Y.; Chen, T.; Jin, Q. Evaluation of heavy metal distribution characteristics of agricultural soil–rice system in a high geological background area according to the influence index of comprehensive quality (IICQ). *Environ. Sci. Pollut. Res.* **2020**, *27*, 20920–20933. [CrossRef]

15. Jia, Z.; Wang, J.; Li, B.; Li, Y.; Zhou, Y.; Tong, G.; Yan, D.; Zhou, S. An integrated methodology for improving heavy metal risk management in soil-rice system. *J. Clean. Prod.* **2020**, *273*, 122797. [CrossRef]
16. Li, J.; Li, K.; Jiao, L.; Zang, F.; Li, X.; Yang, Y.; Mao, X.; Tai, X. Contamination, ecological-health risks, and sources of potentially toxic elements in road-dust sediments and soils. *Environ. Geochem. Health* **2023**, *45*, 8169–8186. [CrossRef] [PubMed]
17. Rossel, R.V.; Walvoort, D.; McBratney, A.; Janik, L.J.; Skjemstad, J. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75. [CrossRef]
18. Van Acker, T.; Theiner, S.; Bolea-Fernandez, E.; Vanhaecke, F.; Koellensperger, G. Inductively coupled plasma mass spectrometry. *Anal. Chem.* **2023**, *95*, 52–64. [CrossRef]
19. Dian, Y.; Le, Y.; Fang, S.; Xu, Y.; Yao, C.; Liu, G. Influence of spectral bandwidth and position on chlorophyll content retrieval at leaf and canopy levels. *J. Indian Soc. Remote Sens.* **2016**, *44*, 583–593. [CrossRef]
20. Wang, Y.; Zou, B.; Chai, L.; Lin, Z.; Feng, H.; Tang, Y.; Tian, R.; Tu, Y.; Zhang, B.; Zou, H. Monitoring of soil heavy metals based on hyperspectral remote sensing: A review. *Earth-Sci. Rev.* **2024**, *254*, 104814. [CrossRef]
21. Shi, T.; Chen, Y.; Liu, Y.; Wu, G. Visible and near-infrared reflectance spectroscopy—An alternative for monitoring soil contamination by heavy metals. *J. Hazard. Mater.* **2014**, *265*, 166–176. [CrossRef] [PubMed]
22. Chen, L.; Lai, J.; Tan, K.; Wang, X.; Chen, Y.; Ding, J. Development of a soil heavy metal estimation method based on a spectral index: Combining fractional-order derivative pretreatment and the absorption mechanism. *Sci. Total Environ.* **2022**, *813*, 151882. [CrossRef] [PubMed]
23. Huang, Y.; Chen, W.; Tan, W.; Deng, Y.; Yang, C.; Zhu, X.; Shen, J.; Liu, N. Transfer learning for enhancing the generality of leaf spectroscopic models. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *139*, 104481. [CrossRef]
24. Padarian, J.; Minasny, B.; McBratney, A.B. Machine learning and soil sciences: A review. *Soil* **2020**, *6*, 35–52. [CrossRef]
25. Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Machine learning predictive models for mineral prospectivity. *Ore Geol. Rev.* **2015**, *71*, 804–818. [CrossRef]
26. Tan, K.; Chen, L.; Wang, H.; Liu, Z.; Ding, J.; Wang, X. Estimation of heavy metal distribution in soil from airborne hyperspectral imagery. *J. Environ. Manag.* **2023**, *347*, 119196. [CrossRef]
27. Ma, X.; Wang, J.; Zhou, K.; Zhang, W.; Zhang, Z.; De Maeyer, P.; Van de Voorde, T. Data-driven estimation of metal elements using hyperspectral and geochemical data. *Ore Geol. Rev.* **2024**, *165*, 105877. [CrossRef]
28. Zhang, T.; Fu, Q.; Tian, R.; Zhang, Y.; Sun, Z. A spectrum contextual self-attention deep learning network for hyperspectral inversion of soil metals. *Ecol. Indic.* **2023**, *152*, 110351. [CrossRef]
29. Liang, H.; Song, Y.; Dai, Z.; Liu, H.; Zhong, K.; Feng, H.; Xu, L. Soil nitrogen and pH estimation based on GNNW-XGBoost. *Spectrochim. Acta A* **2025**, *330*, 125716. [CrossRef]
30. Wang, Y.; Zou, B.; Li, S.; Tian, R.; Zhang, B.; Feng, H.; Tang, Y. A hierarchical residual correction-based hyperspectral inversion method. *J. Hazard. Mater.* **2024**, *479*, 135699. [CrossRef] [PubMed]
31. Zhao, D.; Wang, J.; Miao, J.; Zhen, J.; Wang, J.; Gao, C.; Jiang, J.; Wu, G. Spectral features of Fe and organic carbon in estimating heavy metals. *Geoderma* **2022**, *426*, 116093. [CrossRef]
32. Zhang, S.; Chen, Y.; Zhang, Z.; Wang, S.; Wu, Z.; Hong, Y.; Wang, Y.; Hou, H.; Hu, Z.; Fei, T. VNIR estimation of heavy metals concentrations in soil. *Catena* **2022**, *219*, 106585. [CrossRef]
33. He, F.; Gu, L.; Wang, T.; Zhang, Z. Geo-ecological environmental evaluation using spatiotemporal big data. *J. Clean. Prod.* **2017**, *142*, 854–866. [CrossRef]
34. Mitran, T.; Gunnam, J.R.S.; Gourigari, S.; Kandrika, S. Assessment of soil heavy metals over an industrial area. *J. Geochem. Explor.* **2024**, *257*, 107379. [CrossRef]
35. Xiang, Z.; Wu, S.; Zhu, L.; Yang, K.; Lin, D. Pollution characteristics and source apportionment of heavy metals. *J. Environ. Sci.* **2024**, *143*, 23–34. [CrossRef]
36. Soriano, A.; Pallarés, S.; Pardo, F.; Vicente, A.; Sanfeliu, T.; Bech, J. Deposition of heavy metals in soils of an industrialised area. *J. Geochem. Explor.* **2012**, *113*, 36–44. [CrossRef]
37. Cai, Y.; Peng, C.; Qiu, S.; Li, Y.; Gao, Y. Determination of soil microbial biomass carbon. *Commun. Soil Sci. Plant Anal.* **2011**, *42*, 2824–2834. [CrossRef]
38. Liu, Y.; Fan, Y.; Fan, J.; Yue, J.; Chen, R.; Ma, Y.; Bian, M.; Yang, F.; Feng, H. Combining multiple spectral preprocessing and wavelength optimization methods improves potato aboveground biomass estimation. *Inf. Process. Agric.* **2025**, *12*, 511–521. [CrossRef]
39. Wang, T.; Zhang, Y.; Li, F.; Cao, N. Spectral estimation accuracy of plant phosphorus content. *Comput. Electron. Agric.* **2025**, *234*, 110205. [CrossRef]
40. Huang, Y.; Jiang, D.; Zhuang, D.; Fu, J. Evaluation of hyperspectral indices for chlorophyll-a estimation. *Int. J. Environ. Res. Public Health* **2010**, *7*, 2437–2451. [CrossRef] [PubMed]
41. Shin, H.; Yu, J.; Wang, L.; Jeong, Y.; Kim, J. Spectral interference of heavy metals on moisture signals. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2266–2275.

42. Geng, J.; Lv, J.; Pei, J.; Liao, C.; Tan, Q.; Wang, T.; Fang, H.; Wang, L. Prediction of soil organic carbon using hyperspectral data. *Comput. Electron. Agric.* **2024**, *220*, 108905. [CrossRef]
43. Greenacre, M.; Groenen, P.; Hastie, T.; d'Enza, A.; Markos, A.; Tuzhilina, E. Principal component analysis. *Nat. Rev. Methods Primers* **2022**, *2*, 100. [CrossRef]
44. Hasan, B.; Abdulazeez, A.M. A review of principal component analysis algorithm. *J. Soft Comput. Data Min.* **2021**, *2*, 20–30.
45. Bhagat, S.; Tung, T.; Yaseen, Z. Heavy metal contamination prediction using ensemble model. *J. Hazard. Mater.* **2021**, *403*, 123492. [CrossRef] [PubMed]
46. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T.; et al. Extreme Gradient Boosting [R package xgboost version 1.2. 0.1]. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual, 6–10 July 2020; pp. 13–17.
47. Grekousis, G. Geographical-XGBoost: A new ensemble model. *J. Geogr. Syst.* **2025**, *27*, 169–195. [CrossRef]
48. Xu, X.; Wang, X.; Yang, P.; Meng, Y.; Yu, D.; Li, C. Mapping soil salt contents using satellite imagery. *Catena* **2023**, *223*, 106900. [CrossRef]
49. Xia, K.; Wu, T.; Zhang, S.; Wang, S.; Li, X.; Shen, Q.; Li, T. High-precision estimation of soil organic matter using spectroscopy. *Geoderma* **2024**, *445*, 116877. [CrossRef]
50. Jia, Y.; Jin, S.; Savi, P.; Gao, Y.; Tang, J.; Chen, Y.; Li, W. GNSS-R soil moisture retrieval using XGBoost. *Remote Sens.* **2019**, *11*, 1655. [CrossRef]
51. Hengl, T.; Leenaars, J.; Shepherd, K.; Walsh, M.; Heuvelink, G.; Mamo, T.; Tilahun, H.; Berkhout, E.; Cooper, M.; Fegraus, E. Soil nutrient maps of Sub-Saharan Africa: Assessment of soil nutrient content at 250 m spatial resolution using machine learning. *Nutr. Cycl. Agroecosyst.* **2017**, *109*, 77–102. [CrossRef] [PubMed]
52. Bellon-Maurel, V.; Fernandez-Ahumada, E.; Palagos, B.; Roger, J.-M.; McBratney, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends Anal. Chem.* **2010**, *29*, 1073–1081. [CrossRef]
53. Xu, X.; Wang, Z.; Song, X.; Zhan, W.; Yang, S. A remote sensing-based strategy for mapping potentially toxic elements of soils: Temporal-spatial-spectral covariates combined with random forest. *Environ. Res.* **2024**, *240*, 117570. [CrossRef] [PubMed]
54. Kurkcuoglu, M. Analysis of the energy justice in natural gas distribution with multiscale geographically weighted regression (MGWR). *Energy Rep.* **2023**, *9*, 325–337. [CrossRef]
55. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
56. Ye, M.; Zhu, L.; Li, X.; Ke, Y.; Huang, Y.; Chen, B.; Yu, H.; Li, H.; Feng, H. Estimation of soil arsenic concentration using a geographically weighted XGBoost model based on hyperspectral data. *Sci. Total Environ.* **2023**, *858*, 159798. [CrossRef] [PubMed]
57. Xu, X.; Chen, Y.; Wang, M.; Wang, S.; Li, K.; Li, Y. Improving estimates of soil salt content by using two-date image spectral changes in Yinbei, China. *Remote Sens.* **2021**, *13*, 4165. [CrossRef]
58. Wu, D.; Liu, H.; Wu, J.; Gao, X.; Nyasha, N.K.; Cai, G.; Zhang, W. Bi-directional pollution characteristics and ecological health risk assessment of heavy metals in soil and crops. *Int. J. Environ. Res. Public Health* **2022**, *19*, 9669. [CrossRef] [PubMed]
59. *GB 15618–2018*; Soil Environmental Quality Risk Control Standard for Soil Contamination of Agricultural Land. Ministry of Ecology and Environment of the People's Republic of China: Beijing, China, 2018.
60. Rossel, R.V.; Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **2010**, *158*, 46–54. [CrossRef]
61. Ben-Dor, E.; Inbar, Y.; Chen, Y. The reflectance spectra of organic matter in the visible, near-infrared and shortwave infrared region (400–2500 nm) during a controlled decomposition process. *Remote Sens. Environ.* **1997**, *61*, 1–15. [CrossRef]
62. Xu, M.; Wang, S.; Zhang, S.; Li, F.; Ma, C.; Cheng, Z. Hyperspectral reflectance models for retrieving heavy metal content: Application in archaeological soil. *J. Infrared Millim. Waves* **2011**, *30*, 109–114. [CrossRef]
63. Sun, W.; Zhang, X.; Zou, B.; Wu, T. Exploring the potential of spectral classification in estimation of soil contaminant elements. *Remote Sens.* **2017**, *9*, 632. [CrossRef]
64. Seifi, M.; Ahmadi, A.; Neyshabouri, M.R.; Taghizadeh-Mehrjardi, R.; Bahrami, H.-A. Remote and Vis-NIR spectroscopy for soil salinization estimation. *Remote Sens. Appl. Soc. Environ.* **2020**, *20*, 100398. [CrossRef]
65. Wang, J.; Cui, L.; Gao, W.; Shi, T.; Chen, Y.; Gao, Y. Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy. *Geoderma* **2014**, *216*, 1–9. [CrossRef]
66. Xie, S.; Ding, F.; Chen, S.; Wang, X.; Li, Y.; Ma, K. Prediction of soil organic matter content based on characteristic band selection method. *Spectrochim. Acta A* **2022**, *273*, 120949. [CrossRef]
67. Jiang, Q.; Liu, M.; Wang, J.; Liu, F. Feasibility of using visible and near-infrared reflectance spectroscopy to monitor heavy metal contaminants in urban lake sediment. *Catena* **2018**, *162*, 72–79. [CrossRef]
68. Chen, H.; Jilili, A.; Liu, W.; Chen, J. Correlation between heavy metals, organic matter and pH value in soils along the Bortala River. *Res. Soil Water Conserv.* **2016**, *23*, 210–213.

69. Moosavi, V.; Zuravand, G.; Shamsi, S. Cluster-based local modeling meets deep learning: A novel approach to soil moisture estimation. *J. Hydrol.* **2024**, *635*, 131161. [CrossRef]
70. Shi, L.; Daly, K.; O'Rourke, S. Estimating mineral-associated organic carbon saturation using MIR spectroscopy. *Geoderma* **2025**, *454*, 117181. [CrossRef]
71. Ramirez-Lopez, L.; Schmidt, K.; Behrens, T.; Van Wesemael, B.; Demattê, J.A.; Scholten, T. Sampling optimal calibration sets in soil infrared spectroscopy. *Geoderma* **2014**, *226*, 140–150. [CrossRef]
72. Nawar, S.; Mouazen, A. Optimal sample selection for measurement of soil organic carbon using on-line vis-NIR spectroscopy. *Comput. Electron. Agric.* **2018**, *151*, 469–477. [CrossRef]
73. Fu, Y.; Li, F.; Guo, S.; Zhao, M. Cadmium concentration and fluxes in agricultural soil downstream of sewage irrigation area. *J. Hazard. Mater.* **2021**, *412*, 125203. [CrossRef] [PubMed]
74. Tóth, G.; Hermann, T.; Szatmári, G.; Pásztor, L. Maps of heavy metals in soils of the European Union and priority areas for assessment. *Sci. Total Environ.* **2016**, *565*, 1054–1062. [CrossRef] [PubMed]
75. Tan, K.; Ma, W.; Chen, L.; Wang, H.; Du, Q.; Du, P.; Yan, B.; Liu, R.; Li, H. Estimating distribution trends of soil heavy metals in mining areas using hyperspectral imagery. *J. Hazard. Mater.* **2021**, *401*, 123288. [CrossRef] [PubMed]
76. Li, F.; Xu, L.; Jin, C.; Wang, H. Structure of multi-stage composite genetic algorithm (MSC-GA) and its performance. *Expert Syst. Appl.* **2011**, *38*, 8929–8937. [CrossRef]
77. Shi, T.; Chen, Y.; Liu, H.; Wang, J.; Wu, G. Soil organic carbon estimation using visible–near-infrared spectroscopy with feature selection. *Appl. Spectrosc.* **2014**, *68*, 831–837. [CrossRef] [PubMed]
78. Xu, X.; Wang, Y.; Dai, X.; Shen, Q.; Wu, Q.; Wang, Z.; Cao, J. Improving satellite-based estimation and mapping of soil lead using enhanced spectral features and XGBoost model. *Remote Sens.* **2026**, *18*, 1446. [CrossRef]
79. Li, J.; Zheng, K.; Gao, L.; Ni, L.; Huang, M.; Chanussot, J. Model-Informed Multistage Unsupervised Network for Hyperspectral Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5516117. [CrossRef]
80. Wittstruck, L.; Waske, B.; Jarmer, T. Multi-Modal Vision Transformer for high-resolution soil texture prediction of German agricultural soils using remote sensing imagery. *Remote Sens. Environ.* **2025**, *331*, 114985. [CrossRef]
81. Helfenstein, A.; Mulder, V.; Heuvelink, G.B.; Hack-ten Broeke, M.J. Three-dimensional space and time mapping reveals soil organic matter decreases across anthropogenic landscapes in the Netherlands. *Commun. Earth Environ.* **2024**, *5*, 130. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

remote sensing

MDPI

Article

A Hierarchical Semantic Consistency Constraint Framework for Hyperspectral and LiDAR Data Joint Classification

Jie Shen [1,2], Yimeng Ma [1,2] and Houqun Yang [1,2,*]

[1] School of Computer Science and Technology, Hainan University, Haikou 570228, China; shenjie1234@hainanu.edu.cn (J.S.); yimeng@hainanu.edu.cn (Y.M.)
[2] Haikou Key Laboratory of Deep Learning and Big Data Application Technology, Hainan University, Haikou 570228, China
* Correspondence: yhq@hainanu.edu.cn

Highlights

What are the main findings?

- A hierarchical semantic consistency constraint framework (HSCC) is proposed for joint classification of HSI and LiDAR data.
- The framework progressively strengthens cross-modal interaction through a progressive interactive fusion network (PIFNet) and introduces a semantic consistency constraint (SCC) strategy to explicitly enforce feature similarity for the same object across different modalities and levels, effectively mitigating semantic drift.

What are the implications of the main findings?

- HSCC achieves state-of-the-art classification performance on three public datasets, providing a high-accuracy solution for multi-source remote sensing data fusion in complex land-cover scenes.
- HSCC offers a hierarchical semantic alignment paradigm for heterogeneous feature fusion of multi-source remote sensing data, which can be extended to other multimodal tasks.

Abstract

Hyperspectral image (HSI) and LiDAR data fusion is valuable for land-cover classification in complex surface scenes. Existing methods typically extract features from each modality independently and then consider how to fuse them, ignoring the semantic consistency between features of different modalities and across different hierarchical levels. Moreover, fully mining and exploiting the complementary information between multimodal remote sensing data remains a critical issue. To address these challenges, this paper proposes a hierarchical semantic consistency constraint (HSCC) framework for HSI and LiDAR data joint classification. The framework is co-constructed by a progressive interactive fusion network (PIFNet) and a semantic consistency constraint (SCC) strategy. Specifically, PIFNet progressively calibrates the semantic representations of multimodal features at different abstraction levels through Cross-Modal Shared Attention and Symmetric Cross-Attention mechanisms, promoting information parity in deep interactions. The SCC strategy establishes multi-level semantic associations and employs a semantic consistency constraint loss to guide the network to autonomously maintain the consistency of the same land-cover object across heterogeneous feature representations, thereby further enhancing the discriminative power of the fused features. Experiments on three public datasets, MUUFL, Houston2013, and Augsburg, demonstrate that HSCC outperforms current state-of-the-art

methods, validating its effectiveness in multi-source remote sensing data fusion classification tasks.

Keywords: land-cover classification; hyperspectral image (HSI); light detection and ranging (LiDAR); multimodal fusion; cross-attention

1. Introduction

With the rapid development of Earth observation technology, multi-source remote sensing data fusion has become a key technique for characterizing complex land surface scenes and improving land-cover interpretation accuracy [1,2]. In this context, hyperspectral imaging (HSI) and light detection and ranging (LiDAR) data have attracted sustained attention in the remote sensing community due to their unique and complementary information characteristics [3]. HSI captures fine spectral features of ground objects through dozens to hundreds of continuous narrow bands, providing crucial evidence for material identification [4] and demonstrating significant potential in tasks such as anomaly detection [5] and mixed pixel unmixing [6]. However, in highly complex urban and other surface environments, HSI data are often constrained by the phenomena of "same object with different spectra" and "different objects with same spectrum", as well as limited spatial details [7–9], leading to a bottleneck in classification using only spectral information. In contrast, LiDAR can actively acquire accurate three-dimensional elevation and geometric structure information, exhibiting strong robustness to illumination and atmospheric conditions [10], thus effectively compensating for the limitations of HSI in characterizing object morphology, vertical structure, and spatial distribution [11]. Exploring efficient fusion mechanisms for HSI and LiDAR data to overcome the heterogeneity among multi-source data and fully leverage their complementary advantages has become a frontier topic in multi-source remote sensing information processing, demonstrating great application potential in urban planning [12], ecological environment monitoring [13], precision agriculture management [14], and forest resource inventory [15].

In early studies, classification primarily relied on traditional machine learning methods such as support vector machines [16] and random forests [17]. These approaches typically required manual feature design and selection based on empirical knowledge, for example, extracting morphological profiles [18] or extended multi-attribute profiles [19], followed by simple concatenation of multi-source features before feeding them into a classifier. Although these methods validated the potential of multi-source data fusion, their performance was constrained by feature engineering heavily dependent on prior knowledge, making it difficult to adaptively capture the complex and deep nonlinear correlations between HSI and LiDAR data. Consequently, when processing high-dimensional and heterogeneous remote sensing data, their classification accuracy and generalization ability faced significant bottlenecks [20,21].

With the rise of deep learning, convolutional neural networks (CNNs), leveraging their powerful local perception capability, can effectively extract spatial–spectral features from HSI as well as texture and elevation information from LiDAR data. Roy et al. [22], Feng et al. [23], Zhang et al. [24], and Li et al. [25] designed dual-branch or multi-branch CNN architectures to process heterogeneous data, aiming to learn spectral–spatial and geometric elevation features separately and then fuse them effectively. To enhance feature representation capability, Xu et al. [26] and Li et al. [27] combined traditional morphological profiles with CNN deep features to construct hybrid feature models. However, these CNN-based methods are inherently limited by the local receptive field of CNNs, making it

difficult to effectively capture the long-range contextual associations widely present in remote sensing scenes, and the fused features still suffer from semantic ambiguity.

In recent years, the Transformer, with its global self-attention mechanism, has significantly enhanced the holistic understanding of complex remote sensing scenes [28]. Depending on how multimodal features interact within the Transformer architecture, the existing studies mainly follow two fusion strategies. The first adopts an early fusion approach, where Roy et al. [29], Feng et al. [30], and Yang et al. [31] concatenate or stack features from HSI and LiDAR at the input layer, forming a unified token sequence that is then fed into a Transformer encoder for joint modeling. The second is a deep fusion architecture based on cross-attention. Li et al. [32], Seneviratne et al. [33], and Zhao et al. [34] treat features from one modality as queries and features from the other modality as both keys and values, actively retrieving relevant information through cross-attention layers, thereby achieving deeper feature alignment and complementary information mining.

Although previous works have achieved notable results, two key challenges remain. On the one hand, existing methods typically lack consideration of the semantic consistency across modalities and across hierarchical levels when extracting features from HSI and LiDAR data [35–37]. Ideally, for the same land-cover object, its HSI and LiDAR features should exhibit high similarity at different network levels. However, due to modality heterogeneity and asynchronous learning of the network, this inherent semantic consistency is often disrupted, causing the spectral features and elevation features of the same object to deviate from each other in the feature space. This hierarchical semantic inconsistency problem hinders the follow-up fusion from fully exploiting complementary information and limits the model's ability to mine synergistic discriminative information across modalities. On the other hand, when employing cross-attention mechanisms to facilitate modality interaction, the existing studies [38,39] still mostly adopt unidirectional cross-attention. In this setup, the modality serving as the query dominates the optimization process, while the modality serving as the key/value remains in a passive state [40]. This further exacerbates cross-modal semantic inconsistency, because the features of the passive modality are difficult for the active modality to perceive equally. Moreover, some studies [41,42] try to stack multiple cross-attention modules to deepen fusion, yet they still confine themselves to simple addition or concatenation of the outputs from each branch.

Motivated by the above considerations, we propose a hierarchical semantic consistency constraint (HSCC) framework for the joint classification of HSI and LiDAR data. Through the synergistic collaboration of a progressive interactive fusion network (PIFNet) and a semantic consistency constraint (SCC) strategy, the framework guides the network to autonomously align cross-modal and cross-level semantic representations during multi-stage feature learning, thereby producing more discriminative joint features.

The main contributions of this paper are as follows:

1. We develop a HSCC framework for the joint classification of HSI and LiDAR data. By integrating the PIFNet and the SCC strategy, the framework supports progressive multimodal interaction and fusion, further improving the accuracy of HSI and LiDAR joint classification. Experimental results on three publicly available multimodal remote sensing classification benchmark datasets demonstrate that HSCC achieves optimal performance across multiple evaluation metrics, fully validating its effectiveness.
2. We design a PIFNet network to gradually narrow the modality discrepancy between HSI and LiDAR at the feature level and alleviate the potential single-modality dominance issue during deep interactions. Through the Cross-Modal Shared Attention (CMSA) and Symmetric Cross-Attention (SCA) mechanisms, PIFNet progressively strengthens cross-modal information interaction layer by layer, calibrating the semantic shifts between spectral and elevation information at different levels and thereby

generating more consistent multi-level feature representations. These features are finally fused into a compact and semantically clear joint feature space, effectively supporting the classification task.

3. We propose a SCC strategy that constructs multiple groups of semantic associations between features of different modalities and different levels, and introduces a semantic consistency constraint loss to explicitly guide the network to maintain the intrinsic similarity of features for the same land-cover object. This strategy can effectively alleviate the semantic inconsistency between features from different modalities and enhance the network's perception capability for semantic alignment.

2. Related Work

2.1. CNN–Transformer Networks

Hybrid networks that combine the local perception capability of CNNs with the global context modeling advantage of Transformers have become an important research direction. Sun et al. [43] adopted a multi-scale 3D–2D hybrid CNN for feature extraction and designed a lightweight attention-free Transformer encoder for information integration. Wang et al. [44] used multi-scale asymmetric depthwise convolutions to capture shallow local-global correlations, and then extracted deep joint features via a local–global Transformer. Bai et al. [45] designed a hierarchical convolution module coupled with octave convolution for initial feature extraction and fusion, and captured cross-dimensional dependencies through a triplet rotational multi-head attention mechanism. To fully integrate heterogeneous data, many studies have been devoted to designing more complex cross-modal interaction mechanisms. Zhang et al. [46] introduced a diffusion model to mine latent features of HSI, and then constructed a three-input diffusion-augmented CNN–Transformer network, aggregating multi-source information via a spectral pooling attention interaction mechanism and an inverted residual structure. Wang et al. [47] used CNN and a Transformer in parallel during the feature fusion stage to explore deep features and interaction information across modalities. Although the above hybrid networks have achieved remarkable progress, they typically set up separate feature extraction branches for the two types of remote sensing data, with inter-modal interactions occurring only at deep stages of the network. Such a design makes it difficult to ensure the consistency between spectral and elevation information of the same ground object, thereby limiting the full integration of complementary information to some extent.

2.2. Cross-Attention

The cross-attention mechanism, capable of explicitly establishing semantic correlations between features of different modalities, has become key to achieving deep fusion. Ge et al. [48] combined local convolution with global cross-attention, enhanced convergence via residual aggregation, and simultaneously captured contextual details and long-range dependencies to achieve multi-source feature fusion and relational semantic representation. Liu et al. [49] introduced a neighborhood contrastive learning mechanism to enhance the discriminative ability of feature representations through a cross-source Transformer. Guo et al. [50] addressed feature redundancy in traditional stacking fusion by adopting a dual-branch pyramid structure for multi-scale feature extraction and achieving dynamic alignment of spectral and elevation information via a bidirectional cross-attention mechanism. Li et al. [51] used HSI features as queries and concatenated the key-value pairs from both modalities separately, enabling deep interaction of dual-modal features during attention weight computation. Zhang et al. [52] incorporated a cross-modal attention fusion module into an adaptive multi-stage fusion framework, dynamically capturing cross-modal correlations between hyperspectral and LiDAR features through an adaptive

https://doi.org/10.3390/rs18122058

gating mechanism. To more comprehensively characterize dependencies among multi-modal data, Jiao et al. [53] proposed a Common-Unique Transformer integrating a cross dynamic-agent-attention mechanism, which explicitly separates and fuses shared and unique features from different data sources. The above methods have conducted valuable explorations in cross-modal interaction, but most are limited to unidirectional querying or simple concatenation/linear combination of the two branch outputs. How to achieve deeper attentional synergy while ensuring bidirectional equal interaction remains an open issue for further research.

3. Proposed Method

In this section, we first outline the overall architecture of the HSCC framework, and then introduce the PIFNet and the SCC strategy in detail.

3.1. HSCC Framework

To address the problems of cross-modal semantic inconsistency and imbalanced modality contribution during interaction in HSI and LiDAR joint classification, this paper proposes a HSCC framework. As shown in Figure 1, the framework consists of two core components: PIFNet network and SCC strategy. Specifically, PIFNet is responsible for multi-stage feature alignment and bidirectional interactive fusion, while the SCC strategy imposes consistency constraints on the features output by PIFNet by establishing multi-level semantic correspondences.

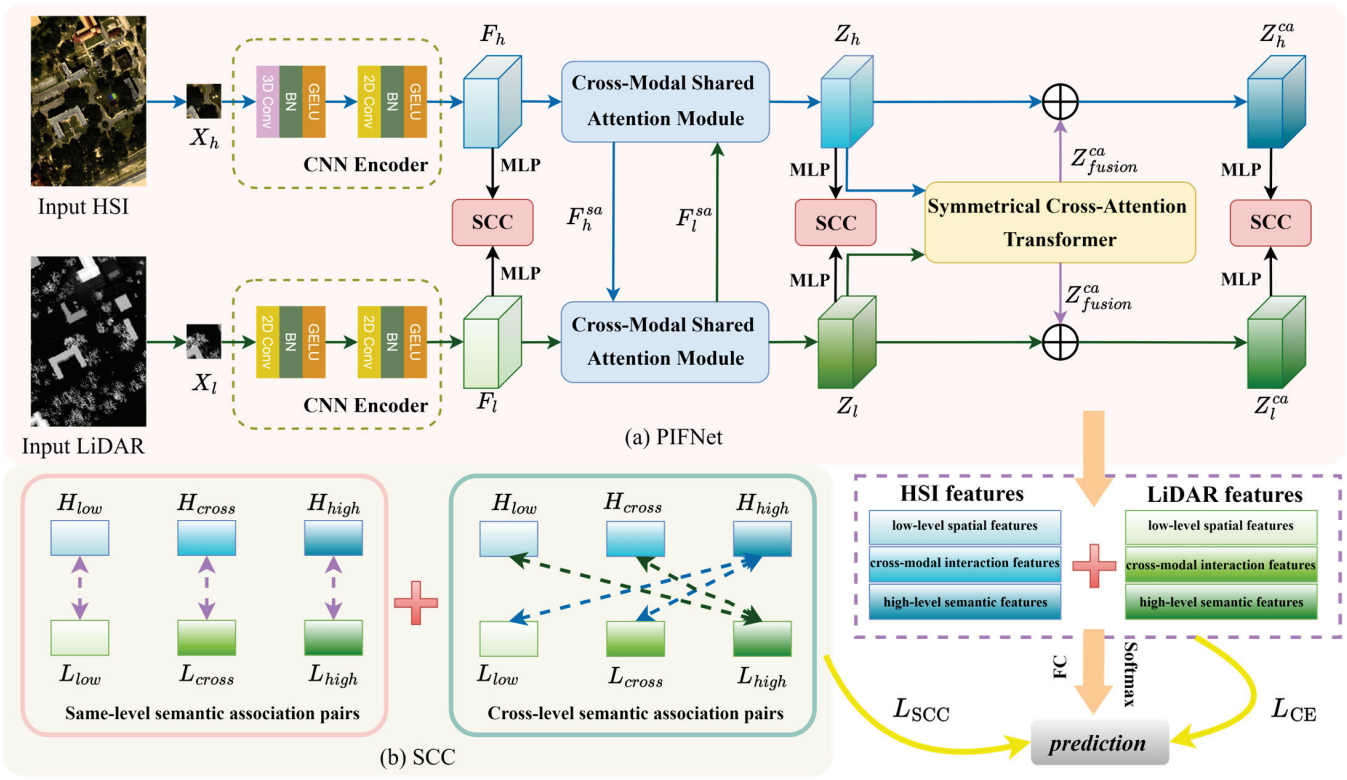


Figure 1. Overall framework of HSCC. (**a**) Architecture of PIFNet. (**b**) SCC strategy.

In PIFNet, a progressive alignment mechanism is adopted to gradually enhance the information interaction between HSI and LiDAR across multiple levels. First, the network performs initial feature extraction and dimension unification for both modalities using simple convolutions. Subsequently, the Cross-Modal Shared Attention (CMSA) module generates and shares the same set of spatial attention maps for HSI and LiDAR, establishing preliminary

cross-modal correlations while preserving the unique attributes of each modality, thereby guiding the network to focus on discriminative regions to which both modalities respond jointly. On this basis, the SCAT module achieves equal cross-modal information integration through Symmetrical Cross-Attention mechanism, effectively suppressing the dominant bias of a single modality. After the above progressive processing, PIFNet gradually produces multi-level aligned features, including low-level spatial features, cross-modal interactive features, and high-level semantic abstraction features. Finally, these multi-level cross-modal features are fused to form a unified joint feature representation for subsequent classification tasks. Meanwhile, the SCC strategy constructs same-level and cross-level semantic association pairs between HSI and LiDAR features, thereby establishing multi-level semantic correspondences. By optimizing the similarity among these semantic associations, SCC effectively suppresses semantic deviations across different modalities and levels, ultimately improving the semantic consistency and representational quality of the fused features.

For the same ground surface area, given a hyperspectral image $H \in \mathbb{R}^{m \times n \times c_h}$ and the corresponding LiDAR image $L \in \mathbb{R}^{m \times n \times c_l}$, where m and n denote the height and width of the images, respectively, and c_h and c_l denote the number of bands of the HSI and LiDAR data, respectively. After padding the image boundaries, for each foreground pixel, a pair of image patches centered at that pixel is extracted, resulting in a HSI cube $X_h \in R^{p \times p \times c_h}$ and a LiDAR cube $X_l \in R^{p \times p \times c_l}$ as inputs, where p is the patch size. Our goal is to predict a class label y for each central pixel, where $y \in \{1, 2, ..., CL\}$ and CL is the total number of classes.

Algorithm 1 presents the detailed implementation of the HSCC framework. In each iteration, PIFNet first progressively aggregates the hierarchical features of the two modalities through multi-stage interactions, and on this basis forms discriminative representations oriented toward classification. Subsequently, the SCC strategy establishes multiple groups of semantic associations within the aligned feature space to guide cross-level and cross-modal feature consistency. The two objectives are jointly constrained under a unified framework, collectively driving the update of network parameters.

Algorithm 1 Training Procedure of the HSCC Framework

Input: HSI cube X_h, LiDAR cube X_l, epochs E
Output: Classification map P

1. Initialize network parameters;
2. for each training epoch e = 1 to E do
    3. Extract HSI shallow features h_{low} and LiDAR shallow features l_{low};
    4. Obtain cross-modal interaction features h_{cross} and l_{cross} for HSI and LiDAR respectively via the CMSA module;
    5. Perform bidirectional symmetric attention interaction through SCAT, output HSI and LiDAR high-level semantic features h_{high}, l_{high};
    6. Integrate multi-level features to compute classification loss L_{CE};
    7. for each feature in $\{h_{low}, l_{low}, h_{cross}, l_{cross}, h_{high}, l_{high}\}$ do
        8. Construct same-level semantic association pairs;
        9. Construct cross-level semantic association pairs;
        10. Compute the similarity of each association pair to obtain the semantic consistency constraint loss L_{SCC};
    11. end for
    12. Evaluate classification loss L_{CE} and semantic consistency constraint loss L_{SCC};
    13. Backpropagate and update network parameters;
14. end for

https://doi.org/10.3390/rs18122058

3.2. PIFNet

The joint classification of HSI and LiDAR is essentially a dual task of cross-modal feature collaboration and semantic unification. However, the existing methods generally adopt a strategy of first encoding each modality independently and then fusing them, which implicitly assumes that the two modalities naturally possess semantic correspondences after being learned separately. Furthermore, when using cross-attention mechanisms for deep fusion, one modality tends to dominate, leading to an imbalance in the contribution of multimodal data. To address these issues, this paper designs the PIFNet. Through a multi-stage symmetric interaction mechanism, PIFNet alleviates modality imbalance while promoting the gradual semantic convergence of multimodal features. As shown in Figure 1a, PIFNet consists of the following four parts.

3.2.1. Feature Alignment Encoding

Considering the rich spectral information contained in HSI data, we first use a $3 \times 3 \times 3$ 3D convolutional layer to extract spectral sequence features, and simultaneously use a 3×3 2D convolutional layer to capture texture and structural features. The numbers of filters in the two layers are set to 4 and 64, respectively. For the LiDAR branch, we extract elevation information using two stacked 3×3 convolutional layers, with the numbers of filters set to 32 and 64, respectively. In addition, to accelerate model convergence, each convolutional layer in both the HSI and LiDAR branches is sequentially followed by a batch normalization (BN) layer and a Gaussian error linear unit (GELU) activation function. After the above processing, the patches of the two modalities are respectively mapped to $F_h \in R^{p \times p \times c}$ and $F_l \in R^{p \times p \times c}$, where $c = 64$.

In this stage, the feature encoding of HSI and LiDAR is performed independently, with interaction occurring only after the feature dimensions are unified. This design avoids semantic confusion caused by the premature fusion of heterogeneous shallow features, while ensuring structural compatibility between the feature representations of the two modalities, thus creating favorable conditions for the subsequent progressive interaction.

3.2.2. Cross-Modal Shared Attention Module

As shown in Figure 2, the proposed mechanism generates and shares the same set of spatial attention weights for HSI and LiDAR, thereby preserving the discriminative features of each modality while guiding the network to focus on spatial regions that elicit common responses from the multi-source data.

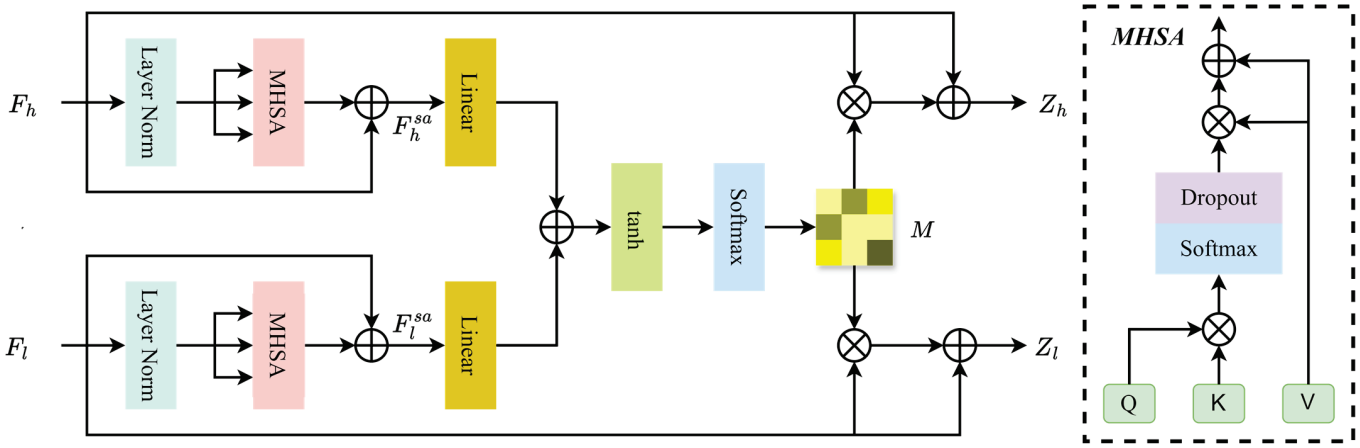


Figure 2. Schematic diagram of the CMSA module.

First, to enhance the internal contextual correlations of each modality's features, we apply a multi-head self-attention mechanism to the previously obtained HSI and LiDAR features respectively for context aggregation. This process can be expressed as:

$$F^{sa} = \mathrm{MHSA}\big(\mathrm{LN}\big(\hat{F}\big)\big) + \hat{F} \tag{1}$$

In the formula, $\hat{F}$ denotes the flattened feature of F, $\mathrm{MHSA}(\cdot)$ represents the multi-head self-attention mechanism, and $\mathrm{LN}(\cdot)$ denotes the layer normalization operation. For each attention head, we use a convolutional layer instead of the linear projection in the standard Transformer to generate Q_i, K_i, and V_i, where i denotes the ith head. The self-attention SA of the ith head can be calculated using the following formula:

$$\mathrm{SA}_i(Q, K, V) = \mathrm{Softmax}\left(\frac{Q_i K_i^T}{\sqrt{d/n}}\right) V_i \tag{2}$$

where T denotes the transpose operation, d denotes the channel dimension, and n is the number of attention heads. Finally, the outputs of all attention heads are concatenated to obtain the context-aggregated features.

Next, the context-aggregated features F_h^{sa} and F_l^{sa} of HSI and LiDAR are respectively projected into a common subspace. After summation and tanh activation, a shared spatial attention weight map M is generated via Softmax normalization. This weight map reflects the regions commonly attended to by both modalities and is applied to the original aggregated features of both modalities in the same manner to recalibrate the spatial dimension. Finally, each modality integrates the calibrated features with the original features through residual connections to obtain the outputs Z_h and Z_l. The specific formulas are as follows:

$$M = \mathrm{Softmax}(\tanh(\mathrm{Linear}(F_h^{sa}) + \mathrm{Linear}(F_l^{sa}))) \tag{3}$$

$$Z_h = M \odot F_h^{sa} + F_h^{sa} \tag{4}$$

$$Z_l = M \odot F_l^{sa} + F_l^{sa} \tag{5}$$

where $M \in \mathbb{R}^{p \times p \times 1}$, $\mathrm{Linear}(\cdot)$ denotes a fully connected layer, $\tanh(\cdot)$ is the hyperbolic tangent activation function, and $\odot$ denotes element-wise multiplication.

This design achieves semantically aware spatial guidance through the shared attention maps, establishing a spatial constraint foundation for cross-modal semantic consistency before entering deep fusion, and thereby effectively suppressing subsequent semantic shifts caused by modality dominance bias.

3.2.3. Symmetrical Cross-Attention Transformer

Attention-based fusion strategies can more flexibly capture semantic correlations between modalities, thereby obtaining more discriminative feature representations. However, the cross-attention mechanism used in existing methods tends to implicitly allow one modality to dominate during deep fusion, while the unique information of the other modality may be suppressed or diluted. This not only compromises the reliability of the fused features but also exacerbates the difficulty of cross-modal semantic alignment.

To address the above modality imbalance problem, this paper proposes a Symmetric Cross-Attention Transformer (SCAT), whose structure is shown in Figure 3. Specifically, we construct equal bidirectional interaction channels for HSI and LiDAR, where both modalities have equal opportunities to serve as the query and the value, and the attention maps generated from both directions are jointly learned and fused, thereby ensuring equal status for the two modalities during information interaction. We use lightweight

https://doi.org/10.3390/rs18122058

convolutions instead of linear layers to generate Q, K, and V for each modality, reducing the number of parameters while preserving spatial structure awareness. Taking the HSI feature Z_h as an example, the generation process is as follows:

$$Q_h = Conv_{1\times1}(DwConv_{3\times3}(LN(Z_h))) \tag{6}$$

$$K_h = Conv_{1\times1}(DwConv_{3\times3}(LN(Z_h))) \tag{7}$$

$$V_h = Conv_{1\times1}(DwConv_{3\times3}(LN(Z_h))) \tag{8}$$

where $LN(\cdot)$ denotes the layer normalization operation, $DwConv(\cdot)$ represents depthwise separable convolution, and the generation process of Q_l, K_l, and V_l for the LiDAR branch is the same as this.

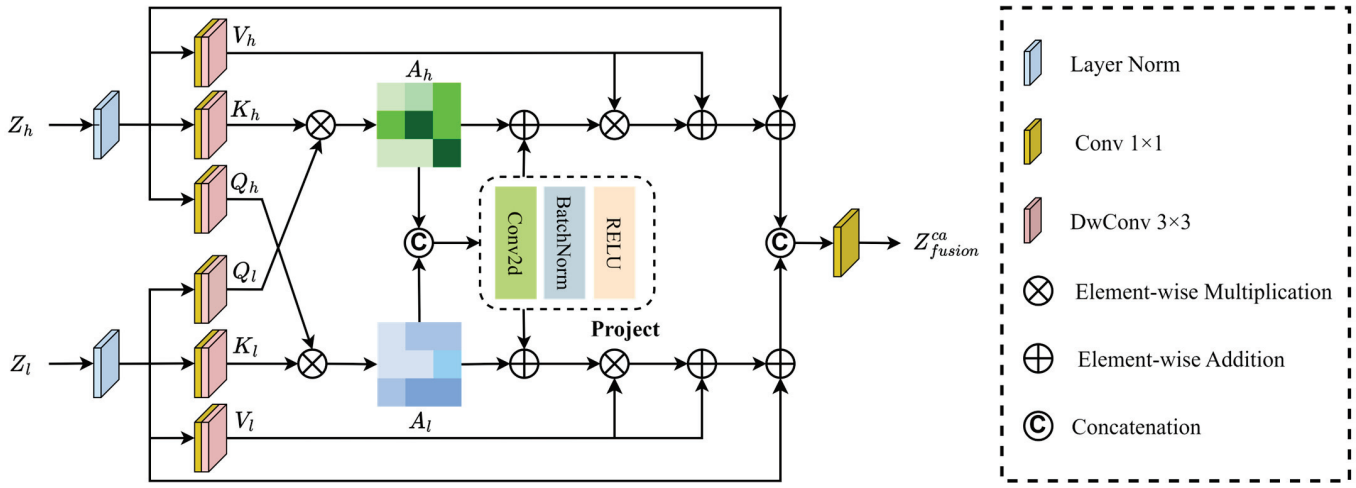


Figure 3. Schematic diagram of SCAT.

Subsequently, the query Q_h of HSI features and the key K_l of LiDAR features are used to compute the attention map A_h, while the query Q_l of LiDAR features and the key K_h of HSI features are used to compute the attention map A_l. To enhance the correlation between the two attention pathways, A_h and A_l are concatenated and passed through a lightweight projection network to extract a shared interaction pattern, generating a collaborative attention component, which is then added to the respective attention maps via residual connections to obtain A'_h and A'_l. The specific calculation process is as follows:

$$A_h = \frac{Q_h^T K_l}{\tau_h} \tag{9}$$

$$A_l = \frac{Q_l^T K_h}{\tau_l} \tag{10}$$

$$A'_h = Dropout(Softmax(Project([A_h, A_l]) + A_h)) \tag{11}$$

$$A'_l = Dropout(Softmax(Project([A_h, A_l]) + A_l)) \tag{12}$$

where τ_h and τ_l are learnable scaling parameters, $[\cdot]$ denotes the concatenation operation, and $Dropout(\cdot)$ is used to control the generalization ability of the model with the dropout rate set to 0.1.

Finally, A'_h and A'_l are used to perform weighted summation on V_h and V_l respectively, and then residual connections are applied to obtain Z'_h and Z'_l. The two are then

concatenated and fused through a 1×1 convolution for dimensionality reduction, yielding the cross-modal fusion feature $Z_{\text{fusion}}^{\text{ca}}$.

$$Z_h' = A_h' V_h + V_h + Z_h \tag{13}$$

$$Z_l' = A_l' V_l + V_l + Z_l \tag{14}$$

$$Z_{\text{fusion}}^{\text{ca}} = \text{Conv}_{1\times1}\left(\left[Z_h', Z_l'\right]\right) \tag{15}$$

The above symmetric fusion design ensures that HSI and LiDAR have equal semantic contribution capabilities during the fusion process, effectively mitigating modality dominance bias and providing a well-aligned cross-modal feature foundation for the semantic consistency constraint.

3.2.4. Classification and Loss Function

To effectively align multi-modal features at different levels under semantic consistency constraints, we perform projection transformations on the low-level features, cross-modal interactive features, and high-level semantic features of HSI and LiDAR, respectively. For the low-level and cross-modal interactive features, we adopt a parameter-shared multilayer perceptron (MLP) to map them into a common semantic subspace. This MLP first expands the original features to a higher dimension to enhance representational capacity, then applies batch normalization and the GELU activation function to stabilize the training process, introduces dropout to mitigate overfitting, and finally compresses the features to the number of classes. For the high-level semantic features, they are refined by another lightweight projection network with a similar structure, except that the features are first appropriately reduced in dimensionality while maintaining alignment with the shallow features. The projected features are fed into the SCC strategy for semantic consistency measurement, forcing representations that were originally separated due to modality or level differences to converge semantically.

We sum the features obtained from HSI and LiDAR at each stage to form a unified multi-level feature representation. The aggregated features are finally fed into a fully connected layer for pixel-wise classification. The overall loss function is defined as shown in Equation (16), which consists of two terms: the classification loss $\mathcal{L}_{\text{CE}}$ and the semantic consistency constraint loss $\mathcal{L}_{\text{SCC}}$. The classification loss adopts the standard cross-entropy form, while the detailed design of $\mathcal{L}_{\text{SCC}}$ will be elaborated in Section 3.3.

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \mu\mathcal{L}_{\text{SCC}} \tag{16}$$

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{cl=1}^{CL} y_{i,cl}\log\left(\hat{y}_{i,cl}\right) \tag{17}$$

where a balancing coefficient μ controls the contribution of the $\mathcal{L}_{\text{SCC}}$, N is the total number of samples, CL is the total number of classes, $y_{i,cl}$ is the ground-truth label indicating whether sample i belongs to class cl, and $\hat{y}_{i,cl}$ is the predicted probability that sample i belongs to class cl.

3.3. SCC Strategy

For HSI and LiDAR data at the same spatial location, features at the same level of abstraction should inherently be consistent. This intra-level semantic correspondence forms the foundation of cross-modal understanding. Furthermore, features from different modalities at different levels should also maintain semantic associations. Cross-level semantic consistency facilitates the full exploitation of complementary information from heterogeneous data sources. To address these requirements, this paper proposes the SCC

https://doi.org/10.3390/rs18122058

strategy, which simultaneously establishes two types of semantic associations. By explicitly increasing the similarity between these semantically corresponding features, the proposed strategy improves the reliability of the fused representation. This process is illustrated in Figure 1b.

In implementation, SCC constructs multiple groups of semantic associations by aligning features at the same level across the two modalities and by matching the high-level semantic features of one modality with the low-level and cross-modal interaction features of the other. Since semantic alignment essentially depends on the directional consistency of feature vectors, and cosine similarity directly measures directional correlation through the cosine of the angle, is invariant to feature scaling, and has a bounded range that facilitates loss optimization, this paper adopts cosine similarity to measure the consistency of each semantic association pair and trains the model using the semantic consistency constraint loss $\mathcal{L}_{SCC}$, encouraging the feature representations of the same pixel to maintain high semantic similarity across different modalities and levels. The specific formulation is as follows:

$$\mathcal{L}_{SCC} = \mathcal{L}_{SL} + \mathcal{L}_{CL1} + \mathcal{L}_{CL2} \tag{18}$$

$$\mathcal{L}_{SL} = (1 - \mathrm{sim}(h_{low}, l_{low})) + (1 - \mathrm{sim}(h_{cross}, l_{cross})) + \left(1 - \mathrm{sim}\left(h_{high}, l_{high}\right)\right) \tag{19}$$

$$\mathcal{L}_{CL1} = \left(1 - \mathrm{sim}\left(h_{high}, l_{low}\right)\right) + \left(1 - \mathrm{sim}\left(h_{high}, l_{cross}\right)\right) \tag{20}$$

$$\mathcal{L}_{CL2} = \left(1 - \mathrm{sim}\left(l_{high}, h_{low}\right)\right) + \left(1 - \mathrm{sim}\left(l_{high}, h_{cross}\right)\right) \tag{21}$$

$$\mathrm{sim}(\alpha, \beta) = \frac{\alpha \cdot \beta}{\|\alpha\| \|\beta\|} \tag{22}$$

where h_{low} and l_{low} denote the low-level features extracted by the CNN encoder; h_{cross} and l_{cross} represent the preliminary cross-modal interaction features generated via the CMSA module; and h_{high} and l_{high} are the high-level semantic features produced by the SCAT module, which contain rich semantic information and deeply fused information.

4. Results

4.1. Dataset

1. MUUFL dataset: The MUUFL dataset was acquired in November 2010 from the Gulf Park campus of the University of Southern Mississippi in Long Beach, Mississippi, USA. This dataset contains two data sources: hyperspectral images and LiDAR data. The hyperspectral data were collected by the ITRES CASI-1500 sensor, consisting of 325 × 220 pixels with 64 effective spectral channels, covering a wavelength range from 0.38 to 1.05 μm. The LiDAR data were simultaneously acquired by the Gemini ALTM sensor with a wavelength of 1.06 μm. The dataset contains a total of 53,687 labeled pixels corresponding to 11 different land cover categories. Figure 4 shows the false-color composite image of the HSI data, the grayscale image of the LiDAR data, and the ground-truth labels.
2. Houston2013 dataset: The Houston2013 dataset was provided by the IEEE GRSS Data Fusion Contest and was acquired over the University of Houston campus and its surrounding urban area in June 2012. The HSI data were collected by the ITRES CASI-1500 sensor, containing 144 spectral bands with wavelengths ranging from 0.38 to 1.05 μm, while the simultaneously acquired LiDAR data are a single-band digital surface model. Both modalities have dimensions of 349 × 1905 pixels with a spatial resolution of 2.5 m. The dataset contains a total of 15,029 labeled pixels corresponding to 15 different land cover types. Figure 5 shows the false-color composite image of the HSI data, the grayscale image of the LiDAR data, and the ground-truth labels.

https://doi.org/10.3390/rs18122058

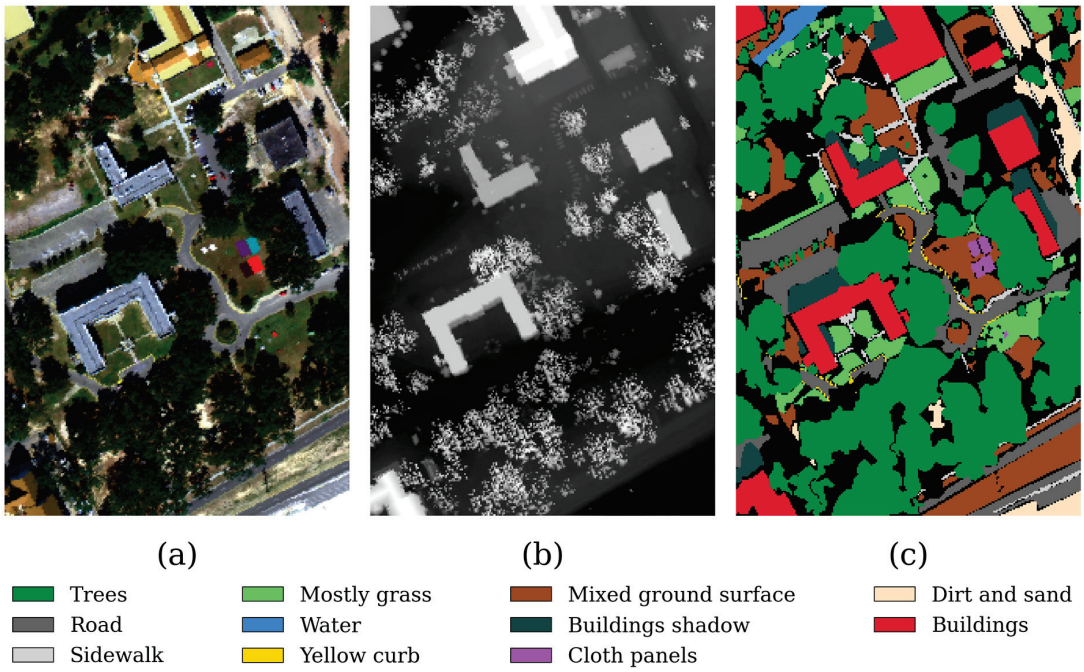


Figure 4. Visualization results of the MUUFL dataset. (**a**) HSI false-color image. (**b**) LiDAR grayscale image. (**c**) Ground-truth label map. Different colors in the legend at the bottom represent different land cover classes.

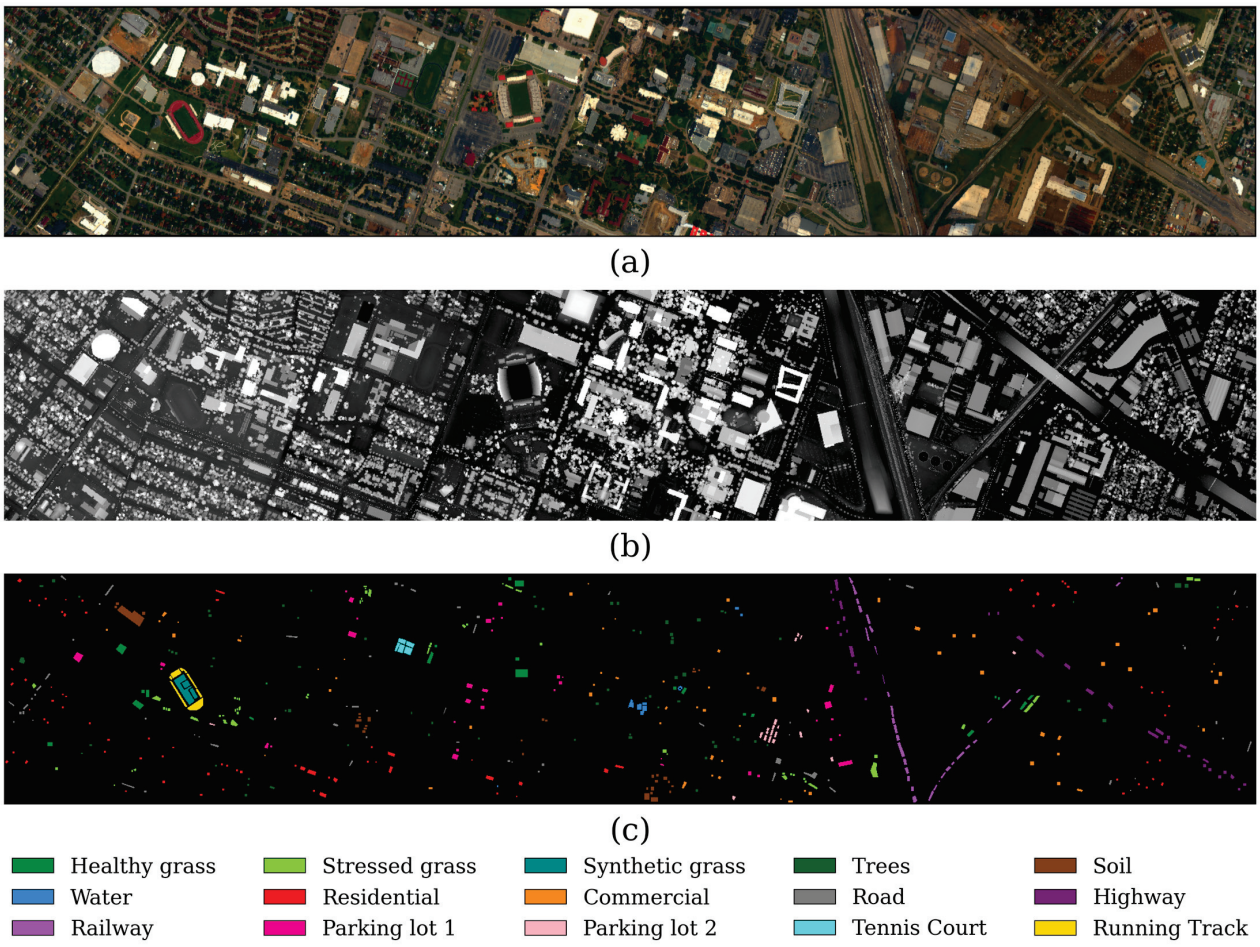


Figure 5. Visualization results of the Houston2013 dataset. (**a**) HSI false-color image. (**b**) LiDAR grayscale image. (**c**) Ground-truth label map. Different colors in the legend at the bottom represent different land cover classes.

https://doi.org/10.3390/rs18122058

3. Augsburg dataset: The Augsburg dataset was acquired over the city of Augsburg, Germany, and its surrounding areas. The HSI data were collected by the DAS-EOC HySpex sensor, covering 180 continuous spectral bands with wavelengths ranging from 0.4 μm to 2.5 μm. The LiDAR data were obtained by the DLR-3K system. The spatial dimensions of the dataset are 332 × 485 pixels, comprising 7 land cover categories. Figure 6 shows the false-color composite image of the HSI data, the grayscale image of the LiDAR data, and the ground-truth labels. Table 1 lists the class names and the number of training and test samples for each class for the three datasets: MUUFL, Houston2013, and Augsburg.

Table 1. Number of samples per class in the three datasets.

No.	MUUFL			Houston2013			Augsburg		
	Class Name	Train	Test	Class Name	Train	Test	Class Name	Train	Test
1	Trees	50	23,196	Healthy grass	20	1231	Forest	270	13,237
2	Mostly grass	50	4220	Stressed grass	20	1234	Residential area	607	29,722
3	Mixed ground surface	50	6832	Synthetic grass	20	677	Industrial area	77	3774
4	Dirt and sand	50	1776	Trees	20	1224	Low plants	537	26,320
5	Road	50	6637	Soil	20	1222	Allotment	12	563
6	Water	50	416	Water	20	305	Commercial area	33	1612
7	Buildings shadow	50	2183	Residential	20	1248	Water	31	1499
8	Buildings	50	6190	Commercial	20	1224			
9	Sidewalk	50	1335	Road	20	1232			
10	Yellow curb	50	133	Highway	20	1207			
11	Cloth panels	50	219	Railway	20	1215			
12				Parking lot 1	20	1213			
13				Parking lot 2	20	449			
14				Tennis Court	20	408			
15				Running Track	20	640			
	Total	550	53,137	Total	300	14,729	Total	1567	76,727

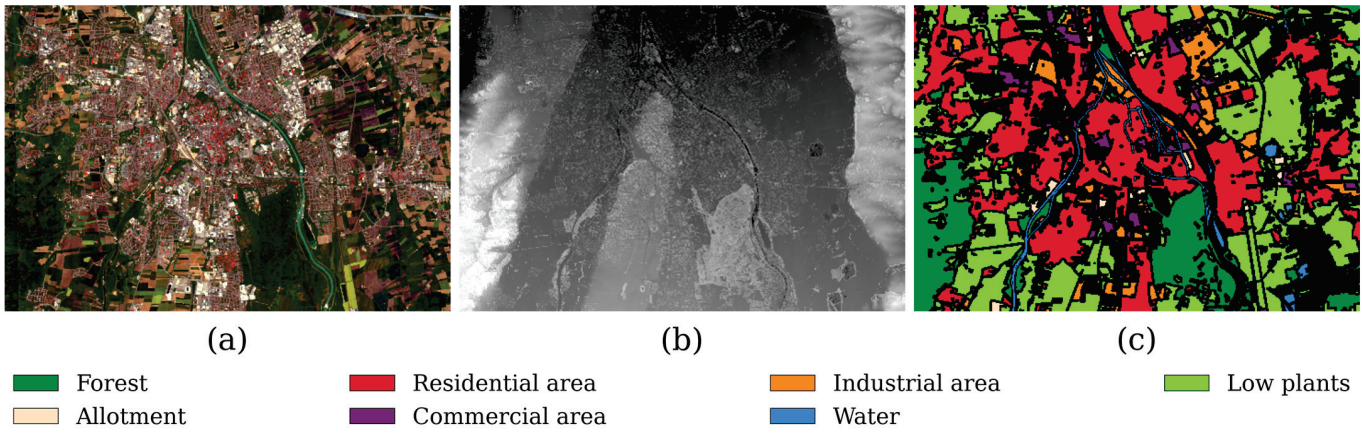


Figure 6. Visualization results of the Augsburg dataset. (**a**) HSI false-color image. (**b**) LiDAR grayscale image. (**c**) Ground-truth label map. Different colors in the legend at the bottom represent different land cover classes.

4.2. Experimental Setup

1. Evaluation metrics: To objectively evaluate the classification performance of the proposed network and other compared methods, we select three widely used quantitative evaluation metrics in the field of remote sensing image classification: Overall Accuracy (OA), Average Accuracy (AA), and Kappa coefficient (Kappa). OA represents the proportion of correctly classified samples to the total number of test samples. AA

denotes the arithmetic mean of classification accuracies for each class. Kappa reflects the consistency between the classification results and the ground-truth distribution. For each metric, a higher value indicates better classification performance.

2. Implementation details: We implement the proposed model using the PyTorch 2.0.1 deep learning framework. All experiments are conducted on a server equipped with a NVIDIA GeForce RTX 4090 GPU (24 GB memory). In the training setup, the number of epochs is set to 500, and the batch size is 64. We adopt the Adam optimizer to update network parameters, and the learning rate scheduler uses the StepLR strategy, decaying the learning rate to 0.9 times its previous value every 50 epochs. To monitor model performance and save the best weights, the model is evaluated on the test set every 10 epochs during training.

4.3. Classification Results and Analysis

To demonstrate the effectiveness of the proposed HSCC, we compared it with eight other state-of-the-art HSI and LiDAR joint classification methods, including IDNet [54], DSymFuser [55], mPMCL [56], FDNet [57], MICFNet [58], AFDSE [59], MCFNet [60], and PICNet [61]. All compared methods were implemented using the original code released by their respective authors, and the model parameters were strictly set according to the recommendations in the corresponding original papers. Furthermore, all methods were evaluated on the same training and test sets, where the data splits were determined by random sampling and kept consistent across all comparative experiments. Each method is described as follows:

IDNet: This method embeds Laplacian high-frequency features into a Transformer architecture, learns high-frequency features through a dynamic range-aware and similarity-based spatial–spectral feature grouping interaction mechanism, and employs differential convolution to enhance high-frequency information during the fusion stage, thereby achieving adaptive dynamic fusion.

DSymFuser: This method adopts a dual-branch Transformer framework, where each branch adaptively aggregates local and global features extracted by CNN and Transformer, and generates bidirectional cross-modal features through a single attention computation.

mPMCL: This method constructs multiple positive sample pairs between HSI and LiDAR features for contrastive training, thereby avoiding the sensitivity to negative sample selection inherent in traditional contrastive learning.

FDNet: This method utilizes a fast Fourier transform-based self-attention mechanism to capture local frequency-domain features of spatial, spectral, and elevation information in HSI and LiDAR data, and embeds deformable convolution offsets into the frequency-domain Transformer to enhance local representation and capture global semantics.

MICFNet: This method achieves spatial information interaction through cross-modal attention map fusion and similarity constraints, dynamically adjusts the learning rate of each modality to alleviate learning imbalance, and exchanges semantic pseudo-labels between modalities for mutually guided fusion.

AFDSE: This method adaptively enhances local detail information using dynamic weights and Laplacian convolution kernels, suppresses redundancy and noise through a sparse attention mechanism and sparse gating, and integrates original and frequency-domain features via a gated fusion strategy, thereby achieving deep joint classification of multi-source remote sensing data.

MCFNet: This method captures local structures in the spatial domain through convolutional networks and employs a cross-domain attention mechanism to interactively fuse frequency-domain and spatial-domain features, thereby enhancing the joint classification capability of hyperspectral and LiDAR data.

PICNet: This method introduces learnable modality prototype vectors to perform cross-attention computation with raw features, thereby leveraging global complementary information to enhance cross-modal feature alignment and fusion.

The classification performance of different methods on the MUUFL dataset is shown in Table 2. For clarity, the best value in each row is shown in bold. The results indicate that our method outperforms the competing methods across all evaluation metrics. For example, compared with the best performing baseline model MICFNet, our model improves OA, AA and Kappa by 1.18%, 0.74% and 1.46% respectively. Although HSCC does not achieve the highest classification accuracy in some classes, the accuracy differences between them and the highest values are marginal. Notably, for the challenging categories "Trees" and "Sidewalk", our method achieves accuracies of 95.96% and 79.85% respectively, which are significantly better than the suboptimal methods. The visualized classification results are shown in Figure 7. It can be observed that methods such as AFDSE and PICNet exhibit obvious noise, and most methods misclassify "Trees". In contrast, the classification maps generated by HSCC have less noise and clearer boundaries, benefiting from its effective utilization of multimodal complementary information.

Table 2. Classification performance of different methods on the MUUFL dataset.

Class	IDNet	DSymFuser	mPMCL	FDNet	MICFNet	AFDSE	MCFNet	PICNet	Our
Trees	89.28	90.00	85.32	89.48	93.68	89.73	86.97	88.53	**95.96**
Mostly grass	**84.50**	74.12	77.63	80.92	74.38	78.06	81.54	71.94	79.12
Mixed ground surface	72.64	81.40	75.91	74.25	**83.90**	76.72	74.99	75.07	77.96
Dirt and sand	91.50	92.68	90.93	87.22	**94.65**	94.30	91.33	87.66	85.98
Road	83.46	80.13	73.54	83.28	77.72	76.16	81.56	**83.73**	83.22
Water	99.28	99.52	96.15	99.76	**100**	99.72	99.52	98.56	99.76
Buildings shadow	93.36	**96.84**	83.97	93.63	94.64	91.97	96.61	93.13	91.89
Buildings	90.05	91.62	91.45	93.78	93.15	90.63	92.23	**93.99**	93.72
Sidewalk	64.94	78.43	58.95	76.33	71.84	57.09	72.36	69.59	**79.85**
Yellow curb	89.47	83.46	78.20	**91.73**	86.47	90.27	90.98	79.70	87.97
Cloth panels	98.63	95.87	**99.54**	99.09	96.35	98.92	**99.54**	96.35	**99.54**
OA	85.87	86.75	82.34	86.46	88.40	85.08	85.27	85.29	**89.58**
AA	87.01	87.64	82.87	88.13	87.89	85.78	87.97	85.30	**88.63**
Kappa	81.81	82.85	77.34	82.50	84.82	80.65	81.10	80.98	**86.28**

Bold indicates the best or tied-best result in each row.

The classification performance of different methods on the Houston2013 dataset is presented in Table 3. Experimental results demonstrate that our model achieves the best performance across all evaluation metrics, with an overall accuracy of 97.28%, an average accuracy of 97.61%, and a Kappa coefficient of 97.06%, significantly surpassing other compared methods. Furthermore, our method attains the highest classification accuracy in 7 out of 15 classes. Two main factors account for this superior performance. On one hand, the proposed PIFNet progressively calibrates the semantic representations of dual-modal features through cross-modal shared attention and symmetric cross-attention mechanisms. Its bidirectional symmetric interaction effectively suppresses the dominant bias of a single modality, ensuring that multi-modal information is fully exploited. On the other hand, the SCC strategy imposes both intra-level and cross-level semantic consistency constraints, tightly aligning multi-modal features at multiple abstraction levels and further enhancing the discriminative power of the fused features. Visualized results are shown in Figure 8. Due to the large spatial extent of the Houston2013 dataset, a region delineated by a red box is locally enlarged to illustrate classification details clearly. From the comparison, it is evident that the classification map generated by HSCC is closest to the ground truth. Within the enlarged region, HSCC achieves significantly better recognition results for the "Health grass" and "Road" categories than other methods, whereas the competing models

generally exhibit considerable misclassifications in this area. These observations further validate the effectiveness of our method.

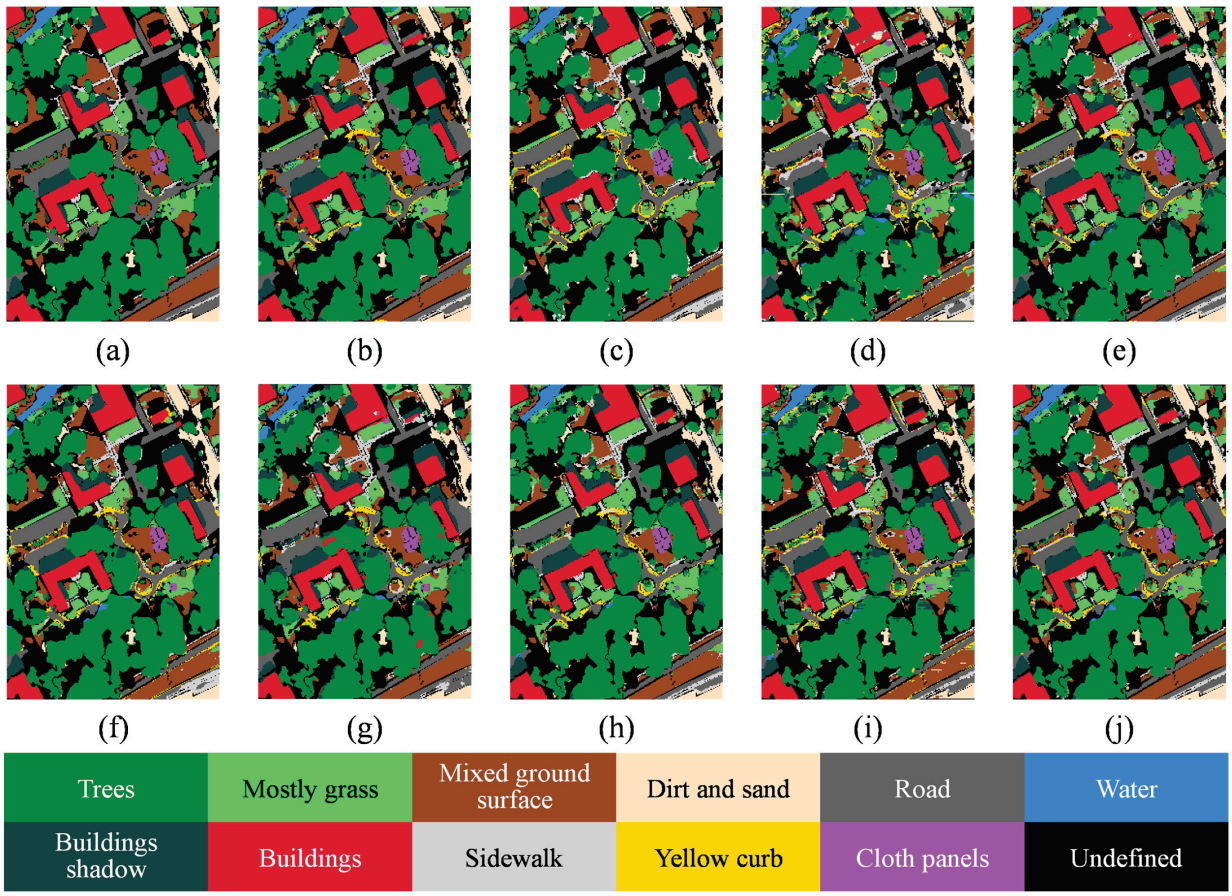


Figure 7. Classification maps generated by different methods on the MUUFL dataset. (**a**) Ground truth. (**b**) IDNet. (**c**) DSymFuser. (**d**) mPMCL. (**e**) FDNet. (**f**) MICFNet. (**g**) AFDSE. (**h**) MCFNet. (**i**) PICNet. (**j**) HSCC (ours).

Table 3. Classification performance of different methods on the Houston2013 dataset.

Class	IDNet	DSymFuser	mPMCL	FDNet	MICFNet	AFDSE	MCFNet	PICNet	Our
Health grass	92.77	80.34	93.01	95.85	92.45	92.83	**98.73**	95.29	98.05
Stressed grass	98.62	**100**	98.54	99.19	98.54	97.24	97.04	98.62	98.38
Synthetic grass	**99.70**	99.60	99.56	98.82	99.56	99.65	99.33	98.23	98.52
Trees	**100**	97.44	96.81	98.61	98.12	98.08	98.26	98.94	98.94
Soil	**100**	98.96	**100**	99.51	**100**	**100**	99.88	99.59	**100**
Water	**100**	**100**	98.69	97.70	98.03	99.61	97.12	94.43	96.72
Residential	95.27	92.54	90.38	95.19	91.51	97.36	96.08	94.71	**98.40**
Commercial	76.06	**92.21**	66.58	87.83	90.03	77.79	83.03	84.74	87.83
Road	90.02	89.90	91.40	85.88	89.61	96.37	90.68	94.24	**98.21**
Highway	97.35	98.07	94.20	99.25	**99.50**	99.12	93.48	92.29	96.11
Railway	**99.84**	95.52	86.17	20.82	98.93	98.35	93.83	98.10	**99.84**
Parking lot 1	94.47	**96.73**	72.22	90.02	91.84	92.63	96.47	75.60	93.16
Parking lot 2	99.55	91.58	98.44	97.77	97.77	**100**	**100**	99.33	**100**
Tennis court	99.51	**100**	99.26	**100**	**100**	**100**	**100**	99.75	**100**
Running track	**100**	93.02	**100**	99.84	**100**	**100**	99.77	**100**	**100**
OA	95.33	94.47	90.69	89.24	95.74	95.80	95.54	94.15	**97.28**
AA	96.21	95.06	92.35	91.09	96.39	96.60	96.25	94.92	**97.61**
Kappa	94.95	94.00	89.95	88.38	95.40	95.46	95.18	93.68	**97.06**

Bold indicates the best or tied-best result in each row.

https://doi.org/10.3390/rs18122058

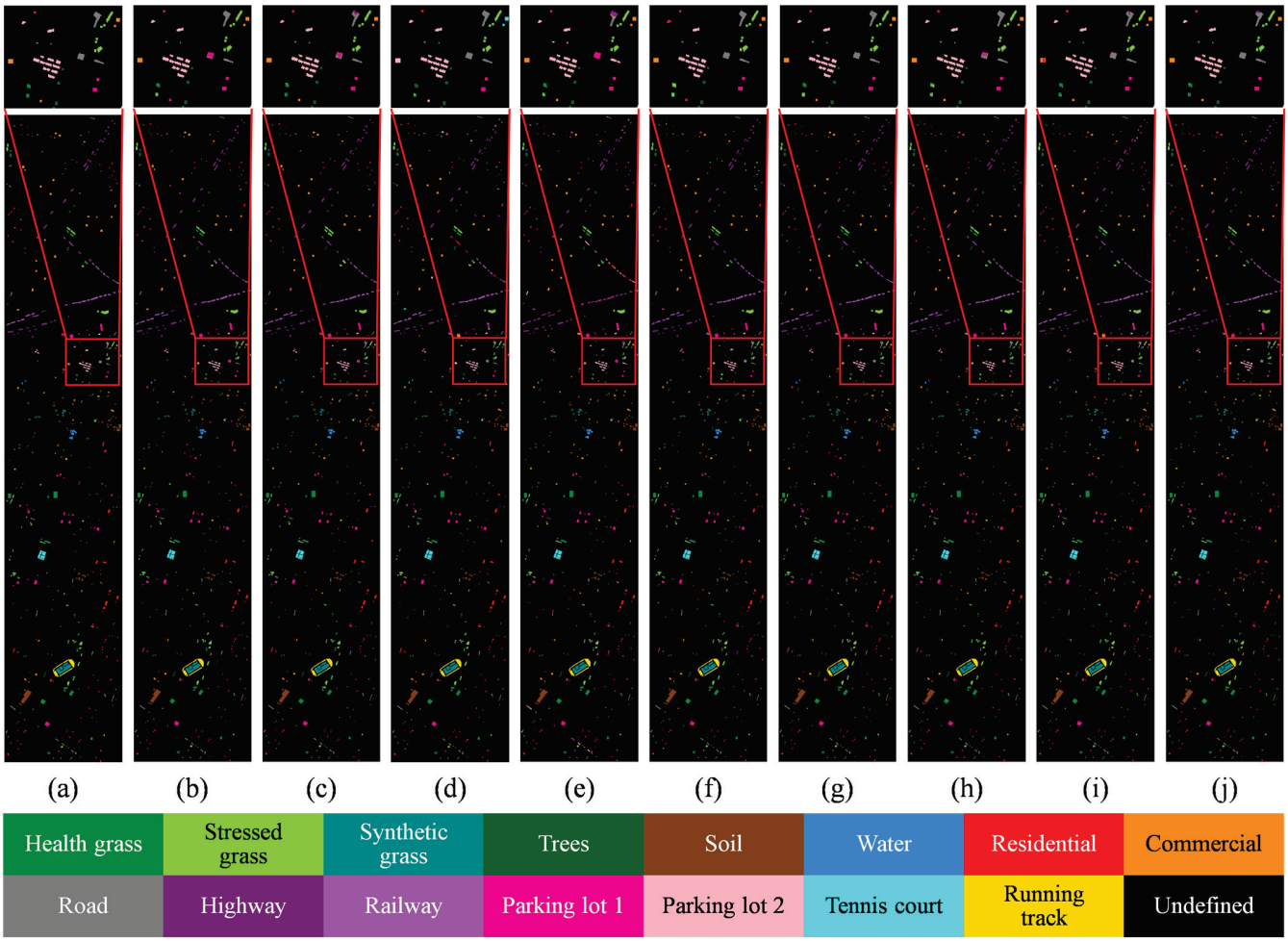


Figure 8. Classification maps generated by different methods on the Houston2013 dataset. (**a**) Ground truth. (**b**) IDNet. (**c**) DSymFuser. (**d**) mPMCL. (**e**) FDNet. (**f**) MICFNet. (**g**) AFDSE. (**h**) MCFNet. (**i**) PICNet. (**j**) HSCC (ours).

The classification performance of different methods on the Augsburg dataset is shown in Table 4. Similar to the previous two datasets, our method achieves the best performance among the three evaluation metrics and attains the highest classification accuracy in three out of all seven classes. Taking AA as an example, HSCC outperforms IDNet, DSymFuser, mPMCL, FDNet, MICFNet, AFDSE, MCFNet, and PICNet by 3.21%, 1.70%, 4.15%, 8.01%, 1.68%, 0.67%, 9.37%, and 8.54%, respectively, representing substantial improvements. Notably, for the two challenging categories, "Industrial area" and "Commercial area", HSCC achieves accuracies of 88.66% and 60.79%, respectively, significantly surpassing the compared methods. Figure 9 presents the classification maps of various methods. Although the quality of the classification maps produced by all methods is acceptable, the prediction results of our method are more accurate. For instance, in the classification results for the "Residential area" category, HSCC yields fewer misclassified pixels and smoother boundaries than other methods.

Table 4. Classification performance of different methods on the Augsburg dataset.

Class	IDNet	DSymFuser	mPMCL	FDNet	MICFNet	AFDSE	MCFNet	PICNet	Our
Forest	99.52	98.58	98.81	**99.64**	97.59	99.02	99.40	96.40	99.18
Residential area	98.59	97.95	96.07	97.89	98.28	97.47	98.55	96.56	**98.79**
Industrial area	84.45	84.03	81.43	78.62	88.31	87.38	76.87	72.98	**88.66**

Table 4. *Cont.*

Class	IDNet	DSymFuser	mPMCL	FDNet	MICFNet	AFDSE	MCFNet	PICNet	Our
Low plants	98.29	98.03	98.34	**98.58**	98.45	97.93	98.35	97.32	98.45
Allotment	72.65	82.34	57.55	55.06	83.66	**84.34**	55.24	51.33	69.63
Commercial area	43.55	44.37	55.52	34.18	38.28	45.11	28.78	57.82	**60.79**
Water	59.37	61.72	62.11	58.84	62.58	62.95	56.10	**69.03**	63.38
OA	95.84	95.48	94.80	95.06	95.67	95.53	94.95	94.27	**96.54**
AA	79.49	81.00	78.55	74.69	81.02	82.03	73.33	74.16	**82.70**
Kappa	94.01	93.52	92.58	92.91	93.78	93.60	92.72	91.79	**95.03**

Bold indicates the best or tied-best result in each row.

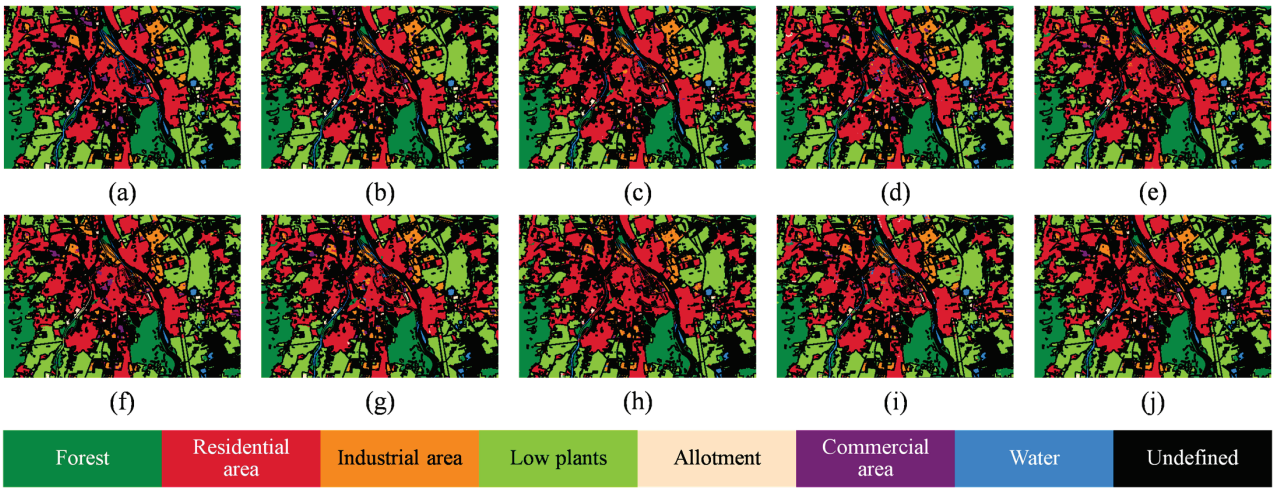


Figure 9. Classification maps generated by different methods on the Augsburg dataset. (**a**) Ground truth. (**b**) IDNet. (**c**) DSymFuser. (**d**) mPMCL. (**e**) FDNet. (**f**) MICFNet. (**g**) AFDSE. (**h**) MCFNet. (**i**) PICNet. (**j**) HSCC (ours).

5. Discussion

5.1. Ablation Analysis of Different Components

To verify the effectiveness of the PIFNet and SCC strategy, we conducted ablation experiments on three datasets, and the results are shown in Table 5. It can be observed that introducing CMSA, SCAT, or SCC individually all improves classification performance, demonstrating that each of the three has its own independent contribution. The complete framework achieves the best performance on all datasets, while removing any component leads to varying degrees of performance degradation, confirming the necessity and synergistic effect of each module.

Table 5. Ablation study results of different components of HSCC on three datasets.

Module			MUUFL			Houston2013			Augsburg		
CMSA	SCAT	SCC	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
×	×	×	87.35	80.14	83.46	94.76	95.33	94.33	95.44	76.95	93.56
√	×	×	87.52	86.99	83.70	94.82	95.37	94.40	95.54	79.16	93.60
×	√	×	88.41	88.37	84.93	96.52	97.14	96.19	96.05	81.45	94.09
×	×	√	88.06	87.12	84.40	96.48	97.04	96.19	95.64	82.06	94.19
×	√	√	88.12	87.94	84.53	96.20	96.64	95.90	96.08	81.78	94.36
√	×	√	88.08	87.72	84.47	95.87	96.53	95.54	96.25	80.18	94.72
√	√	×	88.58	88.07	85.54	96.61	97.13	96.33	95.98	81.64	94.23
√	√	√	**89.58**	**88.63**	**86.28**	**97.28**	**97.61**	**97.06**	**96.54**	**82.70**	**95.03**

Bold indicates the best or tied-best result in each column.

Specifically, when only CMSA is introduced, the OA on the MUUFL, Houston2013, and Augsburg datasets improves by 0.17%, 0.04%, and 0.10%, respectively, compared to the baseline. In contrast, when the CMSA module is removed, the OA on the three datasets decreases by 1.46%, 1.08%, and 0.46%, respectively, indicating that the CMSA module plays a significant role in establishing cross-modal associations at shallow layers and guiding the network to focus on jointly discriminative regions. Its absence causes subsequent semantic alignment to lack a reliable spatial prior. When only SCAT is introduced, the OA on the three datasets improves by 1.06%, 1.76%, and 0.61%, respectively. When standard cross-attention is used in place of the SCAT module, the OA drops by 1.50%, 1.41%, and 0.29%, respectively, demonstrating that the balanced bidirectional interaction achieved by SCAT is crucial for deep fusion, providing a more equitable feature foundation for cross-modal semantic consistency constraints. Notably, when both CMSA and SCAT are introduced together, the OA is significantly superior to using either module alone. For example, on the Augsburg dataset, it improves by 1.00% and 0.49% compared to using CMSA or SCAT individually, respectively. This indicates that shallow spatial guidance and deep bidirectional interaction can effectively collaborate, integrating cross-modal information through multi-stage progressive alignment, thereby enhancing the expressive power of fused features for classification. When only the SCC strategy is introduced, the OA improves by 0.71%, 1.72%, and 0.50%, respectively. After removing the SCC strategy, the OA decreases by 1.00%, 0.67%, and 0.56%, respectively, with similar trends observed for AA and Kappa, confirming that SCC, by explicitly constraining the semantic similarity between cross-level and cross-modal features, can significantly reduce representation discrepancies caused by heterogeneity, thus yielding higher-quality fused representations.

In summary, PIFNet progressively narrows modality differences through the spatial guidance provided by CMSA and the deep balanced interaction ensured by SCAT, while the SCC strategy further enhances feature consistency via semantic consistency constraints. Together, they synergistically support the superior performance of the HSCC framework.

5.2. Analysis of SCC Strategy

To systematically analyze the impact of different semantic associations on the combination, in addition to the full SCC strategy, we also test two other possible semantic association strategies, specifically: (1) Strategy a: constructing only same-level semantic association pairs; (2) Strategy b: constructing only cross-level cross-modal association pairs; (3) Strategy c: the proposed SCC strategy. Table 6 presents the experimental results of different semantic consistency constraint strategies. Strategy a achieves relatively stable performance on all three datasets, but is overall inferior to Strategy c. Strategy b yields slightly better AA than Strategy c on MUUFL, but its AA drops significantly on Houston2013 and Augsburg, indicating that while cross-level cross-modal matching is important, relying solely on cross-level associations is insufficient to handle semantic shifts in complex scenes. In contrast, the proposed SCC strategy constructs a complete set of multi-level semantic correspondences, which can more fully enhance the semantic consistency and classification reliability of the fused features.

Table 6. Ablation study results of different strategies on the three datasets.

Strategy	MUUFL			Houston2013			Augsburg		
	OA	**AA**	**Kappa**	**OA**	**AA**	**Kappa**	**OA**	**AA**	**Kappa**
a	88.27	88.47	84.76	96.78	97.19	96.51	96.36	82.67	94.78
b	88.50	**88.82**	85.01	96.12	96.72	95.81	96.31	79.50	94.70
c	**89.58**	88.63	**86.28**	**97.28**	**97.61**	**97.06**	**96.54**	**82.70**	**95.03**

Bold indicates the best or tied-best result in each column.

https://doi.org/10.3390/rs18122058

5.3. Parameter Analysis

5.3.1. Impact of Learning Rate on Classification Performance

Considering the intrinsic differences in spectral details and geometric structures between HSI and LiDAR data, an appropriate learning rate can effectively coordinate the feature-learning process. Experiments were conducted on the candidate set $\{1 \times 10^{-5}$, 5×10^{-5}, 1×10^{-4}, 5×10^{-4}, 1×10^{-3}, $5 \times 10^{-3}\}$, and the variation in OA with learning rate on the three datasets is shown in Figure 10. The experimental results indicate that when the learning rate is too low, the model converges slowly, resulting in low OA values across all datasets. Conversely, an excessively high learning rate tends to cause training oscillation and leads to a significant performance decline. For the MUUFL dataset, the optimal learning rate is 1×10^{-3}, while the Houston2013 and Augsburg datasets achieve their best performance at 5×10^{-4}.

5.3.2. Impact of Patch Size on Classification Performance

The input patch size determines the range of local context that the model can perceive and is crucial for feature extraction from HSI and LiDAR data. To investigate its impact, we tested six patch sizes of 7, 9, 11, 13, 15 and 17 on the three datasets, with the results shown in Figure 11. The experimental results indicate that the selection of patch size requires a trade-off between local details and global context. A patch size that is too small fails to provide sufficient neighborhood information, preventing the model from adequately capturing the spatial structure and spectral context of land cover objects. Conversely, an excessively large patch size tends to introduce redundant information, interfering with the discrimination of local details and causing performance degradation. The optimal patch size varies across the three datasets. The MUUFL dataset achieves its best performance at $p = 9$, while the Houston2013 and Augsburg datasets perform best at $p = 11$ and $p = 13$, respectively. This is because the Houston2013 dataset features a more complex scene that requires a slightly larger window to incorporate more contextual information, thereby enabling effective differentiation among different land cover classes.

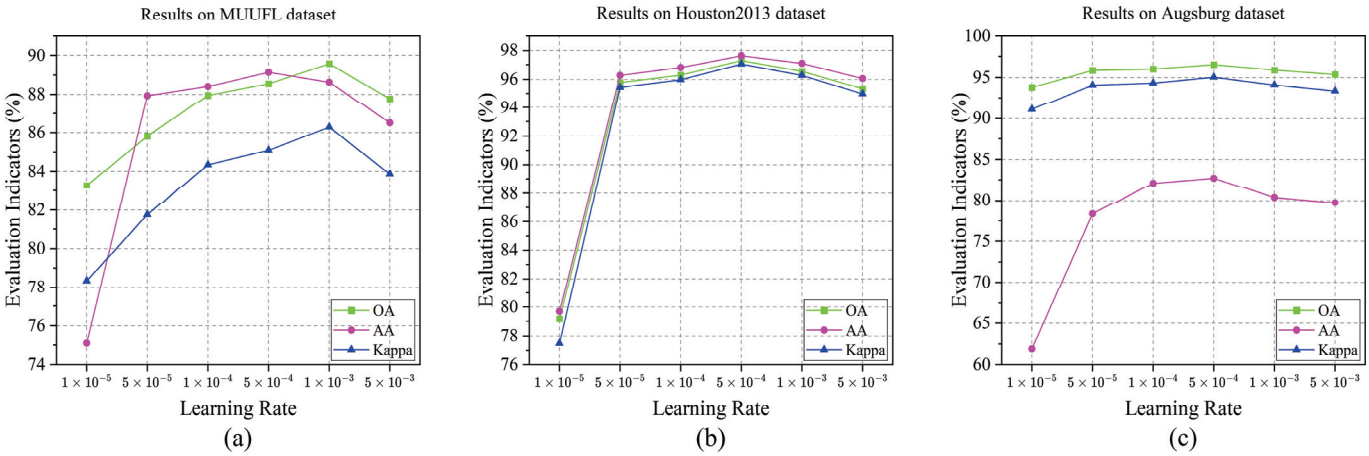


Figure 10. Effect of different learning rates on classification performance. (**a**) MUUFL. (**b**) Houston2013. (**c**) Augsburg.

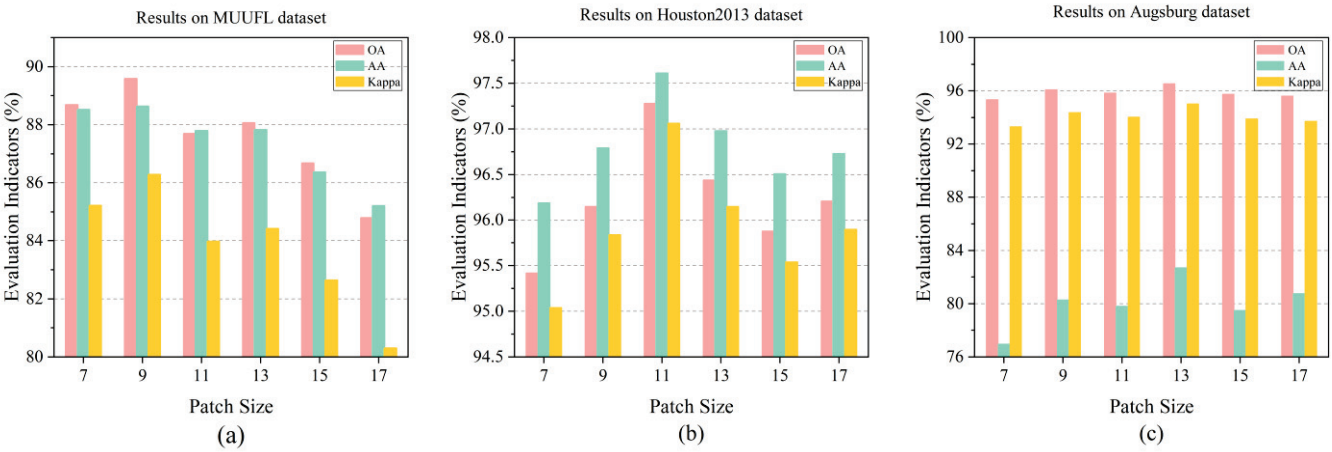


Figure 11. Effect of different patch sizes on classification performance. (**a**) MUUFL. (**b**) Houston2013. (**c**) Augsburg.

5.3.3. Impact of Balance Coefficient μ on Classification Performance

The balance coefficient μ controls the contribution of the semantic consistency constraint loss L_{SCC} to the total loss. When μ is too small, the semantic consistency constraint contributes insufficiently, making it difficult to effectively suppress cross-modal semantic drift. When μ is too large, the dominant role of the classification loss is weakened, and the model may focus excessively on feature alignment at the expense of discriminative ability. The value range of μ was set to {0.1, 0.2, 0.3, 0.4, 0.5}. For each dataset, the model was trained with each candidate value and the resulting OA values were recorded, as shown in Figure 12. The MUUFL dataset achieves the highest OA at $\mu = 0.2$, while the optimal μ for the Houston2013 and Augsburg datasets are 0.3 and 0.1, respectively. This can be attributed to the greater diversity of classes in the Houston2013 dataset, which requires a stronger constraint to guide cross-modal consistency.

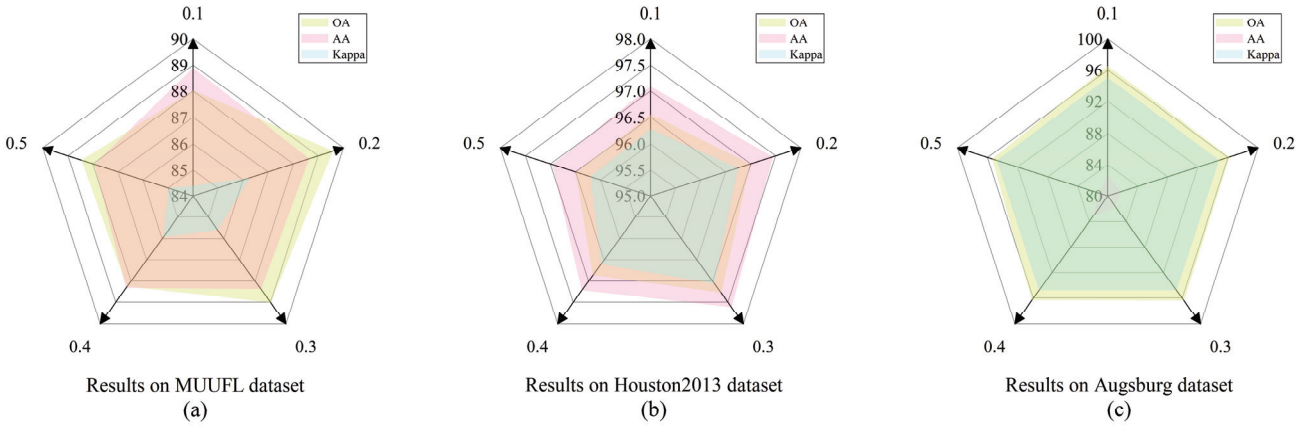


Figure 12. Effect of balance coefficient μ on classification performance. (**a**) MUUFL. (**b**) Houston2013. (**c**) Augsburg.

5.4. Feature Visualization

To further validate the effectiveness of the HSCC framework, we employed the t-distributed stochastic neighbor embedding (t-SNE) algorithm to visualize the learned features and compared them with the best-performing baseline model, MICFNet. Figure 13 shows the feature distributions of the two models on the three datasets, where the first row presents the results of MICFNet and the second row those of HSCC. Overall, HSCC produces clearer class boundaries and higher intra-class compactness. For example, for class

7 (Buildings shadow) of the MUUFL dataset and class 12 (Park lot 1) of the Houston2013 dataset, the feature clusters of HSCC are more compact and well separated from other classes. This is mainly attributed to the SCC strategy, which effectively suppresses cross-modal semantic shift by explicitly pulling closer the semantic similarity of cross-modal and cross-level features. The above visualization results confirm the effectiveness of HSCC in cross-modal feature learning, facilitating more accurate land cover classification.

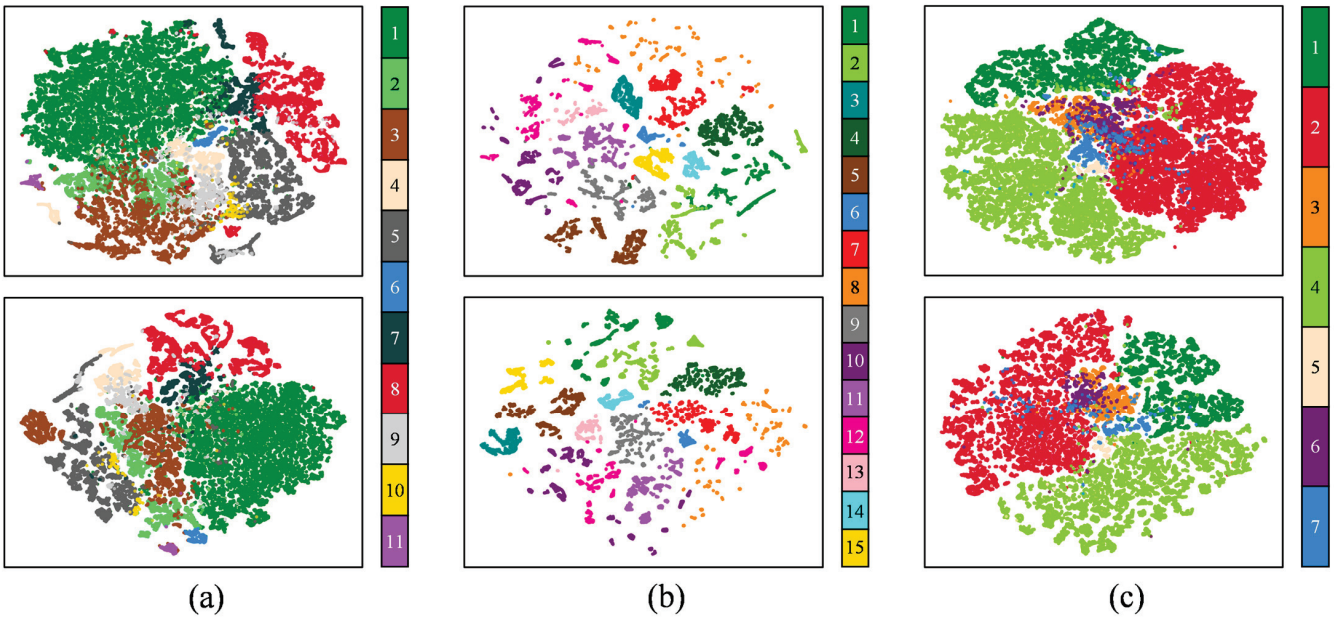


Figure 13. Feature visualization based on the t-SNE algorithm. The first row shows the results of MICFNet, and the second row shows the results of HSCC. (**a**) MUUFL. (**b**) Houston2013. (**c**) Augsburg.

5.5. Computational Complexity Analysis

To comprehensively evaluate the trade-off between performance and computational cost for each method, we report the OA, number of parameters, and floating point operations (FLOPs) on the Houston2013 dataset. The results are shown in Table 7. HSCC achieves the highest OA, significantly outperforming all compared methods. Although mPMCL has the lowest FLOPs and a relatively small number of parameters, its OA is only 90.69%, indicating that excessive computational compression impairs the feature representation capability of the model. In contrast, PICNet, despite its higher computational cost, still yields a lower OA than HSCC, suggesting that its utilization efficiency of computational resources can be further improved. By comparison, HSCC achieves the best accuracy with moderate parameters and FLOPs, validating its feasibility as an efficient fusion framework.

Table 7. Comparison of performance and complexity of different methods on the Houston2013 dataset.

Methods	IDNet	DSymFuser	mPMCL	FDNet	MICFNet	AFDSE	MCFNet	PICNet	Our
OA	95.33	94.47	90.69	89.24	95.74	95.80	95.54	94.15	**97.28**
Parameters	**111.79 K**	848.66 K	172.67 K	345.86 K	167.68 K	441.84 K	252.31 K	27.12 M	425.44 K
FLOPs	21.35 M	112.01 M	**7.59 M**	11.68 M	20.09 M	48.86 M	39.06 M	1.23 G	35.88 M

Bold indicates the best or tied-best result in each column.

6. Conclusions

Aiming at the problems of cross-modal semantic inconsistency and imbalanced modality contribution in HSI and LiDAR joint classification, this paper proposes a hierarchical semantic consistency constraint framework, HSCC. The framework is co-constructed by

PIFNet and the SCC strategy. PIFNet progressively calibrates the semantic representations of multimodal features at different abstraction levels through CMSA and SCAT module to mitigate the information imbalance in deep interactions; the SCC strategy establishes both same-level and cross-level semantic associations to guide the network to autonomously maintain consistency between heterogeneous feature representations. A series of experimental results demonstrate that HSCC achieves superior performance in multi-source remote sensing data fusion classification tasks, validating its effectiveness.

Nevertheless, HSCC still has certain limitations. First, the training of the framework relies on a certain amount of high-quality labeled data; its performance may be constrained in scenarios where annotations are scarce or expensive to obtain. Second, although the introduced multi-attention mechanism helps improve classification accuracy, it also increases the training complexity of the model, requiring further optimization in resource-constrained scenarios. In future work, we will leverage self-supervised or semi-supervised strategies to mine effective supervisory signals from unlabeled data, thereby enhancing the model's generalization ability under few-shot conditions. Meanwhile, we will further explore lightweight network architectures to effectively reduce the computational cost of the model while maintaining performance. In addition, we will consider introducing an uncertainty estimation mechanism to characterize the inherent ambiguity of land cover classes in complex scenes, providing interpretable confidence information for classification decisions.

Author Contributions: Conceptualization, H.Y. and J.S.; methodology, J.S.; software, Y.M.; validation, H.Y., J.S. and Y.M.; formal analysis, J.S.; investigation, J.S. and Y.M.; resources, J.S. and Y.M.; data curation, H.Y.; writing—original draft preparation, J.S.; writing—review and editing, H.Y. and J.S.; visualization, J.S. and Y.M.; supervision, H.Y.; project administration, H.Y.; funding acquisition, H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Hainan Province Science and Technology Special Fund under Grant No. ZDYF2025GXJS184 and the Haikou Science and Technology Plan Project under Grant No. 2022-007.

Data Availability Statement: The original research results presented in this study are all included in this article. If you have any further questions, please feel free to contact the corresponding author.

Acknowledgments: We gratefully appreciate the editor and anonymous reviewers for their efforts and constructive comments, which have greatly improved the technical quality and presentation of this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Samadzadegan, F.; Toosi, A.; Dadrass Javan, F. A Critical Review on Multi-Sensor and Multi-Platform Remote Sensing Data Fusion Approaches: Current Status and Prospects. *Int. J. Remote Sens.* **2025**, *46*, 1327–1402. [CrossRef]
2. Li, Y.; Xiao, X. Deep Learning-Based Fusion of Optical, Radar, and LiDAR Data for Advancing Land Monitoring. *Sensors* **2025**, *25*, 4991. [CrossRef] [PubMed]
3. Wang, F.; Du, X.; Zhang, W.; Nie, L.; Wang, H.; Zhou, S.; Ma, J. Remote Sensing LiDAR and Hyperspectral Classification with Multi-Scale Graph Encoder–Decoder Network. *Remote Sens.* **2024**, *16*, 3912. [CrossRef]
4. Zhao, W.; Du, S. Spectral-Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4544–4554. [CrossRef]
5. Huo, Y.; Dong, Y.; Wang, C.; Zhang, M.; Wang, H. Multi-Scale Memory Network with Separation Training for Hyperspectral Anomaly Detection. *Inf. Process. Manag.* **2026**, *63*, 104494. [CrossRef]
6. Cao, H.; Meng, F.; Sun, H.; Cui, X.; Shao, D. A Theory-Guided Transformer for Interpretable Hyperspectral Unmixing. *Remote Sens.* **2026**, *18*, 886. [CrossRef]
7. Ghamisi, P.; Yokoya, N.; Li, J.; Liao, W.; Liu, S.; Plaza, J. Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 37–78. [CrossRef]

8. Xu, Y.; Zhang, L.; Du, B.; Zhang, F. Spectral-Spatial Unified Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5893–5909. [CrossRef]
9. Chi, B.; Lu, H.; Liu, R.; Yang, Y.; Xu, L.; Wan, W. Multispectral-Hyperspectral Image Fusion via Similarity-Guided Graph Attention and VAE-Transformer. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5515416. [CrossRef]
10. Song, T.; Zeng, Z.; Gao, C.; Chen, H.; Li, J. Joint Classification of Hyperspectral and LiDAR Data Using Height Information Guided Hierarchical Fusion-and-Separation Network. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5505315. [CrossRef]
11. Zhang, Y.; Gao, H.; Zhou, J.; Zhang, C.; Ghamisi, P.; Xu, S.; Li, C.; Zhang, B. A Cross-Modal Feature Aggregation and Enhancement Network for Hyperspectral and LiDAR Joint Classification. *Expert Syst. Appl.* **2024**, *258*, 125145. [CrossRef]
12. Mukundan, A.; Karmakar, R.; Jouhar, J.; Valappil, M.A.E.; Wang, H.-C. Advancing Urban Development: Applications of Hyperspectral Imaging in Smart City Innovations and Sustainable Solutions. *Smart Cities* **2025**, *8*, 51. [CrossRef]
13. Zhang, H.; Meiller, C.; Hueni, A.; Blackman, R.C.; Morsdorf, F.; Helfenstein, I.S.; Schaepman, M.E.; Altermatt, F. Hyperspectral Imagery, LiDAR Point Clouds, and Environmental DNA to Assess Land-Water Linkage of Biodiversity Across Aquatic Functional Feeding Groups. *Remote Sens. Ecol. Conserv.* **2025**, *11*, 618–635. [CrossRef]
14. Saif, M.S.; Chancia, R.; Murphy, S.P.; Pethybridge, S.; van Aardt, J. Exploring UAS Imaging Modalities for Precision Agriculture: Predicting Table Beet Root Yield and Estimating Disease Severity Using Multispectral, Hyperspectral, and LiDAR Sensing. In Proceedings of the SPIE: Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXXI, Orlando, FL, USA, 15–17 April 2025; Volume 13455, p. 1345509. [CrossRef]
15. Kim, N.S.; Lim, C.H. Developing a Novel Method for Vegetation Mapping in Temperate Forests Using Airborne LiDAR and Hyperspectral Imaging. *Forests* **2025**, *16*, 1158. [CrossRef]
16. Ding, S.; Chen, L. Classification of Hyperspectral Remote Sensing Images with Support Vector Machines and Particle Swarm Optimization. In Proceedings of the 2009 International Conference on Information Engineering and Computer Science, Wuhan, China, 19–20 December 2009; pp. 1–5. [CrossRef]
17. Du, P.; Xia, J.; Chanussot, J.; He, X. Hyperspectral Remote Sensing Image Classification Based on the Integration of Support Vector Machine and Random Forest. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 174–177. [CrossRef]
18. Huang, X.; Han, X.; Zhang, L.; Gong, J.; Liao, W.; Benediktsson, J.A. Generalized Differential Morphological Profiles for Remote Sensing Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1736–1751. [CrossRef]
19. Dalla Mura, M.; Atli Benediktsson, J.; Waske, B.; Bruzzone, L. Extended Profiles with Morphological Attribute Filters for the Analysis of Hyperspectral Data. *Int. J. Remote Sens.* **2010**, *31*, 5975–5991. [CrossRef]
20. Kahraman, S.; Bacher, R. A Comprehensive Review of Hyperspectral Data Fusion with LiDAR and SAR Data. *Annu. Rev. Control* **2021**, *51*, 236–253. [CrossRef]
21. Wang, B.; Liu, J.; Li, J.; Li, M. UAV LiDAR and Hyperspectral Data Synergy for Tree Species Classification in the Maoershan Forest Farm Region. *Remote Sens.* **2023**, *15*, 1000. [CrossRef]
22. Roy, S.K.; Deria, A.; Hong, D.; Ahmad, M.; Plaza, A.; Chanussot, J. Hyperspectral and LiDAR Data Classification Using Joint CNNs and Morphological Feature Learning. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5530416. [CrossRef]
23. Feng, Q.; Zhu, D.; Yang, J.; Li, B. Multisource Hyperspectral and LiDAR Data Fusion for Urban Land-Use Mapping Based on a Modified Two-Branch Convolutional Neural Network. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 28. [CrossRef]
24. Zhang, M.; Li, W.; Du, Q.; Gao, L.; Zhang, B. Feature Extraction for Classification of Hyperspectral and LiDAR Data Using Patch-to-Patch CNN. *IEEE Trans. Cybern.* **2020**, *50*, 100–111. [CrossRef] [PubMed]
25. Li, H.; Ghamisi, P.; Soergel, U.; Zhu, X.X. Hyperspectral and LiDAR Fusion Using Deep Three-Stream Convolutional Neural Networks. *Remote Sens.* **2018**, *10*, 1649. [CrossRef]
26. Xu, X.; Li, W.; Ran, Q.; Du, Q.; Gao, L.; Zhang, B. Multisource Remote Sensing Data Classification Based on Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 937–949. [CrossRef]
27. Li, Z.; Sui, H.; Luo, C.; Guo, F. Morphological Convolution and Attention Calibration Network for Hyperspectral and LiDAR Data Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 5728–5740. [CrossRef]
28. Wang, R.; Ma, L.; He, G.; Johnson, B.A.; Yan, Z.; Chang, M.; Liang, Y. Transformers for Remote Sensing: A Systematic Review and Analysis. *Sensors* **2024**, *24*, 3495. [CrossRef] [PubMed]
29. Roy, S.K.; Deria, A.; Hong, D.; Rasti, B.; Plaza, A.; Chanussot, J. Multimodal Fusion Transformer for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5515620. [CrossRef]
30. Feng, Y.; Zhu, J.; Song, R.; Wang, X. S2EFT: Spectral-Spatial-Elevation Fusion Transformer for Hyperspectral Image and LiDAR Classification. *Knowl.-Based Syst.* **2024**, *283*, 111190. [CrossRef]
31. Yang, J.X.; Wang, J.; Li, Z.; Long, C.S.Z.; Zhou, J. HSLiNets: Evaluating Band Ordering Strategies in Hyperspectral and LiDAR Fusion. *arXiv* **2025**, arXiv:2503.21072. [CrossRef]
32. Li, K.; Wang, D.; Wang, X.; Liu, G.; Wu, Z.; Wang, Q. Mixing Self-Attention and Convolution: A Unified Framework for Multisource Remote Sensing Data Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5523216. [CrossRef]

33. Seneviratne, G.; Weerakoon, K.; Elnoor, M.; Rajgopal, V.; Varatharajan, H.; Jaffar, M.K.M.; Pusey, J.; Manocha, D. CROSS-GAiT: Cross-Attention-Based Multimodal Representation Fusion for Parametric Gait Adaptation in Complex Terrains. In Proceedings of the 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hangzhou, China, 19–25 October 2025; pp. 6079–6086. [CrossRef]
34. Zhao, Y.; Bao, W.; Xu, J.; Xu, X. BIHAF-Net: Bilateral Interactive Hierarchical Adaptive Fusion Network for Collaborative Classification of Hyperspectral and LiDAR Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 15971–15988. [CrossRef]
35. Wang, S.; Hou, C.; Chen, Y.; Liu, Z.; Zhang, Z.; Zhang, G. Classification of Hyperspectral and LiDAR Data Using Multi-Modal Transformer Cascaded Fusion Net. *Remote Sens.* **2023**, *15*, 4142. [CrossRef]
36. Ni, K.; Wang, D.; Zheng, Z.; Wang, P. MHST: Multiscale Head Selection Transformer for Hyperspectral and LiDAR Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 5470–5483. [CrossRef]
37. Song, L.; Feng, Z.; Yang, S.; Zhang, X.; Jiao, L. Discrepant Bi-Directional Interaction Fusion Network for Hyperspectral and LiDAR Data Classification. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 5510605. [CrossRef]
38. Hussain, K.M.; Zhao, K.; Zhou, Y.; Ali, A.; Li, Y. Cross Attention Based Dual-Modality Collaboration for Hyperspectral Image and LiDAR Data Classification. *Remote Sens.* **2025**, *17*, 2836. [CrossRef]
39. Hussain, K.M.; Zhao, K.; Pervaiz, S.; Li, Y. Global-Local Mamba-Based Dual-Modality Fusion for Hyperspectral and LiDAR Data Classification. *Remote Sens.* **2026**, *18*, 138. [CrossRef]
40. Yang, J.X.; Zhou, J.; Wang, J.; Tian, H.; Liew, A.W.-C. LiDAR-Guided Cross-Attention Fusion for Hyperspectral Band Selection and Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5515815. [CrossRef]
41. Hui, L.; Xiao-Jun, W. CrossFuse: A Novel Cross Attention Mechanism Based Infrared and Visible Image Fusion Approach. *Inf. Fusion* **2024**, *103*, 102147. [CrossRef]
42. Shen, J.; Chen, Y.; Liu, Y.; Zuo, X.; Fan, H.; Yang, W. ICAFusion: Iterative Cross-Attention Guided Feature Fusion for Multispectral Object Detection. *Pattern Recognit.* **2024**, *145*, 109913. [CrossRef]
43. Sun, L.; Wang, X.; Zheng, Y.; Wu, Z.; Fu, L. Multiscale 3-D–2-D Mixed CNN and Lightweight Attention-Free Transformer for Hyperspectral and LiDAR Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 2100116. [CrossRef]
44. Wang, M.; Sun, Y.; Xiang, J.; Zhong, Y. CITNet: Convolution Interaction Transformer Network for Hyperspectral and LiDAR Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5535918. [CrossRef]
45. Bai, J.; Chen, N.; Peng, J.; Wu, L.; Sun, W.; Ye, Z. HCAFNet: Hierarchical Cross-Modal Attention Fusion Network for HSI and LiDAR Joint Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 9522–9532. [CrossRef]
46. Zhang, G.; Zhang, L.; Zhang, Z.; Deng, J.; Bian, L.; Yang, C. DECT: Diffusion-Enhanced CNN–Transformer for Multisource Remote Sensing Data Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 19288–19301. [CrossRef]
47. Wang, W.; Li, C.; Ren, P.; Lu, X.; Wang, J.; Ren, G.; Liu, B. Dual-Branch Feature Fusion Network Based Cross-Modal Enhanced CNN and Transformer for Hyperspectral and LiDAR Classification. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 5503605. [CrossRef]
48. Ge, H.; Wang, L.; Pan, H.; Liu, Y.; Li, C.; Lv, D.; Ma, H. Cross Attention-Based Multi-Scale Convolutional Fusion Network for Hyperspectral and LiDAR Joint Classification. *Remote Sens.* **2024**, *16*, 4073. [CrossRef]
49. Liu, S.; Yu, T.; Zhou, J.; Xing, G.; Chen, H. Cross-Source Transformer-Based Neighborhood Contrastive Learning for Joint Classification of Hyperspectral and LiDAR Data. *Inf. Fusion* **2025**, *124*, 103225. [CrossRef]
50. Guo, H.; Tian, B.; Liu, W. CCFormer: Cross-Modal Cross-Attention Transformer for Classification of Hyperspectral and LiDAR Data. *Sensors* **2025**, *25*, 5698. [CrossRef] [PubMed]
51. Li, Z.; Liu, R.; Sun, L.; Zheng, Y. Multi-Feature Cross Attention-Induced Transformer Network for Hyperspectral and LiDAR Data Classification. *Remote Sens.* **2024**, *16*, 2775. [CrossRef]
52. Zhang, Y.; Gao, H.; Chen, Z.; Fei, S.; Zhou, J.; Ghamisi, P.; Zhang, B. Adaptive Multi-Stage Fusion of Hyperspectral and LiDAR Data via Selective State Space Models. *Inf. Fusion* **2026**, *125*, 103488. [CrossRef]
53. Jiao, C.; Wang, L.; Hu, C.; Tang, X.; Zhu, H.; Jiao, L. Dynamic Common and Unique Feature Fusion Network for Hyperspectral and LiDAR Data Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5524013. [CrossRef]
54. Liang, C.; Zhao, Y.; Song, Y.; Ni, K. IDNet: Intensity-Constrained Detail-Enhanced Network for Hyperspectral and LiDAR Collaborative Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5518515. [CrossRef]
55. Chang, H.; Bi, H.; Li, F.; Xu, C.; Chanussot, J.; Hong, D. Deep Symmetric Fusion Transformer for Multimodal Remote Sensing Data Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5644115. [CrossRef]
56. Liu, H.; Huang, C.; Xie, T.; Bao, W.; Chen, N.; Yue, J.; Fang, L. Positive Matching Benefits Fusion: A Novel Contrastive Learning Framework for Hyperspectral and LiDAR Data Classification. *IEEE Trans. Geosci. Remote Sens.* **2026**, *64*, 5502218. [CrossRef]
57. Ni, K.; Wang, D.; Zhao, G.; Zheng, Z.; Wang, P. Hyperspectral and LiDAR Classification via Frequency Domain-Based Network. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5525117. [CrossRef]
58. Tang, X.; Zou, Y.; Ma, J.; Zhang, X.; Liu, F.; Jiao, L. Multiple Information Collaborative Fusion Network for Joint Classification of Hyperspectral and LiDAR Data. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5525416. [CrossRef]

59. Wei, X.; Tu, B.; Liu, B.; Li, J.; Plaza, A. Multimodal Data Fusion Classification via Adaptive Frequency Domain Sparse Enhancement. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5525716. [CrossRef]
60. Song, Q.; Mo, F.; Ding, K.; Xiao, L.; Dian, R.; Kang, X.; Li, S. MCFNet: Multiscale Cross-Domain Fusion Network for HSI and LiDAR Data Joint Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4703912. [CrossRef]
61. Gao, F.; Liu, S.; Gong, C.; Zhou, X.; Wang, J.; Dong, J.; Du, Q. Prototype-Based Information Compensation Network for Multisource Remote Sensing Data Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5513615. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

remote sensing

MDPI

Review

Scale-Aware Interpretation of Vegetation Traits and SIF-Based Dynamics in Earth Observation

Jochem Verrelst *, Bhagyashree Verma and Pablo Reyes-Muñoz

Image Processing Laboratory (IPL), University of Valencia, Catedrático Agustín Scardino Benlloch 9, 46980 Paterna, Spain; bhagyashree.verma@uv.es (B.V.); pablo.reyes@uv.es (P.R.-M.)
* Correspondence: jochem.verrelst@uv.es

Highlights

What are the key insights from this review?

- Vegetation Earth observation is an inherently multiscale retrieval problem, with the effective scale determined by the interplay of observation, retrieval, and aggregation processes.
- Increasing spatial, spectral, or temporal resolution does not necessarily improve ecological accuracy (resolution–accuracy paradox), due to non-commutativity and scale-dependent error propagation.

What are the implications emerging from this review?

- Vegetation products must be interpreted and evaluated at their effective scale, using aggregation-consistent and cross-scale diagnostics rather than single-resolution agreement.
- Robust use of current and future hyperspectral missions (e.g., FLEX, CHIME) requires scale-aware and uncertainty-aware modelling frameworks that explicitly account for heterogeneity, nonlinearity, and temporal sampling.

Abstract

Satellite-based vegetation monitoring has evolved from mapping vegetation canopy properties at single points in time toward analysing time-resolved dynamics of vegetation traits and process-related variables. Retrieved traits and solar-induced chlorophyll fluorescence (SIF) are inherently defined by sensor-specific spatial resolution, temporal integration, and spectral response. Modifying these characteristics alters the retrieval problem itself: under nonlinear retrievals and heterogeneous landscapes, aggregation and retrieval are generally non-commutative, and error components scale differently with resolution. Consequently, increasing spatial, spectral, or temporal detail does not guarantee improved ecological accuracy; a phenomenon we term the resolution–accuracy paradox. These interacting processes define the effective scale of vegetation products, which may differ from nominal sensor resolution and governs the interpretation of retrieved vegetation traits. When products with differing resolutions or compositing strategies are combined, scale effects can induce systematic artefacts in spatial patterns and derived dynamic indicators that cannot be resolved through improved per-pixel accuracy alone. This review establishes a scale-aware conceptual framework that treats scale as an explicit diagnostic dimension linking observation characteristics, retrieval formulations, trait definitions, and aggregation operators. We analyse how scale interactions influence spatial patterns, temporal dynamics, disturbance signals, and multiresolution data fusion, and derive diagnostic principles, best-practice guidelines, and research priorities for the scale-consistent interpretation of vegetation trait dynamics and SIF-constrained productivity and stress indicators across spatial and temporal scales. Framed in the context of upcoming hyperspectral missions such as CHIME and FLEX, which increase spectral information content, robust interpretation of

vegetation products requires scale-consistent analysis and uncertainty-aware processing. For practitioners, this implies that vegetation products should be interpreted, validated, and compared at their effective scale rather than assuming that a finer spatial, spectral, or temporal resolution necessarily yields more reliable ecological information.

Keywords: vegetation traits; scale effects; temporal dynamics; solar-induced fluorescence; gross primary productivity; aggregation consistency; hyperspectral remote sensing; CHIME; FLEX

1. Introduction

Satellite observations play a central role in monitoring vegetation structure, function, and dynamics at regional to global scales (see reviews, [1,2]). Advances in Earth observation (EO) have shifted vegetation remote sensing from mapping static canopy states toward analysing time-resolved dynamics of vegetation traits and process-relevant variables (see reviews, [3–5]). Reflectance-based retrieval enables routine estimation of canopy biophysical and biochemical properties such as the leaf area index (LAI), the fraction of Absorbed Photosynthetically Active Radiation (FAPAR), and pigment- and water-related traits, supporting regional to global monitoring and modeling (see reviews, [6–8]). In parallel, solar-induced chlorophyll fluorescence (SIF) has emerged as a complementary constraint on photosynthetic regulation and stress, motivating the retrieval and modelling of higher-level variables such as Gross Primary Productivity (GPP) and vegetation stress (e.g., [4,9–11]).

At present, the currently available satellite SIF products lack the spatial and spectral resolution expected from the upcoming ESA Fluorescence Explorer (FLEX) mission [12,13], as existing spaceborne SIF datasets are largely derived from sensors designed for atmospheric composition monitoring rather than dedicated SIF observation (e.g., [4,14,15]). Forthcoming hyperspectral missions such as FLEX and the Copernicus Hyperspectral Imaging Mission for the Environment (CHIME) will substantially increase spectral resolution, information content, and sensitivity to subtle biochemical and structural signals [13,16,17]. While this expanded capability enables more detailed monitoring of vegetation processes, it also intensifies challenges related to spatial resolution, structural heterogeneity, and scale-dependent inference. In this context, radiative transfer models (RTMs), such as PROSAIL [18] and related canopy–atmosphere formulations, e.g., SCOPE [19], provide the mechanistic link between observed spectra and SIF signals and underlying vegetation properties. As these RTMs interact with finer spatial and spectral sampling, their underlying assumptions (e.g., canopy homogeneity, leaf–canopy decoupling, and linearity in aggregation) become increasingly critical and may no longer hold across heterogeneous landscapes [20,21]. Ensuring physically consistent and interpretable products, therefore, requires careful consideration of scale effects within RTM-based retrieval and inversion frameworks.

These developments increasingly support the analysis of vegetation dynamics rather than static canopy states. Quantitative vegetation products are now widely used to characterise seasonal trajectories, phenological transitions, stress and disturbance responses, recovery behaviour, and long-term ecosystem trends, often derived through curve fitting, phenology frameworks, or temporal segmentation approaches (e.g., [22–25]). Under this shift toward dynamic interpretation, scale emerges as a key factor for understanding retrieved signals and their temporal behaviour [26]. Vegetation traits and SIF are retrieved quantities defined by observation characteristics, including spatial resolution (sensor footprint), temporal integration, spectral response, and preprocessing choices (e.g., [20,27,28]).

https://doi.org/10.3390/rs18121951

Because retrieval mappings are nonlinear and landscapes heterogeneous, aggregation and retrieval are generally non-commutative; changes in spatial, temporal, or spectral resolution modify the retrieval problem itself (e.g., [21,29,30]). As a result, differences in resolution or temporal integration can induce systematic biases in spatial gradients, phenological timing, disturbance signals, and trend estimates (e.g., [31–33]).

The central premise of this review is that increasing spatial, spectral, or temporal resolution does not inherently improve ecological accuracy. While finer resolution may reduce mixing effects, it can also amplify measurement variance, expose unmodelled structural complexity, or alter the representativeness of pixel-scale quantities [34,35]. Accuracy is, therefore, not a monotonic function of nominal pixel size [26,34]. This phenomenon may be described as a *resolution–accuracy paradox* in quantitative vegetation remote sensing: increasing observational detail can change the retrieval problem itself rather than simply improving its solution. The paradox summarises the practical consequences of established phenomena such as scale and resolution effects, mixed pixels, aggregation bias, and non-commutativity. Accordingly, the term is not intended to introduce a new physical principle but, rather, to provide a unifying conceptual description of how these established effects collectively lead to situations in which higher nominal resolution does not necessarily translate into improved ecological accuracy.

The practical consequences of these interacting effects extend throughout operational processing pipelines and time-series analysis frameworks. Preprocessing steps such as atmospheric correction, quality screening, compositing strategies, and periodic reprocessing introduce implicit aggregation and evolving retrieval formulations that can imprint discontinuities or resolution-dependent artefacts in derived dynamic metrics [28,36]. Without explicit consideration of observation resolution and aggregation operators, apparent ecological change may partly reflect methodological decisions rather than ecosystem dynamics.

A substantial body of literature has examined scale, variability, and uncertainty in vegetation remote sensing. Foundational work established the multidimensional nature of scale, including spatial, spectral, directional, and temporal components, and its implications for remote sensing and canopy RT modelling (e.g., [26,27,37,38]). More recent syntheses addressed scaling and representativeness challenges for imaging spectroscopy [20], the expanding SIF literature (see reviews, [4,39]), and validation practices across EO communities [28]. Collectively, these reviews have provided important insights into representativeness, retrieval performance, uncertainty, scaling, and validation. However, they generally treat these topics separately and provide limited guidance on how spatial resolution, temporal integration, aggregation operators, retrieval nonlinearity, and uncertainty interact to shape vegetation spatial patterns, temporal dynamics, and derived ecological indicators. Consequently, scale is rarely treated as an overarching methodological dimension linking observation characteristics, retrieval formulations, aggregation processes, and the interpretation of vegetation dynamics.

In this conceptual review, we frame quantitative vegetation traits and SIF dynamics as an explicitly multiscale retrieval problem. We develop and synthesise a scale-aware framework that distinguishes observation characteristics (footprint, temporal integration, spectral response), retrieval level, trait definition, and aggregation operators, and we analyse how their interaction determines the effective spatial and temporal scale of vegetation products. We propose that scale behaviour should be made explicit through a clear definition of resolution and integration characteristics, aggregation-consistency tests, and scale-stratified evaluation, enabling the interpretation of vegetation dynamics as ecological processes rather than artefacts of resolution mismatch.

The synthesis is framed within the emerging era of imaging spectroscopy missions, including FLEX and CHIME. Their increased spectral dimensionality and temporal density

expand the observable information space but also amplify challenges related to structural heterogeneity within sensor footprints (e.g., patchy mixtures of land-cover types, canopy gaps, and vegetation–soil contrasts) and scale-consistent interpretation. Ensuring that hyperspectral-era products yield robust ecological insight hinges on explicit modelling, testing, and documentation of scale behaviour throughout the processing chain.

2. A Taxonomy of Scales for Trait Retrieval and SIF-Based Products

In quantitative vegetation EO, scale is inherently multidimensional and cannot be reduced to spatial pixel size alone. Vegetation trait retrievals and SIF-based products are influenced by several distinct scale dimensions that jointly determine how vegetation structure and function are represented in satellite-derived variables. Four principal scale dimensions can be distinguished:

1. The spatial, temporal, spectral, and angular resolution of the observations (*observation scale*).
2. The organisational level at which retrieval frameworks derive vegetation properties from observations (*retrieval scale*).
3. The physical interpretation of the target variable itself (*trait-definition scale*).
4. The mathematical operators used to aggregate, resample, or compare observations and retrievals across spatial, temporal, or spectral resolutions (*aggregation operators*).

Together, these elements determine the *effective scale* at which vegetation information is represented in derived products. This effective scale reflects both observation characteristics and the retrieval and aggregation procedures applied during processing.

These scale dimensions are interdependent. Modifying the observation resolution, retrieval formulation, or aggregation strategy can alter the effective meaning of retrieved variables even when nominal units remain unchanged. Explicit differentiation of these scale concepts is thus essential for consistent interpretation of vegetation structure, physiology, and ecosystem dynamics in EO analyses (see reviews, [20,27]). The following subsections elaborate these scale dimensions. We first describe the observation scale (spatial, spectral, temporal, and angular), then distinguish the retrieval scale and the trait-definition scale, and finally, we discuss aggregation operators that link estimates across resolutions and determine the effective scale of vegetation products (see also Figure 1).

Although vegetation traits and SIF-based products differ conceptually, traits represent structural or biochemical canopy properties, whereas SIF-based products arise from fluorescence emission linked to photosynthetic regulation; they are subject to the same scale interactions arising from sensor resolution, RT nonlinearities, and aggregation operators (see reviews, [4,6,8]).

2.1. Observation Scale

Observation scale in satellite remote sensing is defined by the spatial, temporal, spectral, and angular resolution over which the sensor integrates the radiometric signal. These characteristics are governed by: (i) the spatial footprint determined by the instrument point-spread function (PSF) and instantaneous field of view, (ii) the temporal integration or compositing window over which photons are accumulated or observations aggregated, and (iii) the spectral response function (SRF) defining the bandpass over which radiance is convolved (e.g., [40,41]). Satellite observations, therefore, represent spatial, spectral, and temporal integrals of the radiative field rather than point measurements of surface state.

Spatial sampling distance (pixel size) does not necessarily correspond to effective spatial resolution. Instrument PSFs and atmospheric adjacency effects redistribute radiance across neighbouring pixels, so measured signals may originate from areas extending beyond the nominal footprint (e.g., [42–44]). Similarly, temporal compositing modifies the

https://doi.org/10.3390/rs18121951

representation of vegetation dynamics by averaging observations across time windows that may span phenological transitions or disturbance events (e.g., [3,45]). In the spectral domain, finite bandwidth integration smooths narrow absorption features encoding biochemical and structural information, reducing sensitivity to pigments, water content, and other leaf constituents detectable with imaging spectroscopy (e.g., [46,47]).

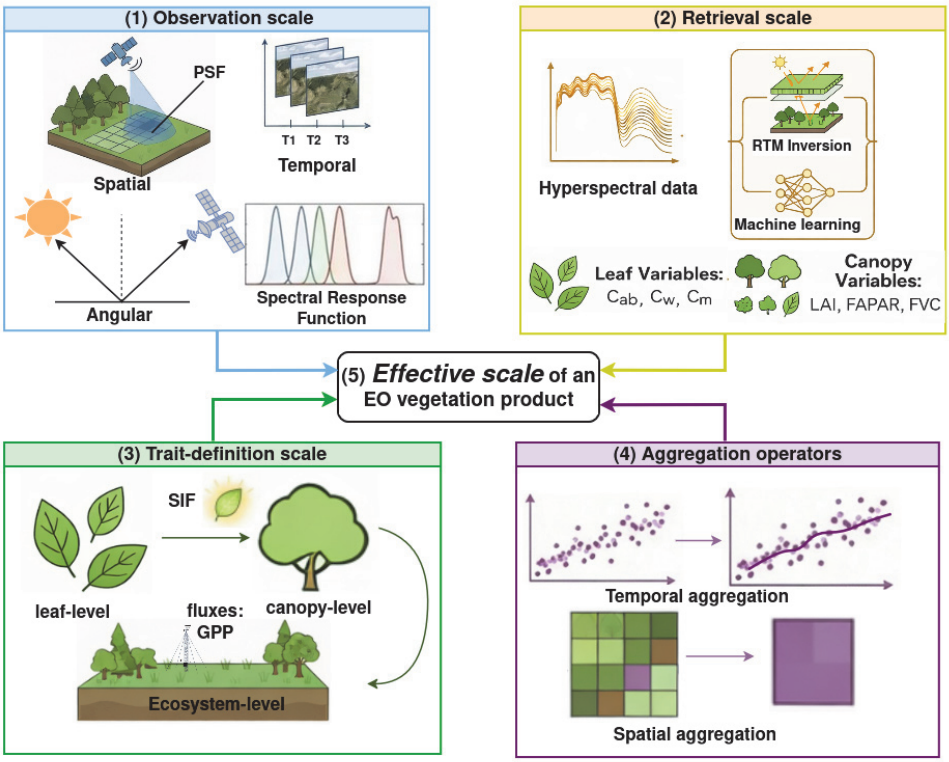


Figure 1. Taxonomy of scale concepts in vegetation EO. (1) Observation scale, defined by point-spread function (PSF), spectral response function (SRF), temporal sampling, and viewing geometry. (2) Retrieval scale, where RTM inversion or ML models map radiometric observations to leaf-, canopy-, or SIF-related variables. (3) Trait-definition scale, emphasising that vegetation variables are biophysical constructs defined at different biological levels, from leaf to ecosystem. (4) Aggregation operators, including spatial averaging, PSF convolution, and temporal compositing, with the important non-commutativity between "aggregate-then-retrieve" and "retrieve-then-aggregate". Together, they determine (5) Effective scale, representing the emergent scale of the final vegetation product resulting from the interaction of observation, retrieval, and aggregation.

Observation sampling is further modulated by illumination and viewing geometry, bidirectional reflectance distribution function (BRDF) behaviour, and normalisation procedures, which influence how canopy structure, background contributions, and anisotropic scattering are expressed in the observed signal (e.g., [21,27,48]). Differences in spatial, temporal, spectral, and angular response, therefore, constitute a primary source of cross-sensor inconsistencies in vegetation trait retrievals and SIF-based products [49]. Operational surface reflectance processing pipelines aim to reduce these geometry- and sensor-induced discrepancies and provide radiometrically consistent inputs for time-series analysis. Those frameworks typically address atmospheric correction, BRDF normalisation, spectral response differences, and quality screening to improve temporal consistency across multi-sensor observations (e.g., [28,36]). While such steps enhance comparability, they do not eliminate scale effects arising from scene heterogeneity, nonlinear RT, and nonlinear retrieval mappings.

https://doi.org/10.3390/rs18121951

In a more operational context, Analysis Ready Data (ARD) concepts formalise these observation characteristics by standardising geometric and radiometric preprocessing, quality masking, and metadata for large-scale time-series analytics [36]. Complementarily, QA4EO (Quality Assurance Framework for Earth Observation) emphasises traceable calibration, validation, and uncertainty characterisation to support interoperability and the provision of standardised, traceable uncertainty, calibration, and validation information across EO products (e.g., [28,50]). These initiatives provide an essential foundation for reproducibility, but do not resolve structural scale-induced biases resulting from aggregation and retrieval non-commutativity.

2.2. Retrieval Scale

Retrieval scale refers to the level of organisation at which inversion or learning models estimate vegetation properties from satellite observations. Some approaches target leaf-level biochemical or structural parameters, such as leaf chlorophyll content (C_{ab}), carotenoids, equivalent water thickness (C_w), or dry matter content, whose optical effects are described using leaf RTMs such as the PROSPECT family (e.g., [51,52]). These parameters are interpreted at canopy scale through canopy RTMs that encode assumptions about canopy architecture, background reflectance, clumping, and vertical structure (e.g., PROSAIL) [18,53]. Other approaches estimate canopy-integrated or effective quantities directly, including LAI, FAPAR, fractional vegetation cover (FVC), or canopy chlorophyll content (CCC; often treated as LAI–C_{ab}) (e.g., [54,55]). SIF-based products provide process-related indicators linked to photosynthetic activity, derived from SIF emission signals and influenced by structural canopy traits (see reviews, [4,39]). The retrieval scale is, therefore, determined not only by the target variable but also by the model structure and mapping between observations and state variables (see reviews, [18,53]).

The retrieval scale does not necessarily coincide with observation resolution. For instance, canopy-level traits may be retrieved from fine-resolution observations, whereas leaf-level parameters are often inferred implicitly within canopy-scale RTMs. In the latter cases, retrieved leaf variables represent effective parameters conditioned on canopy structure, illumination geometry, and model assumptions rather than direct measurements, making their interpretation model-dependent (e.g., [56–58]). The retrieval scale is also influenced by the modelling framework used to relate observations to vegetation variables, including RTM inversion and statistical or machine learning approaches (e.g., [59–62]).

For example, C_{ab} may be retrieved explicitly as a leaf-level parameter through the inversion of coupled leaf–canopy RTMs or implicitly via statistical models trained on canopy reflectance. In both cases, the retrieval operates at different effective scales, even though the variable is nominally defined at the leaf level. Consequently, retrieved C_{ab} often represents a canopy-conditioned, model-dependent quantity rather than a direct measurement of leaf biochemistry. Mismatches between the retrieval scale and the intended interpretation scale can introduce ambiguity and cross-product inconsistencies in vegetation trait estimates [6,8].

2.3. Trait Definition Scale

The trait-definition scale refers to the physical meaning of a vegetation variable, independent of the retrieval algorithm. Vegetation traits are not inherently scale-invariant: leaf biochemical properties (e.g., C_{ab}), canopy-integrated structural variables (e.g., LAI), radiative fractions (e.g., FAPAR), and ecosystem-level flux proxies (e.g., GPP) represent distinct biophysical constructs even when expressed in similar units. The trait-definition scale, therefore, concerns conceptual interpretation, whereas the retrieval scale refers to the modelling and observational framework used to estimate the variable (see reviews, [2,6]).

https://doi.org/10.3390/rs18121951

Nominally identical variables may differ in definition across products. LAI, for example, may refer to true structural leaf area, the effective LAI under turbid-medium assumptions, or the clumping-corrected LAI accounting for non-random foliage distribution (e.g., [63–65]). Likewise, variables retrieved at a similar scale may differ in their physical interpretation. For instance, FAPAR estimates derived from comparable canopy-scale observations may represent absorption by green vegetation only, total canopy absorption including non-photosynthetic elements, or effective radiative fractions conditioned by canopy structure and illumination. Although the retrieval scale is similar in these cases, the underlying trait definitions differ, leading to systematic differences in magnitude, interpretation, and model compatibility.

While the trait-definition scale concerns the conceptual meaning of the variable, its interpretation in practice may be influenced by other scale dimensions. For instance, temporal aggregation can blend phenological stages, illumination conditions, and physiological responses within composite values, thereby altering the apparent meaning of retrieved variables (e.g., [32,66,67]). These effects arise from aggregation (see next) and temporal sampling (see Section 4) rather than from the intrinsic definition of the trait itself.

2.4. Aggregation Operators

Aggregation operators describe how observations or retrieved variables are combined when represented at a coarser spatial or temporal resolution, defining the mapping from fine-scale signals to larger pixels or longer integration windows (e.g., [26,29]). Spatial aggregation includes area-weighted averaging of sub-pixel reflectance, PSF-based radiance integration, or the averaging of vegetation traits across heterogeneous land-cover fractions (e.g., [30,68]). Temporal aggregation commonly arises through the compositing or smoothing of multi-date observations. These operators encode assumptions about how variability propagates across scales, whether through simple averaging, weighted integration, nonlinear transformation, or model-based upscaling (e.g., [26,27]). Because vegetation retrievals are typically nonlinear, aggregation and retrieval are generally non-commutative: retrieving traits from aggregated signals differs from aggregating fine-scale retrievals (e.g., [21,29]). In heterogeneous landscapes, aggregation choices influence spatial gradients, temporal amplitudes, and trend behaviour when comparing or fusing vegetation trait and SIF products across sensors (e.g., [20,27,30,69]).

2.5. Effective Scale

Together, (1) observation characteristics, (2) the retrieval scale, (3) the trait-definition scale, and (4) aggregation operators determine the effective scale of a vegetation product, i.e., the spatial and temporal level at which information is effectively represented after accounting for observation, retrieval, and aggregation processes [27]. In heterogeneous landscapes and under nonlinear retrieval mappings, this effective scale may differ substantially from the nominal pixel size or revisit interval, as it emerges from the interaction between measurement, modelling, and aggregation processes. This makes the explicit articulation of effective scale essential for meaningful comparison, fusion, and ecological interpretation.

In practice, the effective scale cannot generally be inferred from the nominal pixel size alone. Instead, it should be diagnosed by jointly considering (i) the observation resolution defined by footprint size, PSF, temporal integration, and SRF; (ii) the scale at which retrieval variables are formulated; (iii) the ecological scale at which the retrieved quantity is defined and interpreted; (iv) the sensitivity of the product to aggregation, e.g., through retrieve–aggregate versus aggregate–retrieve comparisons; and (v) the propagation of uncertainty

https://doi.org/10.3390/rs18121951

across scales. Products that exhibit strong aggregation sensitivity, non-commutative behaviour, or scale-dependent uncertainty may contain less independent information than implied by their nominal resolution and may, therefore, require interpretation at coarser spatial or temporal scales. Examples of commonly used vegetation variables are provided in Table 1, illustrating how differences in the trait-definition scale and retrieval semantics influence effective-scale considerations and ecological interpretation.

Table 1. Examples of commonly used vegetation variables with different trait-definition scales and associated effective-scale considerations. Variables include leaf chlorophyll content (Cab), canopy chlorophyll content (CCC), leaf area index (LAI), fraction of absorbed photosynthetically active radiation (FAPAR), solar-induced fluorescence (SIF), and gross primary productivity (GPP). Effective scale emerges from the interaction of observation support, retrieval formulation, aggregation behaviour, and uncertainty propagation, and should, therefore, not be interpreted as a fixed property of a product.

Variable	Trait-Definition Scale	Typical Interpretation	Effective-Scale Considerations
Cab	Leaf biochemical trait	Leaf chlorophyll concentration	Sensitive to canopy structure, viewing geometry, and sub-pixel heterogeneity.
CCC	Canopy biochemical stock	Integrated canopy chlorophyll content	Depends on both chlorophyll concentration and canopy density.
True LAI	Canopy structural property	Total one-sided leaf area	Aggregation behaviour influenced by canopy heterogeneity and scale mismatch.
Effective LAI	Radiative-transfer property	Optically effective canopy density	Depends on canopy architecture, clumping, viewing geometry, and RT assumptions.
FAPAR	Canopy–ecosystem functional variable	Fraction of absorbed photosynthetically active radiation	Influenced by canopy structure, illumination conditions, and temporal compositing.
SIF	Canopy–ecosystem functional signal	Photosynthetic activity and stress	Affected by footprint mixing, canopy reabsorption, illumination geometry, and temporal aggregation.
GPP	Ecosystem process variable	Carbon uptake rate	Integrates processes across broader spatial and temporal scales through modelling and data assimilation.

These considerations can be translated into a series of operational questions to diagnose effective scale in real applications: (1) Does the retrieved quantity represent a leaf, canopy, ecosystem, or landscape property? (2) Does retrieval performance change when observations are aggregated before or after retrieval? (3) Do uncertainty estimates increase or change systematically across scales? (4) Are complementary observations and auxiliary data representative of the same spatial and temporal support? (5) Does the interpretation of the product remain stable under changes in resolution or compositing strategy? If these conditions are not met, the nominal resolution may overstate the effective information content of the product, such that reliable interpretation is only possible at coarser spatial or temporal scales.

Beyond the observation–retrieval–aggregation chain discussed above, additional factors can influence the effective scale at which vegetation products are interpreted. In particular, it is useful to distinguish between physical and computational sources of scale dependence. Observation-driven effects originate from sensor characteristics and acquisition geometry, including PSF, SRF, footprint size, temporal integration, and viewing geometry,

https://doi.org/10.3390/rs18121951

which determine how variability is sampled and represented. By contrast, workflow-induced effects emerge after acquisition from the way observations are stored, accessed, and processed. Although these computational operations do not alter the measurements themselves, they can influence which scales of variability are preserved, aggregated, or emphasised during analysis.

These considerations have become increasingly relevant in modern cloud-native EO environments, where additional scale operators—defined here as processes that transform, aggregate, or filter information across spatial, temporal, or spectral dimensions—arise from data representation and computational design (see review, [70]). Chunking strategies, tiling schemes, temporal compositing windows, and distributed processing paradigms implicitly define the spatial and temporal support over which observations are accessed and analysed (e.g., [71–74]). These system-level operations act as effective aggregation operators, conditioning which scales of variability are preserved or suppressed. Consequently, the effective scale is influenced not only by the observation–retrieval–aggregation chain but also by the data and computational architecture through which analyses are executed.

2.6. Scale Dependence of Retrieved Vegetation Quantities

Given these scale dimensions, a key question is whether retrieved quantities preserve their statistical expectation and dynamics under changes in spatial, temporal, or spectral resolution. In reality, most vegetation traits and SIF-based products are structurally scale-variant (e.g., [20,26,27,29]). The root cause lies in the nonlinear nature of RT and retrieval mappings, which makes aggregation and retrieval generally non-commutative (e.g., [21,30,67]):

$$\mathcal{R}(\mathcal{A}(y)) \neq \mathcal{A}(\mathcal{R}(y)), \tag{1}$$

where $\mathcal{R}$ denotes the retrieval mapping, and $\mathcal{A}$ denotes a generic aggregation operator.

Spatial heterogeneity interacts with nonlinear RT processes (e.g., [21,29]), temporal compositing alters seasonal amplitude and phase (e.g., [22,32], and spectral convolution modifies effective absorption features and sensitivities [20]. Only under restrictive conditions, linear retrievals, homogeneous canopies, and aggregation operators aligned with variable definition would scale invariance be approached, conditions rarely met in operational products [4,20,26].

A simple illustrative example highlights this non-commutativity. Consider two equal-area subpixels with LAI values of 1 and 5, and a simplified nonlinear observation response given by $y = 1 - \exp(-0.5 \times \mathrm{LAI})$. Retrieving LAI at the subpixel level and subsequently averaging yields a mean LAI of $(1 + 5)/2 = 3.0$. In contrast, averaging the observation values first gives $\bar{y} = [(1 - \exp(-0.5 \times 1)) + (1 - \exp(-0.5 \times 5))]/2 = 0.656$. Inverting the same observation relationship then yields an apparent LAI of $-\ln(1 - \bar{y})/0.5 = 2.1$. The resulting closure error of 0.9 demonstrates that aggregation and retrieval can produce different outcomes even in the absence of measurement noise, illustrating how nonlinear saturation and spatial heterogeneity generate scale-dependent bias.

In practice, aggregation consistency can be evaluated through closure tests that compare retrieve–aggregate and aggregate–retrieve workflows. The resulting closure error quantifies the sensitivity of a retrieved quantity to aggregation and provides a simple diagnostic of scale dependence. Small closure errors indicate that retrievals are relatively robust to changes in support, whereas larger discrepancies suggest stronger interactions between nonlinear retrieval behaviour and spatial, temporal, or spectral heterogeneity (e.g., [21,26,29,30]).

Because acceptable levels of inconsistency depend on the retrieved variable, retrieval uncertainty, and intended ecological application, universal thresholds are difficult to define. Instead, closure errors should be interpreted relative to the expected uncertainty budget

https://doi.org/10.3390/rs18121951

and the magnitude of variability relevant to the analysis. Discrepancies substantially smaller than retrieval uncertainty or natural variability may be considered negligible, whereas closure errors comparable to or exceeding these quantities indicate that aggregation effects are likely to influence interpretation. For example, a closure error of 0.1 LAI units may be insignificant for regional phenology studies but become important for field-scale calibration or stress detection applications. In a Sentinel-2 LAI product, closure tests can be implemented by comparing LAI retrieved at native 10–20 m resolution and subsequently aggregated to a coarser support with LAI retrieved directly from reflectance aggregated to the same support. Similar tests can be applied temporally by comparing averages derived from daily retrievals with retrievals obtained from temporally composited observations. Such diagnostics provide a practical means of assessing whether a product preserves its interpretation across scales and identifying conditions under which scale-dependent biases become important.

2.7. Diagnostics of Spatial Homogeneity and Representativeness

Diagnostic metrics have been proposed to assess whether pixels or sites satisfy implicit scale assumptions, particularly for moderate-resolution missions such as FLEX and Sentinel-3, where sub-pixel heterogeneity affects retrieval stability and validation [75,76]. Spatial homogeneity can be quantified using higher-resolution imagery through the spatial homogeneity index (SHI), defined as the coefficient of variation of reflectance within a footprint:

$$\mathrm{SHI} = \frac{\sigma(\rho)}{\bar{\rho}}, \tag{2}$$

where $\sigma(\rho)$ and $\bar{\rho}$ denote the standard deviation and mean reflectance [77]. Low SHI values indicate homogeneous canopies, whereas high values signal heterogeneity and increased aggregation bias.

In addition to radiometric variability, landscape configuration metrics provide complementary information on the spatial organisation of heterogeneity. Specifically, the aggregation index (AI) quantifies the degree of clumping or patchiness of similar land-cover classes within a footprint, based on adjacency relationships between pixels [78]:

$$\mathrm{AI} = \frac{g_{ii}}{g_{ii}^{\max}} \times 100, \tag{3}$$

where g_{ii} is the number of like adjacencies between pixels of class i, and $g_{ii}^{\max}$ is the maximum possible number of such adjacencies given the class abundance. High AI values indicate spatially aggregated (canopy-dominated) landscapes with low patchiness, whereas low values reflect fragmented or highly patchy structures. Unlike SHI, which captures within-footprint variability of continuous reflectance, AI characterises the spatial configuration of categorical land-cover patterns. This distinction is particularly relevant for nonlinear retrievals, where patchiness, fragmentation, and edge effects can induce aggregation biases even under moderate radiometric variability.

At the validation level, the spatial representativeness index (SRI) quantifies the mismatch between in situ observations (T_F) and satellite estimates (T_P):

$$\mathrm{SRI} = |T_F - T_P|, \tag{4}$$

where T may represent reflectance, temperature, SIF, or a vegetation trait [79]. Low SRI values indicate strong representativeness, whereas high values reveal footprint mismatch and associated validation bias.

https://doi.org/10.3390/rs18121951

More generally, geostatistical diagnostics such as the variogram range, spatial autocorrelation (e.g., Moran's *I*) [80], and fractal scaling metrics characterise the intrinsic spatial structure of vegetation and reflectance fields [27,29]. These approaches enable an explicit comparison between observation resolution and characteristic landscape scales, helping to identify conditions under which aggregation preserves or distorts signals. While they do not eliminate scale effects, they provide practical criteria for filtering validation sites and designing resolution-aware workflows (e.g., [76,81]). The multidimensional scale taxonomy outlined above motivates concrete methodological responses. Figure 2 summarises practical approaches to resolution-aware analysis, linking the conceptual framework developed here to the scale-induced bias mechanisms analysed in subsequent sections.

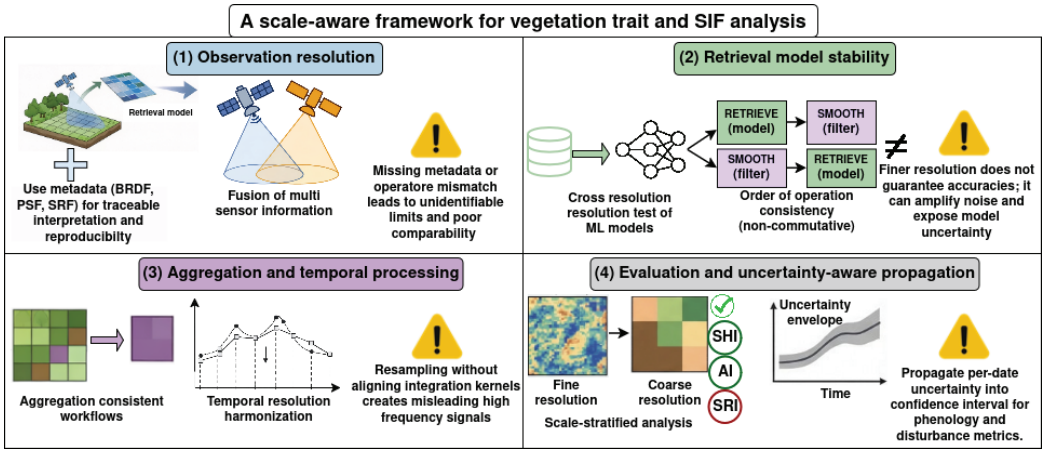


Figure 2. Scale-aware framework for vegetation trait and SIF analysis across the Earth observation processing chain. Four stages are distinguished: (1) observation resolution, defined by PSF, SRF, and acquisition geometry, requiring appropriate metadata for traceable and consistent interpretation; (2) retrieval model stability, emphasising cross-resolution robustness and consistent processing order (non-commutativity); (3) aggregation and temporal processing, for which spatial averaging and temporal harmonisation must align with integration kernels to avoid artefacts; and (4) evaluation and uncertainty-aware propagation, including scale-stratified analysis using SHI (Spatial Homogeneity Index), AI (Aggregation Index), and SRI (Spatial Representativeness Index) diagnostics, together with uncertainty envelopes over time. SHI quantifies within-footprint radiometric heterogeneity, AI characterises the degree of spatial aggregation or patchiness of land-cover classes, and SRI assesses the representativeness mismatch between in situ observations and satellite products.

3. Spatial Scale Effects as Sources of Bias in Trait and SIF-Based Products

3.1. Spatial Resolution and the Conditioning of the Retrieval Problem

Spatial scale fundamentally shapes quantitative vegetation trait retrievals and SIF-based products. Bias arises not only from differences in nominal sensor resolution but from interactions between spatial heterogeneity, nonlinear retrieval mappings, canopy RT processes, and the spatial scale at which observations are aggregated and interpreted (e.g., [21,27,29]). Consequently, retrieved traits and SIF-based indicators are generally not invariant to spatial resolution, even when locally unbiased at their native pixel size. For example, in a mixed vegetation–soil pixel, retrieval from aggregated signals does not generally reproduce the mean of fine-resolution retrievals, because canopy RT processes and retrieval mappings are nonlinear.

A common assumption in remote sensing is that a finer spatial resolution leads to more accurate vegetation trait retrievals, as it reduces sub-pixel mixing and reveals local variability. However, retrieval performance must be defined relative to a specific target quantity and its spatial representation [69]. A retrieval unbiased for an effective canopy variable at 500 m is not necessarily unbiased for a leaf- or crown-scale structure at

10 m. Changing spatial resolution without adapting trait definition, observation operators (e.g., PSF, BRDF normalisation), and aggregation strategy effectively modifies the retrieval target itself (e.g., [40,41]). From this perspective, changes in spatial resolution should be interpreted as changes in the effective scale of the retrieval, rather than as a simple refinement of spatial detail.

At finer spatial scales, retrievals become increasingly sensitive to structural heterogeneity (e.g., shadowing, clumping, background exposure) and measurement noise (e.g., [30,60]). This expands the effective dimensionality of structural and nuisance parameters in the inverse problem, potentially reducing parameter identifiability unless additional constraints are introduced (e.g., [58,59]). Because measurement noise, structural mismatch, and representativeness respond differently to spatial resolution, overall accuracy is not guaranteed to improve monotonically with decreasing pixel size (see reviews, [27,69]). Under these conditions, the sensitivity of the retrieval model to trait variability may play a key role in determining whether finer-scale observations lead to improved trait estimates or simply amplify structural and noise-related effects.

3.2. Bias from Nonlinear Retrievals in Heterogeneous Landscapes

Beyond changes in variance and identifiability, spatial resolution can introduce systematic bias when nonlinear retrieval mappings interact with scene heterogeneity. In heterogeneous canopies, reflectance–trait relationships are rarely linear [26,29,82]. Retrieving traits from aggregated reflectance is, therefore, not equivalent to aggregating fine-resolution retrievals. This non-commutativity arises from nonlinear reflectance–trait relationships and from the distribution of sub-pixel states contributing to the observed signal (e.g., mixtures of cropland, forest, grassland, and soil) (e.g., [21,27,30,83]). Scale effects thus emerge from interactions between nonlinear RT processes and aggregation operators (Section 2.4). Aggregation modifies the distribution of sub-pixel states entering the forward model, altering the effective relationship between observed signals and canopy variables.

Vegetation retrieval problems are commonly formulated as inference problems:

$$y = \mathcal{F}(x, \theta) + \varepsilon, \tag{5}$$

where y denotes observed reflectance or SIF, x the target vegetation variables, θ structural and nuisance parameters (e.g., canopy geometry, background reflectance, illumination), and ε a residual error term representing unresolved measurement, model, and representativeness effects. The mapping $\mathcal{F}$ may represent a physically based RTM, a statistical (e.g., ML) model, or a hybrid formulation (see reviews, [2,6,18,68]).

Accordingly, changing spatial resolution does not merely refine the observation y; it modifies the inference problem itself. Finer pixels expose greater structural variability (e.g., shadowing, crown boundaries, directional effects), increasing the dimensionality of latent and nuisance parameters in θ and the sensitivity of $\mathcal{F}$ to sub-pixel variability. Without additional constraints, this can reduce parameter identifiability and increase predictive variance through the bias–variance trade-off (e.g., [59–61]).

From an error-budget perspective, retrieval uncertainty at a given spatial resolution can be conceptually decomposed as:

$$e_{\text{total}} = e_{\text{meas}} + e_{\text{retr}} + e_{\text{struct}} + e_{\text{repr}}, \tag{6}$$

where e_{meas} denotes measurement and atmospheric-correction uncertainty, e_{retr} denotes algorithmic error, e_{struct} structural model mismatch, and e_{repr} representativeness error due to footprint–scene mismatch [26,28]. These components respond differently to spatial resolution. Finer pixels may reduce mixing but can amplify measurement variance,

https://doi.org/10.3390/rs18121951

increase structural mismatch under simplified models, and alter representativeness. Because these components scale differently and interact nonlinearly, overall accuracy does not necessarily improve with finer resolution or increased model complexity (see reviews, [20,27]). This behaviour is well illustrated by LAI, which exhibits systematic resolution-dependent biases arising from canopy heterogeneity and aggregation effects, and has motivated the development of transformation schemes to relate LAI estimates across scales (e.g., [30,84,85]).

More generally, scaling behaviour has been linked to scene "contexture", defined as the interaction between spatial pattern, heterogeneity, and aggregation operators [86]. As a property of the scene, contexture governs how aggregation and retrieval processes interact with spatial heterogeneity, thereby shaping the effective scale of the resulting product. These principles extend to biochemical traits and SIF-based products, which are similarly nonlinear and sensitive to sub-pixel variability (see reviews, [4,20]).

3.3. Compound Canopy Traits and Scaling

Several commonly retrieved vegetation variables are compound canopy traits, formed by combining leaf-level biochemical properties with canopy structure. Examples include canopy water content (CWC), canopy chlorophyll content (CCC), and canopy nitrogen content (CNC), typically expressed as the product of a leaf trait (e.g., equivalent water thickness, chlorophyll content, or nitrogen content) and the leaf area index (LAI) (e.g., [87–92]). Because compound traits combine structural and biochemical components, their behaviour under spatial aggregation is inherently scale-dependent. Aggregation mixes vegetation states and canopy structural conditions (e.g., LAI, FVC, background), altering the effective absorption and scattering regime governing canopy reflectance. Even when leaf-level properties remain similar, changes in canopy structure can modify the spectral expression of the compound variable (e.g., [89,93,94]). Although compound traits such as CWC or CCC can be robust under controlled canopy-scale conditions (e.g., [87–89]), large-scale applications remain sensitive to mixed pixels, soil/background effects, and fractional vegetation cover (e.g., [91,93]). As compound traits inherit uncertainty from both structural and biochemical components, their error budgets scale nonlinearly with spatial resolution. Spatial aggregation can, therefore, alter both the magnitude and interpretation of retrieved values across resolutions.

3.4. Canopy Structure, RT, and Scale Dependence

Canopy structure mediates how spatial aggregation translates into retrieval bias (e.g., [21,29]). Figure 3 contrasts one-dimensional (1D) and three-dimensional (3D) RTM formulations from a scale-aware perspective, highlighting differences in structural realism, parameter dimensionality, and aggregation behaviour. Two modelling paradigms are commonly distinguished:

- 1D (turbid-medium) RTMs. Canopy structure is represented through effective parameters such as LAI, leaf angle distribution (LAD), and clumping factors. Horizontal heterogeneity is not explicitly resolved, and the canopy is treated as vertically structured but horizontally homogeneous (e.g., [18,95]).
- 3D (discrete-element) RTMs. These models explicitly represent crowns, gaps, shadowing, and adjacency effects, enabling a detailed description of canopy architecture, structural anisotropy, and nonlinear radiative interactions (e.g., [60,61,96]).

The importance of structural effects is inherently scale-dependent. At fine resolution, horizontal canopy structure can dominate reflectance anisotropy, requiring explicit 3D representations to capture crown–gap contrasts and directional effects (e.g., [60,61]). As the resolution increases, multiple canopy elements are integrated within a single footprint,

and reflectance may approach that of an "effective" canopy approximated by vertically averaged formulations (e.g., [30,97]). However, the attenuation of horizontal effects is not universal: illumination geometry, BRDF behaviour, and multiple scattering can preserve structural signatures even at coarser resolutions [21]. Scale-related bias, therefore, arises from interactions between spatial resolution, canopy heterogeneity, and RTM structural assumptions, rather than pixel size alone. Increasing structural realism expands parameter dimensionality and can introduce identifiability challenges when observational constraints are limited (Figure 3).

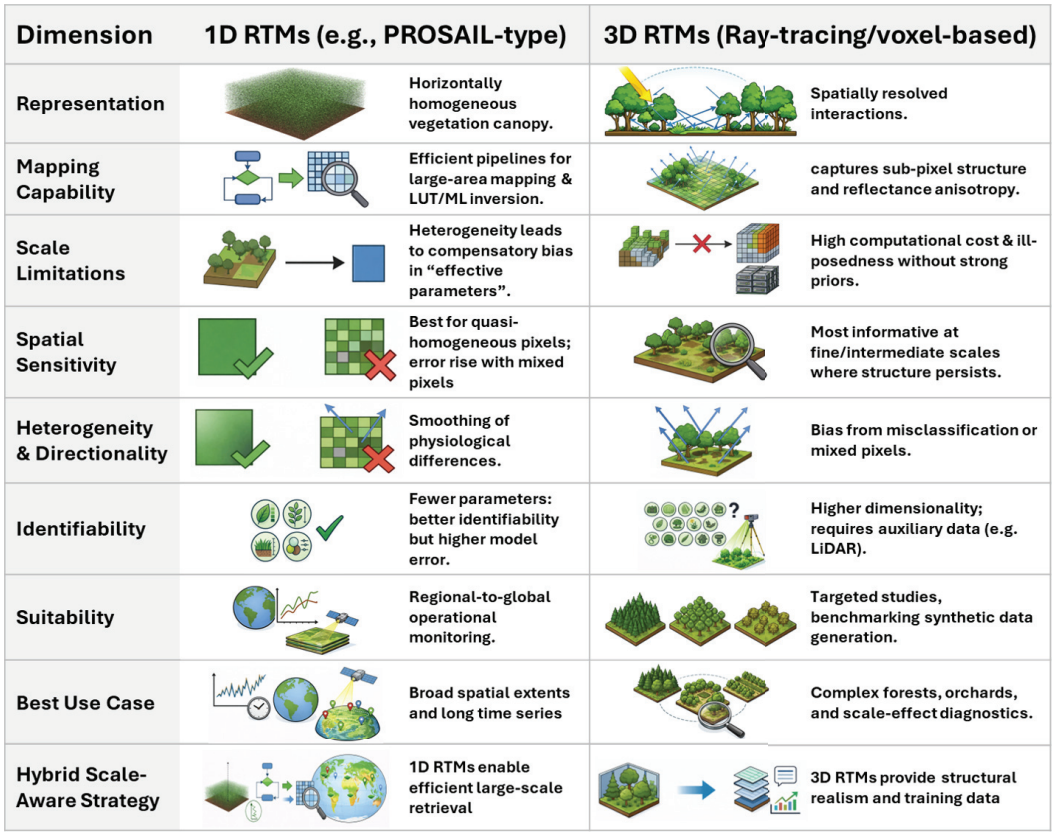


Figure 3. Scale-aware comparison of 1D and 3D RTMs for vegetation monitoring across key dimensions. 1D RTMs represent canopies as homogeneous layers using effective parameters, enabling efficient large-scale retrieval but potentially absorbing structural heterogeneity into biased estimates. In contrast, 3D RTMs explicitly resolve canopy geometry, capturing within-pixel heterogeneity, directional effects, and multiple scattering, albeit at a higher computational cost and with stronger prior requirements. The comparison highlights trade-offs in representativeness, scale sensitivity, and uncertainty and motivates hybrid strategies in which 3D simulations inform or enhance scalable 1D retrieval frameworks.

This explains why simplified turbid-medium models such as PROSAIL often perform adequately at an intermediate-to-coarse resolution (e.g., [18,59]). When heterogeneous canopy elements are integrated within a footprint, simplified formulations can reproduce dominant spectral behaviour, although retrieved parameters represent effective quantities that absorb unresolved structural variability rather than explicitly describing canopy architecture (e.g., [56,58]). In practice, hybrid strategies are common: 3D RTMs are used to analyse structural effects and generate synthetic training data, while computationally efficient 1D models remain the primary tools for operational large-scale retrieval (e.g., [18,60,61]).

3.5. Nonlinear Retrieval Processes: SIF Signals and ML Models Under Scale Change

Beyond the definition of retrieval scale, scale effects also arise from nonlinearities in the retrieval process itself, affecting both SIF- and reflectance-based products. In SIF retrieval, footprint-averaged signals integrate emissions from sunlit and shaded foliage and are modulated by reabsorption, scattering, and canopy structure (e.g., [4,19,98,99]). As the footprint size increases, sub-pixel variability in canopy composition, illumination conditions, and physiological status becomes increasingly aggregated, potentially altering both SIF magnitude and apparent SIF–GPP relationships. In addition, canopy reabsorption and directional effects associated with illumination and viewing geometry may vary across heterogeneous landscapes, causing aggregated signals to differ from the average behaviour of their constituent elements. Temporal aggregation further modifies the interpretation of SIF by smoothing short-term responses to environmental stress and photosynthetic regulation. Collectively, these effects reflect the nonlinear nature of RT and imply that changes in spatial and temporal resolution can modify canopy-leaving SIF magnitude and apparent dynamics independently of physiological regulation (e.g., [21,100]). These effects emerge from the interaction between spatial aggregation (Section 3) and nonlinear retrieval mappings. Similar nonlinear considerations apply to reflectance-based trait retrievals, where increasing spatial resolution reduces averaging but increases sensitivity to canopy anisotropy, sub-pixel heterogeneity, and measurement noise (e.g., [27,29]). As a result, physiological interpretation itself may become resolution-dependent.

ML retrievals exhibit analogous nonlinear behaviour: the learned input–output relationships implicitly reflect the heterogeneity and aggregation characteristics present in the training data (e.g., [6,101,102]). When models are applied across spatial resolutions, the distribution of sub-pixel mixtures and observation conditions changes, whereas the underlying relationships learned from the training data remain unchanged. Consequently, scale-dependent behaviour arises not only in RT-based retrievals but also in ML approaches. As a result, improved performance at finer spatial resolution does not necessarily imply more robust or transferable biophysical information, but may instead reflect resolution-specific structure and mixture effects.

3.6. Implications for Spatial Consistency and Interpretation

Spatial scale effects are structural drivers of bias in quantitative vegetation trait retrievals and SIF-based products. They arise from interactions between landscape heterogeneity, nonlinear retrieval mappings, canopy RT processes, and aggregation operators, and they can occur even when retrievals are unbiased at their native pixel size. Crucially, accuracy is not a monotonic function of spatial resolution or model complexity. Instead, it depends on the alignment between spatial resolution, retrieval formulation, trait definition, and aggregation strategy; that is, on the effective scale at which vegetation information is represented. Robust interpretation of vegetation dynamics, therefore, requires explicit modelling and testing of scale behaviour.

Table 2 synthesises the principal scale-related failure modes discussed in Sections 2 and 3, linking scale dimensions to diagnostic patterns in spatial structure and temporal dynamics. The table serves as a reference for identifying whether apparent discrepancies across products originate from spatial resolution, spectral response, retrieval assumptions, aggregation operators, or algorithm evolution. Together, these mechanisms demonstrate that scale effects are not confined to spatial resolution but emerge from the interaction between retrieval nonlinearity, aggregation, and observation characteristics. These effects extend directly into the temporal domain, where sampling and reconstruction further shape the apparent dynamics of vegetation traits and SIF-based products.

https://doi.org/10.3390/rs18121951

Table 2. Scale-related mechanisms, failure modes, diagnostic symptoms, and mitigation strategies for vegetation trait and SIF-based products. Failure modes are organised according to the interaction between scale dimensions, retrieval assumptions, and ecological interpretation.

Scale Dimension	Affected Traits and Products	Failure Mechanism	Diagnostic Symptoms in Dynamics	Scale-Aware Mitigation Strategies	Representative References
Spatial resolution	LAI, FAPAR, C_{ab}, CWC	Sub-pixel heterogeneity interacts with nonlinear retrieval mappings, so coarse-resolution retrievals are not equivalent to aggregated fine-scale retrievals	Damped seasonal amplitudes; resolution-dependent biases; altered trends across spatial supports	Scale transformation models; heterogeneity stratification; resolution-aware validation using finer-resolution reference data	Tao et al. [84], Tian et al. [103]
Spectral resolution	Pigments (C_{ab}, Car), water-related traits	Bandpass differences and spectral-response mismatch alter effective absorption features and cross-sensor sensitivity	Cross-sensor seasonal offsets; inconsistent interannual variability; reduced comparability of trait anomalies	Spectral response harmonisation; sensor-specific feature selection or harmonised spectral domains; observation-level fusion before retrieval	Blackburn [104], Ustin et al. [105]
Retrieval scale	Leaf- vs. canopy-level traits; CCC, CWC	Mismatch between the scale at which a variable is retrieved and the scale at which it is interpreted	Physically implausible magnitudes; cross-trait inconsistency; erroneous comparisons between leaf- and canopy-scale products	Explicit trait definitions; retrieval formulations consistent with target support; scale-conversion or upscaling relationships where needed	Clevers et al. [87], Gara et al. [106]
Aggregation operator	All nonlinear traits and derived metrics	Non-commutativity of aggregation, resampling, and nonlinear retrieval under heterogeneity leads to biased effective states and distorted dynamics	Bias in mean states; shifted phenology; scale-dependent trends or anomalies	Aggregation-consistent inference; retrieve–then–aggregate versus aggregate–then–retrieve tests; resolution-aware forward modelling	Pinty et al. [21], Jin et al. [107]
Temporal resolution	Traits, SIF-based products, phenology metrics	Temporal compositing and smoothing mix phenological phases and short-duration stress responses across dates	Shifted transition dates; damped seasonal amplitudes; attenuation of short-duration events (e.g., heat or drought stress)	Temporal harmonisation matched to process timescales; segmentation and uncertainty-aware phenology modelling; event-sensitive analyses	Jönsson and Eklundh [22], Verbesselt et al. [24]
Footprint size	SIF-based products, SIF-constrained GPP	Mixed illumination, canopy structure, and physiology within coarse footprints decouple observed SIF from local photosynthetic activity	SIF–GPP decoupling; muted or delayed stress signals; scale-dependent empirical relationships	Aggregation-consistent SIF downscaling; footprint-aware interpretation; consistency constraints between fine and coarse estimates	Duveiller and Cescatti [67], Kang et al. [108]
Canopy structure	LAI, FAPAR, C_{ab}, CWC, SIF-based indicators	Clumping, gaps, and vertical heterogeneity alter photon transport and effective scattering/absorption behaviour, modifying trait sensitivity under aggregation	Resolution-dependent sensitivity; trait–structure inconsistency; directional effects not explained by 1D assumptions	Structure-aware RTM modelling; explicit clumping parameterisation; vertically resolved or heterogeneous canopy representations	Ni-Meister et al. [109], Wang and Li [110]
Algorithm evolution	All traits and SIF-based products	Changes in calibration, atmospheric correction, retrieval algorithms, or processing collections introduce artificial differences unrelated to ecosystem change	Artificial discontinuities; step changes; spurious breaks in long time series; reduced comparability across product versions	Provenance tracking; version-aware analyses; harmonisation across product versions and reprocessing campaigns	Loew et al. [28], Crawford et al. [111]
ML training scale	ML-derived vegetation traits and SIF surrogates	Mismatch between training and application scales, combined with spatial autocorrelation and distribution shift, leads to the breakdown of learned relationships under rescaling and biased generalisation	Overconfident uncertainty estimates; scale-dependent artefacts in temporal dynamics; poor transfer across resolutions, regions, or seasons	Scale-stratified training and evaluation; spatially blocked cross-validation; multi-scale benchmarking and uncertainty calibration	Roberts et al. [112], Valavi et al. [113]

https://doi.org/10.3390/rs18121951

4. Temporal Sampling and Dynamics of Traits and SIF-Based Products

4.1. Temporal Resolution, Sampling Irregularity, and Aliasing

While spatial heterogeneity governs how retrievals respond to landscape structure, temporal sampling determines how vegetation dynamics are reconstructed from discrete satellite observations. Vegetation trait and SIF-based dynamics are not observed continuously but inferred from acquisition sequences. The apparent temporal behaviour reflects the interaction between ecological change and the effective temporal observation scale, determined by sampling interval, integration window, and retrieval formulation.

Let $x(t)$ denote the underlying continuous vegetation state and $y(t_i)$ observations acquired at times t_i. The reconstructed trajectory $\hat{x}(t)$ is obtained through sampling, retrieval, and temporal reconstruction:

$$\hat{x}(t) = \mathcal{T}\Big(\{y(t_i)\}_{i=1}^{N}\Big), \tag{7}$$

where $\mathcal{T}$ denotes the combined retrieval and temporal reconstruction operator. The recoverability of seasonal trajectories and long-term trends depends on sampling density, temporal regularity, and the signal-to-noise ratio.

Accurate reconstruction requires observation frequency sufficiently high relative to dominant process timescales [114]. When revisit intervals exceed these timescales (e.g., rapid stress responses or disturbances), high-frequency dynamics may be missed or aliased into lower-frequency variability [3,114]. In practice, irregular revisit intervals, cloud contamination, and viewing constraints introduce missing data and variable sampling density, complicating trajectory reconstruction and phenological metric retrieval (e.g., [25,115]).

A higher sampling frequency may reveal short-term variability but can also increase sensitivity to noise, retrieval instability, and state-dependent uncertainty across phenological stages (e.g., [115,116]). Derived quantities inherit these constraints: seasonal metrics such as start/end of season (SOS/EOS), transition rates, peak magnitude, or integrals are nonlinear functionals of $\hat{x}(t)$. As a result, sampling density, gap-filling, and smoothing can shift transition dates, damp amplitudes, or alter productivity proxies (e.g., [3,114,115]). Differences across products may reflect reconstruction choices rather than ecological change.

4.2. Temporal Compositing and Smoothing as Aggregation Operators

Temporal compositing and smoothing act as aggregation operators in the temporal domain, analogous to the spatial aggregation operators discussed in Section 3. Compositing combines observations acquired under different illumination and phenological conditions, effectively broadening the temporal integration window [3]. These approaches are widely applied to reflectance-based vegetation traits. Reconstruction frameworks such as TIMESAT [22] or DATimeS [25] are commonly used to extract seasonal trajectories and metrics from irregular time series [115,117]. These methods reshape signal amplitude, timing, and derivatives, thereby modifying the effective temporal scale of reconstructed dynamics.

From a scale-dependent perspective, compositing and smoothing redefine the temporal resolution of observations. Short-term variability may be suppressed, transition dates can shift, and derived metrics become functions of the aggregation strategy rather than direct properties of the underlying process [3]. Let $\mathcal{A}_t$ denote a temporal aggregation operator and $\mathcal{R}$ the retrieval mapping. As in Taxonomy Section 2, these operations are generally non-commutative:

$$\mathcal{R}(\mathcal{A}_t(y)) \neq \mathcal{A}_t(\mathcal{R}(y)), \tag{8}$$

i.e., retrieval and temporal aggregation do not commute. Pre-retrieval compositing modifies the statistical structure of radiometric inputs, whereas post-retrieval smoothing transforms derived variables. Because retrieval mappings are nonlinear, these workflows propa-

https://doi.org/10.3390/rs18121951

gate bias and uncertainty differently (e.g., [2,45,115]). From an error-budget perspective, temporal aggregation alters the balance between noise reduction and representativeness loss. While aggregation can reduce variance, it may also attenuate extremes, shift transitions, and suppress short-duration events, thereby affecting ecological interpretation (e.g., [3,114,115,118,119]).

4.3. Temporal Mismatch Between Trait and SIF-Based Products

SIF-based products introduce additional temporal-scale considerations. Unlike many reflectance-based traits, SIF observations respond rapidly to illumination, photosynthetic regulation, and environmental stress at sub-daily to seasonal timescales (e.g., [120–122]). However, due to signal-to-noise limitations, orbital sampling, and retrieval constraints, current satellite SIF products are commonly analysed and distributed as temporally aggregated composites spanning several days or longer periods (e.g., [123,124]). While this improves stability and resolution, it reduces effective temporal resolution (e.g., [4,67]). These integration windows often differ from those of reflectance-based products, complicating the joint analysis of trait and SIF trajectories, particularly when SIF is used to constrain GPP or stress indicators [10]. Temporal alignment is not merely a resampling problem. Different integration windows imply different temporal filtering of the underlying vegetation state. If $x(t)$ is convolved with distinct kernels for trait and SIF products, the resulting time series represent differently filtered dynamics. Misalignment can shift phase relationships, induce artificial lags, and bias interpretation of coupling between canopy structure and photosynthetic activity [4].

4.4. Algorithm Evolution and Apparent Temporal Change

Temporal inconsistencies may also arise from algorithm evolution and reprocessing of satellite data archives. Updates in atmospheric correction [125], calibration and BRDF normalisation, or cross-sensor harmonisation [126] can introduce discontinuities or spurious trends in long-term time series. Validation frameworks highlight that processing changes across product generations can substantially affect temporal consistency in EO records [28]. Such changes effectively modify the retrieval operator $\mathcal{R}$ over time, even if the revisit frequency remains unchanged. ARD initiatives aim to mitigate these effects through consistent preprocessing and harmonisation [36]. Across sampling, compositing, and reprocessing, the error decomposition in Equation (6) acquires explicit temporal dependence. Measurement, retrieval, structural, and representativeness components become state-dependent functions of time, reflecting seasonal changes in canopy structure, illumination, and observation density. Because derived metrics are nonlinear functionals of $\hat{x}(t)$, distortions introduced during sampling or processing propagate nonlinearly into trends and indicators.

Taken together, temporal sampling, aggregation, and algorithm evolution define the effective temporal scale at which vegetation structure and function are represented. As in the spatial domain, dynamic accuracy is not a monotonic function of sampling density or revisit frequency but depends on alignment between ecological timescales, temporal resolution, retrieval formulation, and aggregation choices (e.g., [114,115]). An explicit consideration of this effective temporal scale is, therefore, essential for a robust interpretation of vegetation trait and SIF dynamics across sensors and missions [3].

4.5. Implications for Temporal Consistency and Interpretation

Temporal-scale effects imply that vegetation dynamics inferred from EO time series cannot be interpreted independently of temporal sampling, integration windows, and compositing strategies. Apparent differences in phenology, stress dynamics, disturbance timing, or trend magnitude may arise from aggregation choices rather than underlying

https://doi.org/10.3390/rs18121951

ecological processes. Robust analysis, therefore, requires: (i) explicit reporting of temporal support and compositing schemes, (ii) consistency between temporal resolution and the characteristic timescales of the target process, and (iii) evaluation of stability under alternative temporal aggregation settings. Temporal scale should thus be treated as an integral component of the retrieval and analysis framework rather than a preprocessing detail.

5. Process-Level Vegetation Dynamics Across Scales

Building on the spatial and temporal mechanisms discussed in Sections 3 and 4, we now examine how scale affects the interpretation of higher-level ecological processes. Processes such as disturbance, stress response, recovery, and ecosystem productivity are not directly observed but inferred from vegetation trait and SIF-based time series. Their apparent dynamics, therefore, inherit the spatial, temporal, and aggregation constraints of the underlying observations and retrievals.

5.1. Disturbance and Stress Detection Under Scale Constraints

Disturbance events and stress dynamics provide a stringent test of scale-aware analysis because they combine rapid structural and physiological change with strong spatial heterogeneity and non-stationary temporal behaviour. Signal responses may differ between reflectance-derived traits and SIF, and scale effects that are subtle during seasonal cycles become amplified under such conditions [27]. Event detectability, timing, and magnitude, therefore, depend strongly on spatial resolution, temporal integration, and the modelling framework used for change detection (e.g., [23,24]). Trait-based disturbance indicators are typically defined as anomalies relative to expected seasonal trajectories (e.g., pigment-related traits, LAI, CWC). At a coarse resolution, the sub-pixel mixing of affected and unaffected vegetation dilutes anomaly magnitude and delays detection [27]. Temporal compositing introduces analogous effects: smoothing or multi-day integration may shift onset dates, suppress short-lived stress responses, or merge multi-phase events (e.g., drought followed by partial recovery) (e.g., [127,128]). These effects are critical when metrics such as time-to-detection, maximum anomaly, or recovery duration are used for attribution or early warning. During drought or physiological stress, ecosystem function may decline before structural changes become detectable in reflectance-based traits, particularly under coarse spatial resolution or temporal smoothing [129]. This implies that interpretation depends on explicit consideration of temporal reconstruction choices and phenological context [3].

5.2. Change Detection and Temporal Segmentation Applied to Traits

The detection and quantification of disturbance and recovery dynamics in retrieved vegetation variables are typically performed using time-series segmentation and change-detection algorithms. Software frameworks such as BFAST [24], LandTrendr [23], and CCDC [130] are commonly applied to analyse disturbances and recovery trajectories in satellite-derived vegetation variables, particularly reflectance-based indices and traits, and hold potential for application to SIF-based products. These methods operate on reconstructed time series and, therefore, inherit the sampling, compositing, and smoothing choices applied during preprocessing. Detected breakpoints and recovery metrics thus reflect the effective temporal scale imposed by aggregation rather than purely intrinsic vegetation dynamics. Additional sensitivities arise from the nonlinear, state-dependent mapping between observed radiance or reflectance and retrieved variables. Phase-dependent variance and residual structure can confound structural change with seasonal variability, while differences in sampling density, gap-filling, or compositing may shift breakpoint timing and alter inferred disturbance magnitude (e.g., [3,115,131]).

https://doi.org/10.3390/rs18121951

At the same time, algorithm updates, preprocessing changes, or archive reprocessing may likewise introduce artificial breakpoints resembling disturbance unless product provenance is carefully tracked (e.g., [132,133]). Robust applications require scale-aware diagnostics, including sensitivity to temporal integration windows, robustness across segmentation settings, and explicit testing of spatial aggregation effects. This introduces an additional layer of scale dependency, as the algorithmic interpretation of change is conditioned by the spatiotemporal representation of the input signal.

5.3. SIF-Based Early Stress Detection and Scale-Dependent Interpretation

A particularly illustrative case of scale-dependent process interpretation arises in the analysis of early stress signals using SIF-based products. These applications aim to resolve subtle temporal lead–lag relationships between physiological and structural responses but are inherently sensitive to spatial and temporal aggregation effects. SIF-based products often respond rapidly to physiological stress and may precede structural changes detectable in reflectance-based traits (e.g., [10,134]). However, SIF is a footprint-integrated radiative signal shaped by canopy structure, illumination geometry, and reabsorption processes (e.g., [4,19,135]). Apparent early stress detection may therefore reflect resolution-dependent integration effects rather than true physiological lead–lag behaviour. Disentangling these mechanisms requires alignment of spatial resolution and temporal integration (Section 4).

Large footprints and multi-day integration can attenuate or delay stress signals, particularly in heterogeneous landscapes (e.g., [123,124]). Conversely, fine-scale or reconstructed SIF products may emphasise variability not preserved under re-aggregation to native resolution (e.g., [39,67]). This implies that interpretation depends on aggregation-consistent modelling (Section 6.3) since enhanced spatial detail does not necessarily imply improved physiological fidelity.

Joint analysis of trait- and SIF-based signals thus relies on careful alignment of spatial resolution and temporal integration to avoid spurious lead–lag interpretations driven by footprint mismatch or differential temporal filtering. In practice, inferred lead–lag relationships should be interpreted as scale-dependent properties of the observation system rather than intrinsic ecosystem behaviour.

5.4. Model Stratification Across Vegetation Types and Scale Effects

At continental-to-global scales, many ecosystem-level variables (e.g., GPP, water fluxes, stress indicators) are not directly retrieved but inferred through models combining vegetation traits, meteorological drivers, and sometimes SIF constraints. These models inherit scale dependencies from both observations and model structure. A key modelling choice is between biome- or plant functional type (PFT)- specific parameterisations and generic continuous models. PFT-based models aim to capture physiological differences (e.g., light-use efficiency (LUE), temperature response, phenology, water stress sensitivity) (e.g., [136,137]) but rely on categorical assumptions derived from land-cover maps (e.g., [138,139]). Because such maps are resolution-dependent and uncertain, PFT stratification can introduce artificial discontinuities and confound temporal interpretation.

This is evident in GPP modelling, where PFT-dependent parameterisations are widely used (e.g., [136,137]). Apparent productivity changes may partly reflect shifts across PFT boundaries rather than ecosystem dynamics. Misclassification and mixed pixels can introduce systematic bias, particularly near class transitions [138]. From a scale-dependent perspective, PFT stratification acts as a categorical aggregation operator, partitioning continuous ecological gradients into discrete classes determined by map resolution and classification accuracy.

https://doi.org/10.3390/rs18121951

- PFT-stratified models. Partitioning variability into discrete categories may obscure continuous ecological gradients (e.g., [139,140]). Increasing complexity through PFT differentiation does not necessarily improve accuracy; instead, it redistributes structural and representativeness errors and increases sensitivity to land-cover uncertainty (e.g., [137,141]).
- Generic models. Generic models preserve continuous gradients but may underrepresent physiological heterogeneity when differences between vegetation types are large (e.g., [142–144]).

Figure 4 summarises these trade-offs in scale-aware terms, highlighting how modelling strategies redistribute scale sensitivity across the processing chain rather than eliminate it. Differences between modelling approaches should be interpreted not only in terms of their ability to capture ecological processes but also in terms of how they represent spatial heterogeneity, aggregation behaviour, and temporal dynamics under varying observation scales. In particular, PFT-based approaches introduce discrete boundaries that can amplify scale sensitivity through classification errors and aggregation effects, whereas continuous models tend to preserve smooth gradients across resolutions. In this sense, model choice implicitly defines the effective scale at which ecosystem dynamics are represented and must be consistent with the resolution and aggregation characteristics of the input data.

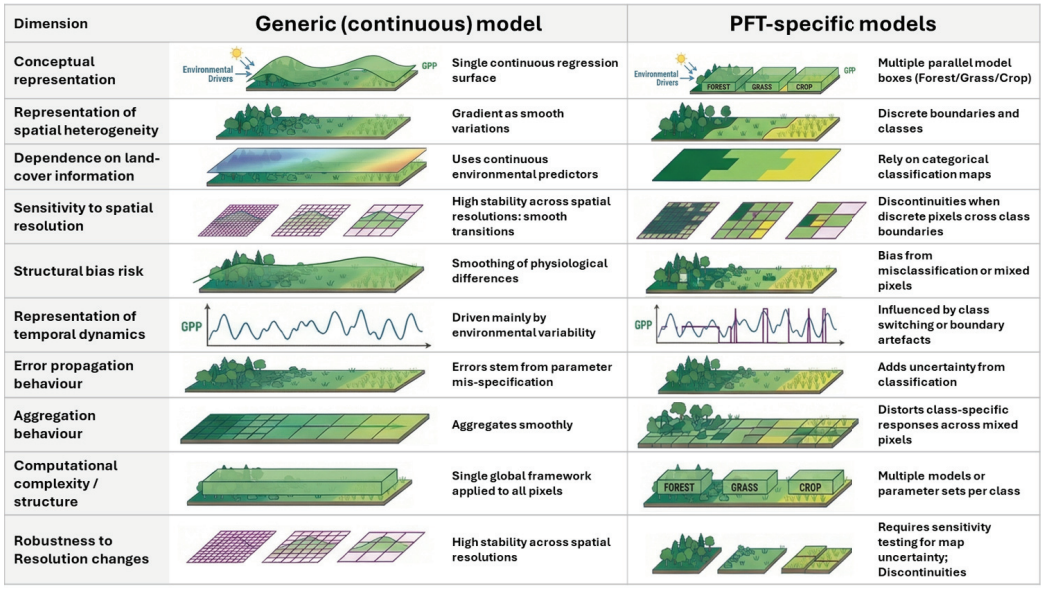


Figure 4. Scale-aware comparison of generic (continuous) and PFT-specific (categorical) ecosystem models for productivity estimation. Generic models represent ecosystems through continuous responses to environmental drivers, yielding smooth, stable behaviour across scales but potentially masking structural heterogeneity. PFT-specific models use class-based parameterisations, capturing functional differences but introducing sensitivity to land-cover classification, mixed pixels, and resolution-dependent discontinuities. The comparison demonstrates distinct error propagation pathways and aggregation behaviour, illustrating that increasing spatial resolution does not necessarily improve accuracy and may amplify inconsistencies.

5.5. Implications for Process-Level Consistency and Interpretation

Process-level interpretation of vegetation traits and SIF-based products requires consistency between the scale of retrieved variables and the scale at which ecological processes operate. Mismatches between observation, retrieval, and process representation can lead to biased or misleading inference, particularly when nonlinear responses are aggregated across heterogeneous conditions. Scale-aware modelling should ensure: (i) alignment between trait definition and process representation, (ii) explicit treatment of aggregation

effects within model–data integration, and (iii) uncertainty propagation across scales. Process-level conclusions should, therefore, be interpreted as scale-dependent inferences rather than scale-invariant properties of the ecosystem.

6. Scale-Aware Data Fusion and Multiresolution Modelling

Building on the spatial, temporal, and process-level scale effects discussed in previous sections, data fusion approaches aim to reconcile observations across sensors and resolutions within a consistent modelling framework. In this context, fusion refers to methodological approaches that combine observations from multiple sensors, resolutions, or data sources to produce spatially and temporally consistent estimates of vegetation traits or process-level variables. Data fusion strategies differ in where scale effects are addressed within the processing chain and whether aggregation consistency and uncertainty traceability are explicitly enforced. Rather than comparing products post hoc, scale-aware data fusion integrates observations across sensors and resolutions within a unified modelling framework.

Data fusion can operate at different stages of the processing chain: (i) at the observation level (radiance or surface reflectance), (ii) at the trait level (retrieved vegetation variables), or (iii) at the process or latent-state level (e.g., joint trait–SIF retrieval). Each level implies distinct assumptions about nonlinearity, heterogeneity, and uncertainty propagation and, therefore, different exposure to scale-induced bias. Figure 5 summarises the principal methodological strategies for handling scale in quantitative vegetation trait retrievals and SIF-based products. The table contrasts where scale is addressed in the processing chain, whether aggregation consistency is enforced, and which spatial, temporal, or spectral dimensions are explicitly modelled. The following subsections elaborate these strategies and their implications for scale-consistent dynamic retrieval schemes.

6.1. Observation-Level Spatiotemporal Fusion (Reflectance-First)

Observation-level fusion generates image time series that are both spatially fine and temporally dense by combining frequent moderate-resolution observations (e.g., Sentinel-3) with less frequent high-resolution data (e.g., Sentinel-2 or CHIME). Representative approaches include neighbourhood-weighted methods such as STARFM (Spatial and Temporal Adaptive Reflectance Fusion Model) and its extensions [145–147], as well as unmixing-based and Bayesian formulations that explicitly model mixed coarse pixels and their temporal evolution (e.g., [148,149]). For vegetation traits, reflectance-first fusion offers two advantages. First, a single retrieval model can be applied consistently across time, reducing algorithmic non-stationarity. Second, uncertainty propagation can be initiated from well-characterised radiometric error models (see review, [150]), which is critical for analysing derivatives, trends, and seasonal metrics. However, fusion operators must remain radiometrically and geometrically consistent. Pre-existing effects such as BRDF, canopy anisotropy, illumination geometry, and land-cover heterogeneity are not always explicitly modelled and can interact with fusion assumptions, particularly when combining observations acquired under differing viewing geometries, sun–sensor configurations, or spatial supports. These directional effects may alter both apparent reflectance and effective observation support, introducing scale-dependent biases that propagate into retrieved traits and complicate cross-sensor consistency (e.g., [126,151]). Also, increasing spatial resolution through fusion does not necessarily improve interpretability. Without explicit aggregation operators and uncertainty characterisation, fusion-derived fine-resolution products may amplify sub-pixel heterogeneity and cross-sensor inconsistencies, leading to apparent fine-scale variability that is not physically meaningful (e.g., [27,29]).

https://doi.org/10.3390/rs18121951

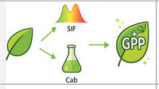
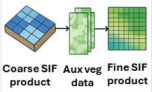
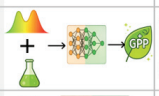
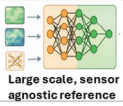
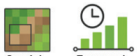
Strategy	Operates on	Scale dimensions	Aggregation consistency	Typical Application	Key limitations
Observation level spatiotemporal fusion			✓ (if radiometric consistency maintained)	 Time series phenology analysis and trend monitoring	Sensitive to radiometric inconsistencies, BRDF effects, and land-cover heterogeneity ⚠
Trait level fusion			⊘ Not guaranteed	 Product harmonization, multi-sensor comparison	Potential propagation of retrieval errors from individual products ⚠
SIF downscaling and reconstruction			Method-dependent	 GPP estimation and stress diagnosis	Illumination geometry and sub-footprint composition can dominate apparent dynamics ⚠
Joint trait & SIF retrieval			✓ (if likelihood-based retrieval is used)	 Physiology-aware GPP and stress products	Model complexity: requires consistent uncertainty characterization ⚠
Hierarchical Bayesian fusion			✓ Explicitly enforced	Statistically rigorous fusion across diverse scales	⚠ High computational cost and complexity
Multi-fidelity learning	 Large scale, sensor agnostic reference		Model-dependent	 Large-scale dynamic inference across sensors	Validation and interpretability across scales remain challenging ⚠

Figure 5. Overview of scale-aware strategies for vegetation monitoring across observation, retrieval, and inference levels. The figure compares approaches including observation-level spatiotemporal fusion, trait-level fusion, SIF downscaling and reconstruction, joint trait–SIF retrieval, hierarchical Bayesian fusion, and multi-fidelity learning. Each strategy is summarised using a minimal visual scheme indicating the data domain, scale dimensions addressed (spatial, temporal, spectral), aggregation consistency, typical application, and key limitations. The comparison highlights fundamental trade-offs between consistency across scales, modelling complexity, and applicability, and underscores the importance of approaches that explicitly handle cross-scale interactions in vegetation trait and SIF workflows.

6.2. Trait-Level Fusion: Retrieve–Then–Fuse vs. Fuse–Then–Retrieve

Trait-level fusion combines retrieved variables (e.g., LAI, FAPAR, pigment- or water-related traits) across sensors and spatial resolutions. This approach is particularly relevant when spectral response functions differ or when retrieval algorithms are mission-specific. In this context, non-commutativity becomes central: for nonlinear mappings, retrieving traits from aggregated reflectance is not equivalent to aggregating traits retrieved at a finer resolution (e.g., [21,27,29]). As a result, retrieve-then-fuse strategies require either: (i) explicit scale-transfer relationships linking trait definitions across resolutions or (ii) aggregation operators that account for sub-pixel heterogeneity, clumping, and background effects [152].

Reflectance-first approaches are often preferred in operational systems to maintain radiometric consistency and enable the use of a single retrieval model at the target scale. Nevertheless, trait-level fusion remains essential for constructing long-term, multi-mission records and for integrating products derived from different retrieval paradigms (e.g., [153,154]). This becomes particularly relevant when combining vegetation traits with SIF-based diagnostics, where differences in spatial resolution and retrieval assumptions can introduce additional inconsistencies [76]. Under a scale-aware framework, trait-level fusion entails explicit treatment of aggregation behaviour to avoid propagating resolution-dependent biases into derived vegetation products and long-term time series.

https://doi.org/10.3390/rs18121951

6.3. SIF Downscaling and Scale-Consistent Constraints on GPP and Stress

Over the past decade, satellite SIF observations have been primarily available at coarse spatial resolution (e.g., GOME-2, OCO-2, TROPOMI), which has motivated the development of downscaling approaches to generate spatially resolved products (e.g., [11,155,156]). Even with the upcoming FLEX, scale mismatches between SIF, reflectance, and ecosystem processes will persist, sustaining the need for multiresolution inference and aggregation-consistent approaches. Existing methods typically combine coarse SIF with higher-resolution reflectance or meteorological predictors through regression (e.g., ML) or LUE-based formulations (see review, [39]). When accounting for scale effects, the key question is not whether finer detail can be generated but whether reconstructed fields remain aggregation-consistent; i.e., whether fine-scale estimates reproduce the original coarse observation when re-aggregated through the appropriate operator (e.g., PSF-consistent integration). Because canopy-leaving SIF is an integrated radiative signal shaped by canopy structure, reabsorption, and illumination (e.g., [19,67,135]), redistribution across heterogeneous landscapes can introduce bias if aggregation behaviour is not explicitly constrained.

Semi-empirical LUE and regression-based approaches often improve agreement with flux-tower GPP (e.g., [100,156,157]), but such improvements do not necessarily imply physically consistent redistribution. SIF–GPP relationships are empirical and scale-dependent, as both integrate processes across heterogeneous canopies and observation footprints (e.g., [4,10,158]). Enhanced correlations may, in practice, reflect resolution-induced artefacts rather than improved physiological inference.

Beyond statistical agreement with GPP or flux-tower measurements, the physical consistency of downscaled SIF products can be assessed through several complementary diagnostics. First, aggregation consistency requires that fine-scale estimates reproduce the original coarse SIF observations when integrated through the appropriate footprint and PSF operator. Second, reconstructed SIF patterns should remain compatible with physically plausible relationships among absorbed radiation, canopy structure, fluorescence yield, and energy balance (e.g., [19,135]). Third, inferred temporal variability should preserve expected responses to environmental forcing and avoid introducing artificial high-frequency fluctuations or spatial artefacts. Finally, consistency across complementary observations, including reflectance, thermal measurements, and vegetation traits, provides an additional check of whether reconstructed SIF fields represent plausible vegetation functioning rather than merely statistical redistribution.

A principled alternative embeds downscaling within a joint inference framework linking fine-scale latent states to each sensor via explicit measurement operators:

$$y_{\mathrm{SIF}}^{\mathrm{coarse}} = \mathcal{A}_s\Big(\mathcal{H}_{\mathrm{SIF}}\Big(x^{\mathrm{fine}}\Big)\Big), \tag{9}$$

where H_{SIF} denotes a generic SIF observation operator that maps fine-scale vegetation states to canopy-leaving SIF, e.g., through an RTM such as SCOPE [19]. The operator A_s represents spatial aggregation (Section 2), accounting for sensor footprint and PSF effects.

Within such formulations, scale consistency is enforced directly in the observation model, ensuring coherence between fine-scale estimates and coarse observations [159]. This avoids inconsistencies introduced by post hoc resampling or regression-based redistribution. This reframes SIF downscaling as a multiresolution retrieval problem in which aggregation behaviour, uncertainty propagation, and physical interpretability are jointly constrained. These considerations motivate modelling strategies in which scale is treated explicitly within the inference formulation rather than corrected post hoc, as discussed in the following subsection.

https://doi.org/10.3390/rs18121951

6.4. Hierarchical and Multi-Fidelity Approaches for Multi-Resolution Modelling

Some retrieval frameworks address scale effects directly within the inference formulation. Hierarchical statistical models represent latent vegetation states alongside sensor-specific observation operators (e.g., PSFs, spectral response, revisit characteristics). Observations from different sensors are linked through a shared latent process, with uncertainty and scale effects incorporated explicitly within the hierarchical formulation (e.g., [160,161]). This enables consistent propagation of uncertainty and scale effects across resolutions.

Multi-fidelity modelling provides a related strategy by combining observations or simulations of differing resolution or complexity. Frequent coarse observations may be integrated with sparse high-resolution data, with relationships learned jointly across fidelity levels (e.g., [162,163]). Similarly, simplified canopy models can be coupled with detailed RT simulations, allowing discrepancies between fidelity levels to be explicitly estimated [164].

In both approaches, scale behaviour becomes an explicit modelling dimension. Aggregation consistency, cross-scale coherence, and uncertainty propagation are enforced within the inference framework rather than evaluated post hoc.

6.5. Recurrent Failure Modes and Practical Guidance

Across data fusion strategies, recurrent failure modes arise from mismatches between observation scale, retrieval formulation, and aggregation operators. Three failure modes are particularly common: (i) training–application scale mismatch in ML models, in which models trained at one resolution are applied at another without accounting for changes in heterogeneity, observation support, or sensor characteristics (see reviews, [20,27]). Such mismatches may arise when training and application domains differ in spatial resolution, footprint composition, or spectral response, causing learned relationships to become scale-dependent. Robust evaluation, therefore, benefits from validation strategies that account for spatial autocorrelation and scale structure (e.g., spatially blocked validation) [112,165,166], as well as explicit assessment of model transferability across sensors and resolutions; (ii) aggregation inconsistency, in which fine-scale reconstructions fail to reproduce original coarse observations when re-aggregated (e.g., [26,29,167]); and (iii) unmodelled anisotropy (BRDF effects), which can alias angular variability into apparent temporal signals (e.g., [21,168]).

7. Evaluation and Validation of Scale-Aware Trait and SIF-Based Dynamics

Assessing the reliability of vegetation trait retrievals and SIF-based products calls for evaluation frameworks that explicitly account for spatial resolution, temporal sampling, aggregation operators, and uncertainty. Because retrievals are nonlinear and landscapes are heterogeneous, differences between products may arise not only from retrieval accuracy but also from resolution mismatch, aggregation effects, and representativeness constraints. In this context, we distinguish between validation and evaluation. Validation refers to comparison with independent reference observations (e.g., field measurements, airborne campaigns, flux towers) to assess physical accuracy. Evaluation is used more broadly and includes internal consistency tests across sensors, resolutions, retrieval frameworks, or processing strategies, particularly when reference data are unavailable or incomplete [28].

A range of complementary evaluation strategies is required to diagnose scale effects. These include: assessment of spatial representativeness between pixel and field measurements, cross-resolution consistency through re-aggregation tests, the temporal consistency of time-series dynamics, and agreement across sensors with differing observation characteristics. In addition, process-level consistency between SIF, vegetation traits, and productivity proxies provides an important constraint on ecological interpretation. Uncertainty evalua-

tion further requires assessing whether predicted uncertainty remains consistent across scales and retrieval formulations.

Table 3 summarises key evaluation targets, associated scale dimensions, comparison approaches, recommended metrics, and common pitfalls. A key insight is that many discrepancies between products arise from scale interactions rather than retrieval errors alone. In particular, aggregation–retrieval non-commutativity, temporal compositing effects, and observation inconsistencies can induce systematic biases in spatial patterns and temporal metrics. As outlined in Table 3, a scale-aware evaluation examines whether inferred vegetation dynamics remain stable under controlled changes in spatial and temporal resolution, and whether derived indicators (e.g., phenology, disturbance timing, productivity estimates) remain consistent across sensors and retrieval formulations. Robust evaluation frameworks should include aggregation-consistency diagnostics, resolution-stratified analysis, and explicit treatment of uncertainty propagation across scales. Ultimately, evaluation should not be interpreted as agreement at a single nominal resolution but as the stability and coherence of inferred vegetation dynamics across admissible changes in the observation and aggregation scale.

https://doi.org/10.3390/rs18121951

Table 3. Scale-aware evaluation and diagnostic strategies for vegetation trait and SIF-based products. The table links evaluation targets to scale dimensions, comparison approaches, recommended metrics, and common failure modes, providing practical diagnostics for identifying scale-induced inconsistencies and representativeness issues across resolutions. It highlights how evaluation frameworks can be used to assess aggregation consistency, transferability, uncertainty behaviour, and cross-scale agreement while recognising that many apparent discrepancies arise from interactions between observation, retrieval, aggregation, and data/compute design rather than from retrieval errors alone.

Evaluation Target	Scale Dimension	What Is Compared	Recommended Metrics	Typical Failure Modes
Spatial representativeness	Spatial (pixel vs. field)	Retrieved traits vs. in situ or upscaled references	RMSE, bias, representativeness error	Footprint mismatch; sub-pixel heterogeneity
Cross-resolution consistency	Spatial + aggregation	Fine-resolution retrievals vs. re-aggregated coarse observations	Correlation, bias, closure error	Ignoring aggregation operators; non-commutativity
Aggregation consistency	Aggregation	Re-aggregated fine-scale estimates vs. original coarse observations	Closure error, conservation metrics	Violation of conservation or aggregation assumptions
Temporal dynamics	Temporal sampling	Time-series trajectories across sensors or products	Correlation, phase shift, amplitude differences	Over-smoothing; temporal aliasing; inconsistent compositing
Phenology metrics	Temporal aggregation	Seasonal indicators (SOS, EOS, peak timing)	Timing error, RMSE, confidence intervals	Sensitivity to smoothing and temporal resolution
Disturbance detection	Spatiotemporal	Disturbance timing and magnitude	Detection delay, magnitude error, detection probability	Mixed pixels; missed short-duration events
Recovery dynamics	Temporal trajectory	Post-disturbance trajectories across products	Recovery rate, trajectory similarity, time-to-recovery	Baseline effects; inconsistent temporal support
Multi-sensor agreement	Observation (spectral + angular)	Products from different sensors	Bias, correlation, uncertainty overlap	Spectral mismatch; BRDF effects; preprocessing differences
SIF–trait/process consistency	Retrieval + process	SIF vs. traits or GPP proxies	Correlation, lag analysis, causality metrics	Scale mismatch; temporal misalignment; confounding drivers
ML-based retrieval behaviour	Training vs. application scale	Predictions across resolutions or domains	Cross-scale stability, calibration, generalisation error	Information leakage; domain shift; scale mismatch
Uncertainty consistency	Multi-scale uncertainty	Predicted vs. propagated uncertainty	Coverage probability, calibration curves, sharpness	Ignoring scale dependence; inconsistent uncertainty propagation
Resolution sensitivity	Spatial + temporal	Controlled resampling experiments	Stability metrics, variance ratios, sensitivity indices	Assuming monotonic accuracy with resolution
Data/compute-induced scale effects	Data + computational	Alternative layouts or processing strategies	Consistency and reproducibility metrics	Hidden scale operators; reproducibility issues

https://doi.org/10.3390/rs18121951

8. Best-Practice Guidelines for Scale-Aware Trait and SIF-Based Dynamics

The preceding sections show that scale effects arise from interacting choices in spatial resolution, temporal integration, retrieval formulation, aggregation operators, and uncertainty modelling. These interactions define the effective observation scale of vegetation products and condition the interpretation of vegetation traits and SIF-based indicators [27]. Because retrievals involve nonlinear RT processes, heterogeneous landscapes, and model-dependent inference, finer observations often expose additional sources of variability, heterogeneity, and uncertainty rather than yielding proportionally more reliable ecological information [26]. Rather than simply improving retrieval accuracy, changes in resolution modify the conditioning of the inverse problem and redistribute uncertainty across measurement noise, structural mismatch, model approximation, and representativeness components [21]. Consequently, robust vegetation monitoring calls for scale-aware strategies for retrieval design, evaluation, and interpretation.

Robust quantitative vegetation monitoring thus depends on explicit alignment between observation resolution, retrieval model structure, aggregation operators, and uncertainty treatment throughout the processing chain. Practical guidance for implementing such scale-aware workflows can be drawn from EO validation frameworks [28], uncertainty propagation theory (see review, [150]), and vegetation retrieval and scaling studies (see review, [6,20]). Operationally, scale-aware retrieval aims to maintain consistency between three interacting elements: (i) the spatial and temporal resolution of the observations, (ii) the structural assumptions embedded in the retrieval model, and (iii) the ecological scale at which the vegetation trait is defined. As addressed before, the effective scale of a vegetation product is the scale at which these elements are mutually consistent. When observation scale, retrieval formulation, and trait-definition scale are misaligned, apparent spatial patterns and temporal dynamics may reflect resolution artefacts rather than ecosystem processes.

In line with this framework, Table 4 summarises best-practice guidelines for scale-aware analysis of vegetation traits and SIF-based products. The table links the scale dimensions introduced in Section 2 with practical workflow recommendations and highlights common pitfalls that may introduce resolution-induced artefacts. Taken together, the guidelines emphasise that robust analysis requires explicit alignment between observation characteristics, retrieval formulation, and aggregation operators and that many apparent product differences can be traced back to inconsistencies in how these elements are combined. In particular, failures to account for scale interactions often manifest as biases in spatial patterns, temporal dynamics, and cross-sensor comparability.

https://doi.org/10.3390/rs18121951

Table 4. Best-practice guidelines for scale-aware analysis of vegetation traits and SIF-based products. The table summarises recommended practices, the scale dimensions addressed (cf. Section 2), and typical pitfalls that introduce scale-induced artefacts. It provides operational recommendations for improving the interpretation, validation, comparison, and application of vegetation trait and SIF products across spatial, temporal, spectral, and retrieval scales.

Guideline Domain	Scale Dimension	Recommended Practice	Typical Pitfalls
Observation definition	Observation scale	Characterise footprint, PSF, SRF, temporal integration, and angular effects before analysis.	Treating pixel size as effective resolution; ignoring PSF or compositing effects
Trait definition and retrieval scale	Trait-definition and retrieval scales	Match retrieval outputs to the intended ecological interpretation and distinguish structural from effective quantities.	Mixing incompatible variables or definitions
Aggregation and nonlinearity	Aggregation operators; spatial scale	Evaluate aggregation consistency and test retrieve–aggregate versus aggregate–retrieve workflows.	Assuming commutativity; neglecting sub-pixel heterogeneity
Model complexity and identifiability	Retrieval scale; model complexity	Match model complexity to information content and apply appropriate constraints.	Overparameterisation; equifinality
Temporal consistency	Temporal observation and aggregation	Harmonise temporal sampling and compositing windows before comparing products.	Incompatible temporal supports or smoothing strategies
Effective scale alignment	Observation, retrieval, and trait-definition scales	Assess cross-resolution stability and aggregation consistency to diagnose effective scale.	Assuming finer resolution guarantees better interpretation
Uncertainty propagation	Uncertainty scaling; effective scale	Propagate uncertainties across retrieval and aggregation steps and distinguish aleatoric from epistemic uncertainty.	Ignoring scale dependence or mixing uncertainty sources
Validation and representativeness	Representativeness; effective scale	Account for spatial and temporal representativeness when comparing with reference data.	Direct point-to-pixel comparisons
ML training and evaluation	Training-domain and retrieval scales	Use spatially and temporally stratified validation and assess cross-scale robustness.	Information leakage; poor transferability
Reproducibility and provenance	Processing scale; effective scale	Document preprocessing, compositing, and retrieval settings and maintain traceable workflows.	Undocumented processing changes; artificial trends

https://doi.org/10.3390/rs18121951

9. Outlook: Research Priorities for Scale-Aware Vegetation Dynamics in the CHIME and FLEX Era

Overall, scale-aware interpretation is not only a methodological necessity but also a major challenge for next-generation vegetation monitoring. Upcoming hyperspectral missions such as CHIME and FLEX will substantially expand spectral information content and temporal sampling, enabling improved inference of vegetation biochemical traits and photosynthetic function (e.g., [9,169,170]). However, as discussed throughout this review, increased spatial, spectral, or temporal detail does not necessarily translate into improved ecological accuracy. Changes in the observation scale alter the conditioning of the inverse problem and the identifiability of retrieved variables and redistribute uncertainty across measurement, structural, algorithmic, and representativeness components (see reviews, [20,27]).

In the CHIME/FLEX era, the challenge lies not only in increased information content but also in interpreting vegetation traits and SIF-based products within an explicit scale- and uncertainty-aware framework. The hyperspectral era shifts the bottleneck from data availability to scale-consistent inference, where interpretation is governed by the representation of observation characteristics, retrieval assumptions, and aggregation processes. Addressing this challenge calls for retrieval strategies, evaluation protocols, and data standards that explicitly account for spatial, temporal, and spectral resolution, together with the aggregation operators and uncertainty structures linking observations across scales. Guidance can be drawn from EO validation frameworks [28], ARD concepts [36], and advances in imaging spectroscopy retrieval methods (see reviews, [2,5,39]).

Looking ahead reveals that a key methodological priority is the explicit optimisation of the effective scale of retrievals. Rather than maximising nominal resolution, CHIME/FLEX-era workflows should align observation resolution (including its effective spatial and temporal support), retrieval assumptions, and trait-definition scale to ensure that inferred vegetation dynamics remain stable under admissible changes in spatial resolution and temporal aggregation. Following this rationale, Figure 6 organises future research directions around six scale-related questions that emerge from this review. These concern: (i) how ARD standards can better document scale characteristics and processing assumptions; (ii) how spectral resolution influences trait identifiability and retrieval stability; (iii) how retrieval performance should be evaluated across spatial, temporal, and spectral supports; (iv) how observations acquired at different resolutions can be integrated through joint multiresolution inference frameworks; (v) how ML retrievals generalise across resolutions and observation domains; and (vi) how uncertainty propagates into dynamic indicators such as phenology, disturbance metrics, and ecosystem trends. For each question, Figure 6 highlights methodological directions that explicitly incorporate scale behaviour into retrieval design, validation, data fusion, and interpretation.

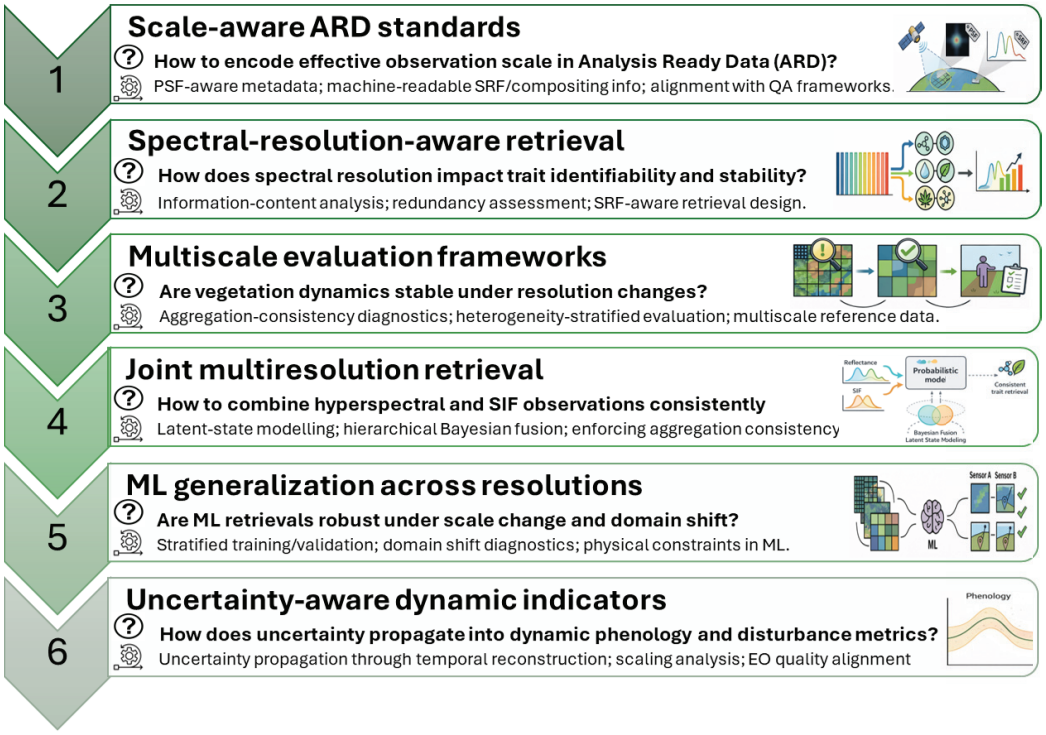


Figure 6. Research roadmap for scale-aware vegetation and SIF monitoring in the CHIME and FLEX era. The figure summarises key research priorities, including scale-aware ARD standards, spectral-resolution-aware retrieval, multiscale evaluation, joint multiresolution modelling, ML generalisation across scales, and uncertainty-aware dynamic indicators. Each priority is linked to guiding questions and associated methodological directions: information-content analysis, SRF-aware retrieval design, heterogeneity-stratified validation, hierarchical Bayesian fusion, domain-shift diagnostics, and uncertainty propagation via temporal reconstruction. The roadmap highlights that robust multi-sensor integration and consistent interpretation of vegetation dynamics require explicit treatment of observation, retrieval, aggregation, and uncertainty scales across the EO–model chain.

10. Conclusions

This review has framed satellite-based retrieval of vegetation traits and SIF-based products as an explicitly multiscale retrieval problem. Observation characteristics (spatial resolution, temporal integration, spectral response), retrieval formulation, trait definition, and aggregation operators jointly determine the scale at which vegetation structure and function are represented. Across spatial, temporal, and spectral domains, scale effects emerge as structural properties of the observation–retrieval system, arising from scene heterogeneity, nonlinear RT processes, and aggregation operators.

Within this perspective, three main contributions are synthesised: (i) a formalisation of effective scale arising from the combined effects of observation, retrieval, and aggregation processes; (ii) a unifying interpretation of the resolution–accuracy paradox, showing that increasing observational detail can alter the retrieval problem rather than improve its solution, thereby synthesising resolution effects, mixed-pixel behaviour, aggregation bias, and non-commutativity within a common conceptual framework; and (iii) a scale-aware diagnostic and methodological framework linking EO retrievals, aggregation operators, and process-level modelling. A key insight is that ecological accuracy is not a monotonic function of spatial, spectral, or temporal detail. Changes in resolution or integration windows alter the conditioning of the inverse problem and redistribute errors among measurement noise, structural mismatch, model approximation, and representativeness components. As a result, vegetation trait and SIF-based products cannot be interpreted independently of their effective scale.

https://doi.org/10.3390/rs18121951

These considerations have direct implications for evaluation and application. Ecological validity should be assessed through aggregation-consistent diagnostics and the stability of inferred behaviour under controlled changes in spatial and temporal scale, rather than agreement at a single nominal resolution. In particular, metrics derived from time series (e.g., phenology, SIF dynamics, disturbances, and trends) must be interpreted in a scale-consistent manner, as their values depend on temporal sampling, compositing, and aggregation choices.

For users of vegetation trait and SIF-based products, several practical implications follow from this framework. Effective scale should be explicitly reported alongside nominal spatial and temporal resolution, aggregation consistency should be evaluated through closure tests or cross-scale comparisons where possible, uncertainty should be propagated and interpreted across scales rather than at a single resolution, and validation should account for representativeness and scale mismatch. These practices help distinguish ecological signals from artefacts introduced by observation support, retrieval formulation, and aggregation processes.

Altogether, explicit modelling, testing, and reporting of effective scale should become standard practice in quantitative vegetation EO. In the era of hyperspectral missions such as CHIME and FLEX, progress will hinge not only on increased observational capability but also on the community's ability to interpret vegetation dynamics within scale-aware and uncertainty-aware retrieval frameworks. In this perspective, scale is not a limitation to be corrected, but a defining dimension of the retrieval problem that must be explicitly modelled, evaluated, and interpreted.

Author Contributions: Conceptualisation, J.V.; methodology, J.V.; investigation, J.V.; resources, J.V.; writing—original draft preparation, J.V.; writing—review and editing, J.V., B.V. and P.R.-M.; visualisation, J.V. and B.V.; supervision, J.V.; funding acquisition, J.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Research Council (ERC) under the FLEXINEL project: grant number 101086622. The views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Data Availability Statement: No new data were created or analysed in this study. Data sharing is not applicable to this article.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (v. GPT-5.4, OpenAI) for drafting explanatory text and refining technical descriptions and also used generative AI tools to assist in the creation of figures. The authors have reviewed and edited all outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Frappart, F.; Ramillien, G. Monitoring of the Terrestrial Vegetation Dynamics from Satellite Remote Sensing: A Review of Vegetation Indices. *Remote Sens.* **2020**, *12*, 2915. [CrossRef]
2. Verrelst, J.; Malenovský, Z.; van der Tol, C.; Camps-Valls, G.; Gastellu-Etchegorry, J.P.; Lewis, P.; North, P.; Moreno, J. Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods. *Surv. Geophys.* **2019**, *40*, 589–629. [CrossRef]
3. Zeng, L.; Wardlow, B.D.; Xiang, D.; Hu, S.; Li, D. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sens. Environ.* **2020**, *237*, 111511. [CrossRef]
4. Mohammed, G.H.; Colombo, R.; Middleton, E.M.; Rascher, U.; van der Tol, C.; Nedbal, L.; Goulas, Y.; Pérez-Priego, O.; Damm, A.; Meroni, M.; et al. Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. *Remote Sens. Environ.* **2019**, *231*, 111177. [CrossRef] [PubMed]

5. Verrelst, J.; Kovács, D.D.; Rivera-Caicedo, J.P. *Vegetation Trait Mapping with Optical Remote Sensing: Recent Advances in Methods and Applications*; Elsevier: Amsterdam, The Netherlands, 2026. [CrossRef]
6. Verrelst, J.; Camps-Valls, G.; Muñoz-Marí, J.; Rivera, J.P.; Veroustraete, F.; Clevers, J.G.; Moreno, J. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties—A review. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 273–290. [CrossRef]
7. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, 111402. [CrossRef]
8. Homolová, L.; Malenovský, Z.; Clevers, J.G.P.W.; García-Santos, G.; Schaepman, M.E. Review of optical-based remote sensing for plant trait mapping. *Ecol. Complex.* **2013**, *15*, 1–16. [CrossRef]
9. Cogliati, S.; Verhoef, W.; Kraft, S.; Sabater, N.; Alonso, L.; Vicent, J.; Moreno, J.; Drusch, M.; Colombo, R. Retrieval of sun-induced fluorescence using advanced spectral fitting methods. *Remote Sens. Environ.* **2015**, *169*, 344–357. [CrossRef]
10. Damm, A.; Guanter, L.; Paul-Limoges, E.; van der Tol, C.; Hueni, A.; Buchmann, N.; Eugster, W.; Ammann, C.; Schaepman, M.E. Far-red sun-induced chlorophyll fluorescence shows ecosystem-specific relationships to gross primary production: An assessment based on observational and modeling approaches. *Remote Sens. Environ.* **2015**, *166*, 91–105. [CrossRef]
11. Sun, Y.; Frankenberg, C.; Jung, M.; Joiner, J.; Guanter, L.; Köhler, P.; Magney, T. Overview of solar-induced chlorophyll fluorescence (SIF) from the Orbiting Carbon Observatory-2: Retrieval, cross-mission comparison, and global monitoring for GPP. *Remote Sens. Environ.* **2018**, *209*, 808–823. [CrossRef]
12. Coppo, P.; Taiti, A.; Pettinato, L.; Francois, M.; Taccola, M.; Drusch, M. Fluorescence imaging spectrometer (FLORIS) for ESA FLEX mission. *Remote Sens.* **2017**, *9*, 649. [CrossRef]
13. Drusch, M.; Moreno, J.; Del Bello, U.; Franco, R.; Goulas, Y.; Huth, A.; Kraft, S.; Middleton, E.M.; Miglietta, F.; Mohammed, G.; et al. The fluorescence explorer mission concept—ESA's earth explorer 8. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 1273–1284. [CrossRef]
14. Du, S.; Liu, L.; Liu, X.; Zhang, X.; Gao, X.; Wang, W. The solar-induced chlorophyll fluorescence imaging spectrometer (SIFIS) onboard the first terrestrial ecosystem carbon inventory satellite (TECIS-1): Specifications and prospects. *Sensors* **2020**, *20*, 815. [CrossRef]
15. Bacour, C.; Maignan, F.; Peylin, P.; Macbean, N.; Bastrikov, V.; Joiner, J.; Köhler, P.; Guanter, L.; Frankenberg, C. Differences between OCO-2 and GOME-2 SIF products from a model-data fusion perspective. *J. Geophys. Res. Biogeosci.* **2019**, *124*, 3143–3157. [CrossRef]
16. Nieke, J.; Despoisse, L.; Gabriele, A.; Weber, H.; Strese, H.; Ghasemi, N.; Gascon, F.; Alonso, K.; Boccia, V.; Tsonevska, B.; et al. The Copernicus Hyperspectral Imaging Mission for the Environment (CHIME): An Overview of Its Mission, System and Planning Status. In Proceedings of the Sensors, Systems, and Next-Generation Satellites XXVII, Amsterdam, The Netherlands, 3–6 September 2023. [CrossRef]
17. García-Soria, J.L.; Morata, M.; Berger, K.; Pascual-Venteo, A.B.; Rivera-Caicedo, J.P.; Verrelst, J. Evaluating epistemic uncertainty estimation strategies in vegetation trait retrieval using hybrid models and imaging spectroscopy data. *Remote Sens. Environ.* **2024**, *310*, 114228. [CrossRef]
18. Jacquemoud, S.; Verhoef, W.; Baret, F.; Bacour, C.; Zarco-Tejada, P.; Asner, G.; François, C.; Ustin, S. PROSPECT + SAIL Models: A Review of Use for Vegetation Characterization. *Remote Sens. Environ.* **2009**, *113*, S56–S66. [CrossRef]
19. van der Tol, C.; Verhoef, W.; Timmermans, J.; Verhoef, A.; Su, Z. An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. *Biogeosciences* **2009**, *6*, 3109–3129. [CrossRef]
20. Malenovský, Z.; Homolová, L.; Lukeš, P.; Buddenbaum, H.; Verrelst, J.; Alonso, L.; Schaepman, M.E.; Lauret, N.; Gastellu-Etchegorry, J.P. Variability and Uncertainty Challenges in Scaling Imaging Spectroscopy Retrievals and Validations from Leaves Up to Vegetation Canopies. *Surv. Geophys.* **2019**, *40*, 631–656. [CrossRef] [PubMed]
21. Pinty, B.; Lavergne, T.; Dickinson, R.E.; Widlowski, J.; Gobron, N.; Verstraete, M.M. Simplifying the interaction of land surfaces with radiation for relating remote sensing products to climate models. *J. Geophys. Res. Atmos.* **2006**, *111*, D02116. [CrossRef]
22. Jönsson, P.; Eklundh, L. TIMESAT—a program for analyzing time-series of satellite sensor data. *Comput. Geosci.* **2004**, *30*, 833–845. [CrossRef]
23. Kennedy, R.E.; Yang, Z.; Cohen, W.B. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. *Remote Sens. Environ.* **2010**, *114*, 2897–2910. [CrossRef]
24. Verbesselt, J.; Hyndman, R.; Newnham, G.; Culvenor, D. Detecting trend and seasonal changes in satellite image time series. *Remote Sens. Environ.* **2010**, *114*, 106–115. [CrossRef]
25. Belda, S.; Pipia, L.; Morcillo-Pallarés, P.; Rivera-Caicedo, J.P.; Amin, E.; De Grave, C.; Verrelst, J. DATimeS: A machine learning time series GUI toolbox for gap-filling and vegetation phenology trends detection. *Environ. Model. Softw.* **2020**, *127*, 104666. [CrossRef] [PubMed]
26. Atkinson, P.M.; Curran, P.J. Defining an optimal size of support for remote sensing investigations. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 768–776. [CrossRef]

27. Wu, H.; Li, Z. Scale Issues in Remote Sensing: A Review on Analysis, Processing and Modeling. *Sensors* **2009**, *9*, 1768–1793. [CrossRef]
28. Loew, A.; Bell, W.; Brocca, L.; Bulgin, C.E.; Burdanowitz, J.; Calbet, X.; Donner, R.V.; Gelaro, R.; Ghent, D.; Gruber, A.; et al. Validation practices for satellite-based Earth observation data across communities. *Rev. Geophys.* **2017**, *55*, 779–817. [CrossRef]
29. Woodcock, C.E.; Strahler, A.H. The Factor of Scale in Remote Sensing. *Remote Sens. Environ.* **1987**, *21*, 311–332. [CrossRef]
30. Garrigues, S.; Allard, D.; Baret, F.; Weiss, M. Influence of landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data. *Remote Sens. Environ.* **2006**, *105*, 286–298. [CrossRef]
31. Zhang, X.; Wang, J.; Gao, F.; Liu, Y.; Schaaf, C.; Friedl, M.; Yu, Y.; Jayavelu, S.; Gray, J.; Liu, L.; et al. Exploration of scaling effects on coarse resolution land surface phenology. *Remote Sens. Environ.* **2017**, *190*, 318–330. [CrossRef]
32. Cui, T.; Martz, L.; Zhao, L.; Guo, X. Investigating the impact of the temporal resolution of MODIS vegetation indices on land surface phenology estimation. *GISci. Remote Sens.* **2020**, *57*, 395–410. [CrossRef]
33. Jin, S.; Sader, S.A. MODIS time-series imagery for forest disturbance detection and quantification of patch size effects. *Remote Sens. Environ.* **2005**, *99*, 462–470. [CrossRef]
34. Schott, J.R.; Gerace, A.; Woodcock, C.E.; Wang, S.; Zhu, Z.; Wynne, R.H.; Blinn, C.E. The impact of improved signal-to-noise ratios on algorithm performance: Case studies for Landsat class instruments. *Remote Sens. Environ.* **2016**, *185*, 37–45. [CrossRef]
35. Tan, B.; Woodcock, C.E.; Hu, J.; Zhang, P.; Ozdogan, M.; Huang, D.; Yang, W.; Knyazikhin, Y.; Myneni, R.B. The impact of gridding artifacts on the local spatial properties of MODIS data. *Remote Sens. Environ.* **2006**, *105*, 98–114. [CrossRef]
36. Dwyer, J.L.; Roy, D.P.; Sauer, B.; Jenkerson, C.B.; Zhang, H.K.; Lymburner, L. Analysis Ready Data: Enabling Analysis of the Landsat Archive. *Remote Sens.* **2018**, *10*, 1363. [CrossRef]
37. Aplin, P. On Scales and Dynamics in Observing the Environment. *Int. J. Remote Sens.* **2006**, *27*, 2123–2140. [CrossRef]
38. Ge, Y.; Jin, Y.; Stein, A.; Chen, Y.; Wang, J.; Wang, J.; Cheng, Q.; Bai, H.; Liu, M.; Atkinson, P.M. Principles and methods of scaling geospatial Earth science data. *Earth-Sci. Rev.* **2019**, *197*, 102897. [CrossRef]
39. Verrelst, J.; Zhang, Y.; Morata, M.; De Clerck, E.; Liu, L. Machine Learning for Satellite Solar-Induced Fluorescence: Retrieval, Reconstruction, Downscaling, and Applications. *Remote Sens.* **2026**, *18*, 553. [CrossRef]
40. Duggin, M.; Robinove, C. Assumptions implicit in remote sensing data acquisition and analysis. *Remote Sens.* **1990**, *11*, 1669–1694. [CrossRef]
41. Schowengerdt, R.A. *Remote Sensing: Models and Methods for Image Processing*, 3rd ed.; Academic Press: Burlington, MA, USA, 2007.
42. Cracknell, A.P. Review article: Synergy in remote sensing—What's in a pixel? *Int. J. Remote Sens.* **1998**, *19*, 2025–2047. [CrossRef]
43. Duveiller, G.; Defourny, P. A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. *Remote Sens. Environ.* **2010**, *114*, 2637–2650. [CrossRef]
44. Huang, C.; Townshend, J.R.G.; Liang, S.; Kalluri, S.N.V.; DeFries, R.S. Impact of sensor's point spread function on land cover characterization: Assessment and deconvolution. *Remote Sens. Environ.* **2002**, *80*, 203–212. [CrossRef]
45. Holben, B.N. Characteristics of maximum-value composite images from temporal AVHRR data. *Int. J. Remote Sens.* **1986**, *7*, 1417–1434. [CrossRef]
46. Kokaly, R.F.; Clark, R.N. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sens. Environ.* **1999**, *67*, 267–287. [CrossRef]
47. Thenkabail, P.S.; Smith, R.B.; De Pauw, E. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens. Environ.* **2000**, *71*, 158–182. [CrossRef]
48. Verrelst, J.; Schaepman, M.E.; Koetz, B.; Kneubühler, M. Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data. *Remote Sens. Environ.* **2008**, *112*, 2341–2353. [CrossRef]
49. Tian, H.; Gao, Y.; Zhang, Y.; Li, H.; Verrelst, J.; Zeng, Y. Quantifying angular sensitivity and synergistic effects of SIF and vegetation indices for robust winter wheat yield estimation. *Int. J. Appl. Earth Obs. Geoinf.* **2026**, *148*, 105235. [CrossRef]
50. Group on Earth Observations. *A Quality Assurance Framework for Earth Observation (QA4EO): Principles and Guidelines*; QA4EO Guideline Document (Report); Group on Earth Observations: Geneva, Switzerland, 2010.
51. Jacquemoud, S.; Baret, F. PROSPECT: A Model of Leaf Optical Properties Spectra. *Remote Sens. Environ.* **1990**, *34*, 75–91. [CrossRef]
52. Féret, J.B.; François, C.; Asner, G.; Gitelson, A.; Martin, R.; Bidel, L.; Ustin, S.; le Maire, G.; Jacquemoud, S. PROSPECT-4 and 5: Advances in the Leaf Optical Properties Model Separating Photosynthetic Pigments. *Remote Sens. Environ.* **2008**, *112*, 3030–3043. [CrossRef]
53. Berger, K.; Atzberger, C.; Danner, M.; D'Urso, G.; Mauser, W.; Vuolo, F.; Hank, T. Evaluation of the PROSAIL model capabilities for future hyperspectral model environments: A review study. *Remote Sens.* **2018**, *10*, 85. [CrossRef]
54. Bacour, C.; Baret, F.; Béal, D. Neural Network Estimation of LAI, fAPAR, fCover and LAI×Cab, from Top of Canopy MERIS Reflectance Data: Principles and Validation. *Remote Sens. Environ.* **2006**, *105*, 313–325. [CrossRef]
55. Clevers, J.; Kooistra, L.; van den Brande, M. Using Sentinel-2 Data for Retrieving LAI and Leaf and Canopy Chlorophyll Content of a Potato Crop. *Remote Sens.* **2017**, *9*, 405. [CrossRef]

56. Jacquemoud, S.; Bacour, C.; Poilvé, H.; Frangi, J.P. Comparison of Four Radiative Transfer Models to Simulate Plant Canopies Reflectance: Direct and Inverse Mode. *Remote Sens. Environ.* **2000**, *74*, 471–481. [CrossRef]
57. Darvishzadeh, R.; Skidmore, A.; Schlerf, M.; Atzberger, C. Inversion of a Radiative Transfer Model for Estimating Vegetation LAI and Chlorophyll in a Heterogeneous Grassland. *Remote Sens. Environ.* **2008**, *112*, 2592–2604. [CrossRef]
58. Zurita-Milla, R.; Laurent, V.; van Gijsel, J. Visualizing the Ill-Posedness of the Inversion of a Canopy Radiative Transfer Model: A Case Study for Sentinel-2. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *43*, 7–18. [CrossRef]
59. Combal, B.; Baret, F.; Weiss, M.; Trubuil, A.; Macé, D.; Pragnère, A.; Myneni, R.B.; Knyazikhin, Y.; Wang, L. Retrieval of Canopy Biophysical Variables from Bidirectional Reflectance: Using Prior Information to Solve the Ill-Posed Inverse Problem. *Remote Sens. Environ.* **2003**, *84*, 1–15. [CrossRef]
60. Pinty, B.; Gobron, N.; Widlowski, J.L.; Lavergne, T.; Verstraete, M. Synergy between 1-D and 3-D Radiation Transfer Models to Retrieve Vegetation Canopy Properties from Remote Sensing Data. *J. Geophys. Res. Atmos.* **2004**, *109*, D21205. [CrossRef]
61. Widlowski, J.L.; Taberner, M.; Pinty, B.; Bruniquel-Pinel, V.; Disney, M.; Fernandes, R.; Gastellu-Etchegorry, J.P.; Gobron, N.; Kuusk, A.; Lavergne, T.; et al. The Third RAdiation Transfer Model Intercomparison (RAMI) Exercise: Documenting Progress in Canopy Reflectance Models. *J. Geophys. Res. Atmos.* **2007**, *112*, D09111. [CrossRef]
62. Berger, K.; Rivera Caicedo, J.P.; Martino, L.; Wocher, M.; Hank, T.; Verrelst, J. A Survey of Active Learning for Quantifying Vegetation Traits from Terrestrial Earth Observation Data. *Remote Sens.* **2021**, *13*, 287. [CrossRef] [PubMed]
63. Chen, J.; Black, T. Defining Leaf Area Index for Non-Flat Leaves. *Plant Cell Environ.* **1992**, *15*, 421–429. [CrossRef]
64. Weiss, M.; Baret, F.; Smith, G.; Jonckheere, I.; Coppin, P. Review of Methods for In Situ Leaf Area Index (LAI) Determination: Part II. Estimation of LAI, Errors and Sampling. *Agric. For. Meteorol.* **2004**, *121*, 37–53. [CrossRef]
65. Zheng, G.; Moskal, L.M. Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors. *Sensors* **2009**, *9*, 2719–2745. [CrossRef]
66. Oliphant, A.J. Terrestrial ecosystem-atmosphere exchange of CO_2, water and energy from FLUXNET; review and meta-analysis of a global in-situ observatory. *Geogr. Compass* **2012**, *6*, 689–705. [CrossRef]
67. Duveiller, G.; Cescatti, A. Spatially Downscaling Sun-Induced Chlorophyll Fluorescence Leads to an Improved Temporal Correlation with Gross Primary Productivity. *Remote Sens. Environ.* **2016**, *182*, 72–89. [CrossRef]
68. Liang, S. *Quantitative Remote Sensing of Land Surfaces*; John Wiley & Sons: Hoboken, NJ, USA, 2003. [CrossRef]
69. Markham, K.; Frazier, A.E.; Singh, K.K.; Madden, M. A review of methods for scaling remotely sensed data for spatial pattern analysis. *Landsc. Ecol.* **2023**, *38*, 619–635. [CrossRef]
70. Verrelst, J.; Clerck, E.D.; Verma, B.; Mishra, K.; Caballero, G. Cloud-Native Earth Observation for Quantitative Vegetation Science: Architectures, Workflows, and Scientific Implications. *Remote Sens.* **2026**, *18*, 1154. [CrossRef]
71. Lewis, A.; Oliver, S.; Lymburner, L.; Evans, B.; Wyborn, L.; Mueller, N.; Raevksi, G.; Hooke, J.; Woodcock, R.; Sixsmith, J.; et al. The Australian geoscience data cube—foundations and lessons learned. *Remote Sens. Environ.* **2017**, *202*, 276–292. [CrossRef]
72. Sudmanns, M.; Augustin, H.; Killough, B.; Giuliani, G.; Tiede, D.; Leith, A.; Yuan, F.; Lewis, A. Think global, cube local: An Earth Observation Data Cube's contribution to the Digital Earth vision. *Big Earth Data* **2023**, *7*, 831–859. [CrossRef]
73. Hoyer, S.; Hamman, J. xarray: ND labeled arrays and datasets in Python. *J. Open Res. Softw.* **2017**, *5*, 10. [CrossRef]
74. Rocklin, M. Dask: Parallel Computation with Blocked algorithms and Task Scheduling. In Proceedings of the 14th Python in Science Conference, Austin, TX, USA, 6–12 July 2015. [CrossRef]
75. De Grave, C.; Pipia, L.; Siegmann, B.; Morcillo-Pallarés, P.; Rivera-Caicedo, J.P.; Moreno, J.; Verrelst, J. Retrieving and validating leaf and canopy chlorophyll content at moderate resolution: A multiscale analysis with the sentinel-3 OLCI sensor. *Remote Sens.* **2021**, *13*, 1419. [CrossRef]
76. Rossini, M.; Celesti, M.; Bramati, G.; Migliavacca, M.; Cogliati, S.; Rascher, U.; Colombo, R. Evaluation of the Spatial Representativeness of In Situ SIF Observations for the Validation of Medium-Resolution Satellite SIF Products. *Remote Sens.* **2022**, *14*, 5107. [CrossRef]
77. Cremer, N.; Alonso, K.; Doxani, G.; Chlus, A.; Thompson, D.; Brodrick, P.; Townsend, P.; Palombo, A.; Santini, F.; Gao, B.C.; et al. Atmospheric Correction Inter-Comparison eXercise, ACIX-III Land: An Assessment of Atmospheric Correction Processors for EnMAP and PRISMA over Land. *Remote Sens.* **2025**, *17*, 3790. [CrossRef]
78. He, H.S.; DeZonia, B.E.; Mladenoff, D.J. An Aggregation Index (AI) to Quantify Spatial Patterns of Landscapes. *Landsc. Ecol.* **2000**, *15*, 591–601. [CrossRef]
79. Ma, J.; Zhou, J.; Liu, S.; Göttsche, F.M.; Zhang, X.; Wang, S.; Li, M. Continuous Evaluation of the Spatial Representativeness of Land Surface Temperature Validation Sites. *Remote Sens. Environ.* **2021**, *265*, 112669. [CrossRef]
80. Moran, P.A. Notes on continuous stochastic phenomena. *Biometrika* **1950**, *37*, 17–23. [CrossRef]
81. Jantol, N.; Prikaziuk, E.; Celesti, M.; Hernandez-Sequeira, I.; Tomelleri, E.; Pacheco-Labrador, J.; Van Wittenberghe, S.; Pla, F.; Bandopadhyay, S.; Koren, G.; et al. Using Sentinel-2-based metrics to characterize the spatial heterogeneity of FLEX sun-induced chlorophyll fluorescence on sub-pixel scale. *Remote Sens.* **2023**, *15*, 4835. [CrossRef]

82. Verrelst, J.; Schaepman, M.E.; Malenovskỳ, Z.; Clevers, J.G. Effects of woody elements on simulated canopy reflectance: Implications for forest chlorophyll content retrieval. *Remote Sens. Environ.* **2010**, *114*, 647–656. [CrossRef]
83. Monitoring and assessing of landscape heterogeneity at different scales. *Environ. Monit. Assess.* **2013**, *185*, 9419–9434. [CrossRef]
84. Tao, X.; Yan, B.; Wang, K.; Wu, D.; Fan, W.; Xu, X.; Liang, S. Scale transformation of Leaf Area Index product retrieved from multiresolution remotely sensed data: Analysis and case studies. *Int. J. Remote Sens.* **2009**, *30*, 5383–5395. [CrossRef]
85. Xu, X.; Fan, W.; Tao, X. The spatial scaling effect of continuous canopy Leaves Area Index retrieved by remote sensing. *Sci. China Earth Sci.* **2009**, *52*, 393–401. [CrossRef]
86. Chen, J. Spatial scaling of a remotely sensed surface parameter by contexture. *Remote Sens. Environ.* **1999**, *69*, 30–42. [CrossRef]
87. Clevers, J.G.P.W.; Kooistra, L.; Schaepman, M.E. Estimating canopy water content using hyperspectral remote sensing data. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, 119–125. [CrossRef]
88. Gitelson, A.A.; Viña, A.; Ciganda, V.; Rundquist, D.C.; Arkebauer, T.J. Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* **2005**, *32*, L08403. [CrossRef]
89. Inoue, Y.; Guérif, M.; Baret, F.; Skidmore, A.; Gitelson, A.; Schlerf, M.; Darvishzadeh, R.; Olioso, A. Simple and robust methods for remote sensing of canopy chlorophyll content: A comparative analysis of hyperspectral data for different types of vegetation. *Plant Cell Environ.* **2016**, *39*, 2609–2623. [CrossRef] [PubMed]
90. Li, Z.; Jin, X.; Yang, G.; Drummond, J.; Yang, H.; Clark, B.; Li, Z.; Zhao, C. Remote Sensing of Leaf and Canopy Nitrogen Status in Winter Wheat (Triticum aestivum L.) Based on N-PROSAIL Model. *Remote Sens.* **2018**, *10*, 1463. [CrossRef]
91. Brown, L.A.; Ogutu, B.O.; Dash, J. Estimating Forest Leaf Area Index and Canopy Chlorophyll Content with Sentinel-2: An Evaluation of Two Hybrid Retrieval Algorithms. *Remote Sens.* **2019**, *11*, 1752. [CrossRef]
92. Verrelst, J.; Rivera-Caicedo, J.P.; Reyes-Muñoz, P.; Morata, M.; Amin, E.; Tagliabue, G.; Panigada, C.; Hank, T.; Berger, K. Mapping landscape canopy nitrogen content from space using PRISMA data. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 382–395. [CrossRef]
93. Bøegh, E.; Houborg, R.; Bienkowski, J.; Braban, C.F.; Dalgaard, T.; van Dijk, N.; Dragosits, U.; Holmes, E.; Finch, J.W.; Glud, R.N.; et al. Remote sensing of LAI, chlorophyll and leaf nitrogen pools of crop- and grasslands in five European landscapes. *Biogeosciences* **2013**, *10*, 6279–6307. [CrossRef]
94. Delloye, C.; Weiss, M.; Defourny, P. Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. *Remote Sens. Environ.* **2018**, *216*, 245–261. [CrossRef]
95. Verhoef, W. Light Scattering by Leaf Layers with Application to Canopy Reflectance Modeling: The SAIL Model. *Remote Sens. Environ.* **1984**, *16*, 125–141. [CrossRef]
96. Gastellu-Etchegorry, J.P.; Demarez, V.; Pinel, V.; Zagolski, F. Modeling Radiative Transfer in Heterogeneous 3-D Vegetation Canopies. *Remote Sens. Environ.* **1996**, *58*, 131–156. [CrossRef]
97. Nilson, T. A Theoretical Analysis of the Frequency of Gaps in Plant Stands. *Agric. Meteorol.* **1971**, *8*, 25–38. [CrossRef]
98. Frankenberg, C.; Butz, A.; Toon, G.C. Disentangling chlorophyll fluorescence from atmospheric scattering effects in O_2 A-band spectra of reflected sunlight. *Geophys. Res. Lett.* **2011**, *38*, L03801. [CrossRef]
99. Verrelst, J.; Rivera, J.P.; van der Tol, C.; Magnani, F.; Mohammed, G.; Moreno, J. Global sensitivity analysis of the SCOPE model: What drives simulated canopy-leaving sun-induced fluorescence? *Remote Sens. Environ.* **2015**, *166*, 8–21. [CrossRef]
100. Duveiller, G.; Filipponi, F.; Walther, S.; Köhler, P.; Frankenberg, C.; Guanter, L.; Cescatti, A. A spatially downscaled sun-induced fluorescence global product for enhanced monitoring of vegetation productivity. *Earth Syst. Sci. Data* **2020**, *12*, 1101–1116. [CrossRef]
101. Foody, G.M.; Mathur, A. Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification. *Remote Sens. Environ.* **2004**, *93*, 107–117. [CrossRef]
102. Verrelst, J.; Muñoz, J.; Alonso, L.; Delegido, J.; Rivera, J.P.; Camps-Valls, G.; Moreno, J. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and-3. *Remote Sens. Environ.* **2012**, *118*, 127–139. [CrossRef]
103. Tian, Y.; Woodcock, C.E.; Wang, Y.; Privette, J.L.; Shabanov, N.V.; Zhou, L.; Zhang, Y.; Buermann, W.; Dong, J.; Veikkanen, B.; et al. Multiscale analysis and validation of the MODIS LAI product: I. Uncertainty assessment. *Remote Sens. Environ.* **2002**, *83*, 414–430. [CrossRef]
104. Blackburn, G.A. Hyperspectral remote sensing of plant pigments. *J. Exp. Bot.* **2007**, *58*, 855–867. [CrossRef] [PubMed]
105. Ustin, S.L.; Gitelson, A.A.; Jacquemoud, S.; Schaepman, M.E.; Asner, G.P.; Gamon, J.A.; Zarco-Tejada, P. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sens. Environ.* **2009**, *113*, S67–S77. [CrossRef]
106. Gara, T.W.; Skidmore, A.K.; Darvishzadeh, R.; Wang, T. Leaf to canopy upscaling approach affects the estimation of canopy traits. *GISci. Remote Sens.* **2019**, *56*, 554–575. [CrossRef]
107. Jin, Z.; Tian, Q.; Chen, J.; Chen, M. Spatial scaling between leaf area index maps of different resolutions. *J. Environ. Manag.* **2007**, *85*, 628–637. [CrossRef]
108. Kang, X.; Huang, C.; Zhang, L.; Zhang, Z.; Lv, X. Downscaling solar-induced chlorophyll fluorescence for field-scale cotton yield estimation by a two-step convolutional neural network. *Comput. Electron. Agric.* **2022**, *201*, 107260. [CrossRef]

109. Ni-Meister, W.; Yang, W.; Kiang, N.Y. A clumped-foliage canopy radiative transfer model for a global dynamic terrestrial ecosystem model. I: Theory. *Agric. For. Meteorol.* **2010**, *150*, 881–894. [CrossRef]
110. Wang, Q.; Li, P. Canopy vertical heterogeneity plays a critical role in reflectance simulation. *Agric. For. Meteorol.* **2013**, *169*, 111–121. [CrossRef]
111. Crawford, C.J.; Roy, D.P.; Arab, S.; Barnes, C.; Vermote, E.; Hulley, G.; Gerace, A.; Choate, M.; Engebretson, C.; Micijevic, E.; et al. The 50-year Landsat collection 2 archive. *Sci. Remote Sens.* **2023**, *8*, 100103. [CrossRef]
112. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
113. Valavi, R.; Elith, J.; Lahoz-Monfort, J.J.; Guillera-Arroita, G. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **2019**, *10*, 225–232. [CrossRef]
114. Zhang, X.; Friedl, M.A.; Schaaf, C.B. Sensitivity of vegetation phenology detection to the temporal resolution of satellite data. *Int. J. Remote Sens.* **2009**, *30*, 2061–2074. [CrossRef]
115. Atkinson, P.M.; Jeganathan, C.; Dash, J.; Atzberger, C. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sens. Environ.* **2012**, *123*, 400–417. [CrossRef]
116. Levitan, N.; Kang, Y.; Özdoğan, M.; Magliulo, V.; Castillo, P.; Moshary, F.; Gross, B. Evaluation of the Uncertainty in Satellite-Based Crop State Variable Retrievals Due to Site and Growth Stage Specific Factors and Their Potential in Coupling with Crop Growth Models. *Remote Sens.* **2019**, *11*, 1928. [CrossRef] [PubMed]
117. Pipia, L.; Belda, S.; Franch, B.; Verrelst, J. Trends in satellite sensors and image time series processing methods for crop phenology monitoring. In *Information and Communication Technologies for Agriculture—Theme I: Sensors*; Springer: Cham, Switzerland, 2022; pp. 199–231. [CrossRef]
118. White, M.A.; Thornton, P.E.; Running, S.W. A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Glob. Biogeochem. Cycles* **2009**, *21*, 217–234. [CrossRef]
119. Richardson, A.D.; Keenan, T.F.; Migliavacca, M.; Ryu, Y.; Sonnentag, O.; Toomey, M. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agric. For. Meteorol.* **2013**, *169*, 156–173. [CrossRef]
120. Guanter, L.; Zhang, Y.; Jung, M.; Joiner, J.; Voigt, M.; Berry, J.A.; Frankenberg, C.; Huete, A.R.; Zarco-Tejada, P.; Lee, J.E.; et al. Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E1327–E1333. [CrossRef]
121. Rossini, M.; Meroni, M.; Celesti, M.; Cogliati, S.; Julitta, T.; Panigada, C.; Rascher, U.; van der Tol, C.; Colombo, R. Analysis of red and far-red sun-induced chlorophyll fluorescence and their ratio in different canopies based on observed and modeled data. *Remote Sens.* **2016**, *8*, 412. [CrossRef]
122. Liu, X.; Liu, L.; Guan, K.; Du, S.; Wang, S. Tracing the seasonality of photosynthesis in croplands using remotely sensed sun-induced fluorescence. *Agric. For. Meteorol.* **2017**, *232*, 237–248. [CrossRef]
123. Köhler, P.; Frankenberg, C.; Joiner, J.; Guanter, L. Global retrievals of solar-induced chlorophyll fluorescence with TROPOMI: First results and intersensor comparison to OCO-2. *Geophys. Res. Lett.* **2018**, *45*, 10456–10463. [CrossRef]
124. Guanter, L.; Bacour, C.; Schneider, A.; Aben, I.; van Kempen, T.A.; Maignan, F.; Retscher, C.; Köhler, P.; Frankenberg, C.; Joiner, J.; et al. The TROPOSIF global sun-induced fluorescence data set from the TROPOMI instrument. *Earth Syst. Sci. Data* **2021**, *13*, 5423–5440. [CrossRef]
125. Liang, S.; Fang, H.; Chen, M. Atmospheric correction of Landsat ETM+ land surface imagery. I. Methods. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 2490–2498. [CrossRef]
126. Claverie, M.; Ju, J.; Masek, J.G.; Dungan, J.L.; Vermote, E.F.; Roger, J.; Skakun, S.V.; Justice, C. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sens. Environ.* **2018**, *219*, 145–161. [CrossRef]
127. Pouliot, D.; Latifovic, R.; Fernandes, R.; Olthof, I. Evaluation of compositing period and AVHRR and MERIS spectral bands for phenology monitoring. *Remote Sens. Environ.* **2011**, *115*, 158–166. [CrossRef]
128. Cai, Z.; Jönsson, P.; Jin, H.; Eklundh, L. Performance of Smoothing Methods for Reconstructing NDVI Time-Series and Estimating Vegetation Phenology from MODIS Data. *Remote Sens.* **2017**, *9*, 1271. [CrossRef]
129. Vicca, S.; Balzarolo, M.; Filella, I.; Granier, A.; Herbst, M.; Knohl, A.; Longdoz, B.; Mund, M.; Nagy, Z.; Pintér, K.; et al. Remotely-sensed detection of effects of extreme droughts on gross primary production. *Sci. Rep.* **2016**, *6*, 28269. [CrossRef]
130. Zhu, Z.; Woodcock, C.E. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* **2014**, *144*, 152–171. [CrossRef]
131. Jamali, S.; Jönsson, P.; Eklundh, L.; Ardö, J.; Seaquist, J. Detecting changes in vegetation trends using time series segmentation. *Remote Sens. Environ.* **2015**, *156*, 182–195. [CrossRef]
132. Gómez, C.; White, J.C.; Wulder, M.A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 55–72. [CrossRef]

133. Wulder, M.A.; Masek, J.G.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* **2012**, *112*, 2–10. [CrossRef]
134. Sun, Y.; Frankenberg, C.; Wood, J.D.; Schimel, D.S.; Jung, M.; Guanter, L.; Drewry, D.; Verma, M.; Porcar-Castell, A.; Griffis, T.J.; et al. OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. *Science* **2017**, *358*, eaam5747. [CrossRef]
135. Porcar-Castell, A.; Tyystjärvi, E.; Atherton, J.; Van der Tol, C.; Flexas, J.; Pfündel, E.; Moreno, J.; Frankenberg, C.; Berry, J. Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: Mechanisms and challenges. *J. Exp. Bot.* **2014**, *65*, 4065–4095. [CrossRef]
136. Running, S.W.; Nemani, R.R.; Heinsch, F.A.; Zhao, M.; Reeves, M.; Hashimoto, H. A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production. *BioScience* **2004**, *54*, 547–560. [CrossRef]
137. Yuan, W.; Liu, S.; Zhou, G.; Zhou, G.; Tieszen, L.L.; Baldocchi, D.; Bernhofer, C.; Gholz, H.; Goldstein, A.H.; Goulden, M.; et al. Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. *Agric. For. Meteorol.* **2007**, *143*, 189–207. [CrossRef]
138. Turner, D.P.; Ritts, W.D.; Styles, J.M.; Yang, Z.; Cohen, W.B.; Law, B.E.; Thornton, P.E. Evaluation of MODIS NPP and GPP products across multiple biomes. *Remote Sens. Environ.* **2006**, *102*, 282–292. [CrossRef]
139. Cranko Page, J.; Abramowitz, G.; De Kauwe, M.G.; Pitman, A.J. Are Plant Functional Types Fit for Purpose? *Geophys. Res. Lett.* **2024**, *51*, e2023GL104962. [CrossRef]
140. Xie, X.; Li, A.; Chen, J.; Guan, X.; Leng, J. Quantifying Scaling Effect on Gross Primary Productivity Estimation in the Upscaling Process of Surface Heterogeneity. *J. Geophys. Res. Biogeosci.* **2022**, *127*, e2021JG006775. [CrossRef]
141. Ma, X.; Huete, A.R.; Yu, Q.; Restrepo-Coupe, N.; Beringer, J.; Hutley, L.B.; Kanniah, K.D.; Cleverly, J.; Eamus, D. Parameterization of an ecosystem light-use-efficiency model for predicting savanna GPP using MODIS EVI. *Remote Sens. Environ.* **2014**, *154*, 253–271. [CrossRef]
142. Xiao, J.; Zhuang, Q.; Law, B.E.; Chen, J.; Baldocchi, D.D.; Cook, D.R.; Oren, R.; Richardson, A.D.; Wharton, S.; Ma, S.; et al. A continuous measure of gross primary productivity for the conterminous U.S. derived from MODIS and AmeriFlux data. *Remote Sens. Environ.* **2010**, *114*, 576–591. [CrossRef]
143. Reyes-Muñoz, P.; Kovács, D.D.; Berger, K.; Pipia, L.; Belda, S.; Rivera-Caicedo, J.P.; Verrelst, J. Inferring global terrestrial carbon fluxes from the synergy of Sentinel 3 & 5P with Gaussian process hybrid models. *Remote Sens. Environ.* **2024**, *305*, 114072. [CrossRef]
144. Reyes-Muñoz, P.; DKovács, D.; Verrelst, J. Tower-to-global upscaling of terrestrial carbon fluxes driven by MODIS-LAI, Sentinel-3-LAI and ERA5-Land data. *Ecol. Indic.* **2025**, *177*, 113597. [CrossRef]
145. Gao, F.; Masek, J.; Schwaller, M.; Hall, F. On the Blending of the Landsat and MODIS Surface Reflectance: Predicting Daily Landsat Surface Reflectance. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2207–2218. [CrossRef]
146. Zhu, X.; Chen, J.; Gao, F.; Chen, X.; Masek, J.G. An Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model for Complex Heterogeneous Regions. *Remote Sens. Environ.* **2010**, *114*, 2610–2623. [CrossRef]
147. Zhu, X.; Helmer, E.H.; Gao, F.; Liu, D.; Chen, J.; Lefsky, M.A. A Flexible Spatiotemporal Method for Fusing Satellite Images with Different Resolutions. *Remote Sens. Environ.* **2016**, *172*, 165–177. [CrossRef]
148. Gevaert, C.M.; García-Haro, F.J. A Comparison of STARFM and an Unmixing-Based Algorithm for Landsat and MODIS Data Fusion. *Remote Sens. Environ.* **2015**, *156*, 34–44. [CrossRef]
149. Xue, J.; Leung, Y.; Fung, T. A Bayesian Data Fusion Approach to Spatio-Temporal Fusion of Remotely Sensed Images. *Remote Sens.* **2017**, *9*, 1310. [CrossRef]
150. Verrelst, J.; García-Soria, J.L.; Reyes-Muñoz, P.; De Clerck, E.; Morata, M.; Rivera-Caicedo, J.P. Epistemic and aleatoric uncertainty in optical vegetation trait retrieval: Concepts, Methods, and Outlook. *ISPRS J. Photogramm. Remote Sens.* **2026**, *234*, 20–45. [CrossRef]
151. Roy, D.P.; Li, J.; Zhang, H.K.; Yan, L.; Huang, H.; Li, Z. Examination of Sentinel-2A Multi-Spectral Instrument (MSI) Reflectance Anisotropy and the Suitability of a General Method to Normalize MSI Reflectance to Nadir BRDF Adjusted Reflectance. *Remote Sens. Environ.* **2017**, *199*, 25–38. [CrossRef]
152. Fang, H.; Jiang, C.; Li, W.; Wei, S.; Baret, F.; Chen, J.M.; García-Haro, F.J.; Liang, S.; Liu, R.; Myneni, R.B.; et al. Characterization and intercomparison of global moderate resolution leaf area index (LAI) products: Analysis of climatologies and theoretical uncertainties. *J. Geophys. Res. Biogeosci.* **2013**, *118*, 529–548. [CrossRef]
153. Verger, A.; Baret, F.; Weiss, M. A multisensor fusion approach to improve LAI time series. *Remote Sens. Environ.* **2011**, *115*, 2460–2470. [CrossRef]
154. Baret, F.; Weiss, M.; Lacaze, R.; Camacho, F.; Makhmara, H.; Pacholcyzk, P.; Smets, B. GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 1: Principles of development and production. *Remote Sens. Environ.* **2013**, *137*, 299–309. [CrossRef]

155. Joiner, J.; Yoshida, Y.; Vasilkov, A.P.; Yoshida, Y.; Corp, L.A.; Middleton, E.M. First observations of global and seasonal terrestrial chlorophyll fluorescence from space. *Biogeosciences* **2011**, *8*, 637–651. [CrossRef]
156. Li, X.; Xiao, J. A Global, 0.05-Degree Product of Solar-Induced Chlorophyll Fluorescence Derived from OCO-2, MODIS, and Reanalysis Data. *Remote Sens.* **2019**, *11*, 517. [CrossRef]
157. Shekhar, A.; Buchmann, N.; Gharun, M. How well do recently reconstructed solar-induced fluorescence datasets model gross primary productivity? *Remote Sens. Environ.* **2022**, *283*, 113282. [CrossRef]
158. Magney, T.S.; Bowling, D.R.; Logan, B.A.; Grossmann, K.; Stutz, J.; Blanken, P.D.; Burns, S.P.; Cheng, R.; Garcia, M.A.; Köhler, P.; et al. Mechanistic evidence for tracking the seasonality of photosynthesis with solar-induced fluorescence. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 11640–11645. [CrossRef]
159. Kaminski, T.; Knorr, W.; Rayner, P.J.; Heimann, M. Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle. *Glob. Biogeochem. Cycles* **2002**, *4*, 14-1–14-16. [CrossRef]
160. Wikle, C.K. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology* **2003**, *84*, 1382–1394. [CrossRef]
161. Cressie, N.; Wikle, C.K. *Statistics for Spatio-Temporal Data*; Wiley: Hoboken, NJ, USA, 2011. [CrossRef]
162. Kennedy, M.C.; O'Hagan, A. Predicting the Output from a Complex Computer Code when Fast Approximations are Available. *Biometrika* **2000**, *87*, 1–13. [CrossRef]
163. Bilionis, I.; Zabaras, N.; Konomi, B.A.; Lin, G. Multi-output separable Gaussian process: Towards an efficient, fully Bayesian paradigm for uncertainty quantification. *J. Comput. Phys.* **2013**, *241*, 212–239. [CrossRef]
164. Perdikaris, P.; Raissi, M.; Damianou, A.; Lawrence, N.D.; Karniadakis, G.E. Nonlinear Information Fusion Algorithms for Data-Efficient Multi-Fidelity Modelling. *Proc. R. Soc. A* **2017**, *473*, 20160751. [CrossRef]
165. Meyer, H.; Reudenbach, C.; Wöllauer, S.; Nauss, T. Importance of spatial predictor variable selection in machine learning applications—Moving from data reproduction to spatial prediction. *Ecol. Model.* **2019**, *411*, 108815. [CrossRef]
166. Ploton, P.; Mortier, F.; Réjou-Méchain, M.; Barbier, N.; Picard, N.; Rossi, V. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **2020**, *11*, 4540. [CrossRef] [PubMed]
167. Rastetter, E.B.; King, A.W.; Cosby, B.J.; Hornberger, G.M.; O'Neill, R.V.; Hobbie, J.E. Aggregating fine-scale ecological knowledge to model coarser-scale attributes of ecosystems. *Ecol. Appl.* **1992**, *2*, 55–70. [CrossRef] [PubMed]
168. Schaepman-Strub, G.; Schaepman, M.E.; Painter, T.H.; Dangel, S.; Martonchik, J.V. Reflectance quantities in optical remote sensing—Definitions and case studies. *Remote Sens. Environ.* **2006**, *103*, 27–42. [CrossRef]
169. Gascon, F.; Michael, R.A.S.T.; Nieke, J.; Celesti, M.; Bogaarts, C. CHIME: Une Mission Copernicus d'Imagerie Hyperspectrale Pour l'Environment. *Revue Française Photogrammétrie Télédétection* **2022**, *224*, 5–8. [CrossRef]
170. De Grave, C.; Verrelst, J.; Morcillo-Pallarés, P.; Pipia, L.; Rivera-Caicedo, J.P.; Amin, E.; Belda, S.; Moreno, J. Quantifying vegetation biophysical variables from the Sentinel-3/FLEX tandem mission: Evaluation of the synergy of OLCI and FLORIS data sources. *Remote Sens. Environ.* **2020**, *251*, 112101. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

remote sensing

MDPI

Article

Bathymetric Inversion of Tibetan Plateau Lakes Using Hyperspectral Imagery and ICESat-2 Data

Chang Zhong [1,2], Yu Zhao [2,*], Mengchun Pan [1], Qi Zhang [1], Xinxin Sui [2], Li Chen [2], Ning Wang [2] and Fan Bu [2]

[1] College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410003, China; zhongchang@mail.cgs.gov.cn (C.Z.); panmengchun@mail.nudt.edu.cn (M.P.); zhangqi@mail.nudt.edu.cn (Q.Z.)
[2] China Aero Geophysical Survey and Remote Sensing Center for Natural Resources, Beijing 100083, China; suixinxin@mail.cgs.gov.cn (X.S.); cli_a@mail.cgs.gov.cn (L.C.); wangning@mail.cgs.gov.cn (N.W.); bufan@mail.cgs.gov.cn (F.B.)
* Correspondence: zhaoyu_hy@mail.cgs.gov.cn

Highlights

What are the main findings?

- A hyperspectral–ICESat-2 joint framework was proposed for Tibetan Plateau lake bathymetric inversion.
- The method achieved more accurate and spatially consistent bathymetric reconstruction than representative baseline methods.

What are the implications of the main findings?

- Multi-source constraints can improve bathymetric inversion in plateau lakes with limited in situ measurements.
- The framework is useful for lake basin mapping and water storage studies in high-altitude inland environments.

Abstract

Lake depth is a fundamental parameter for estimating lake storage, analyzing basin morphology, and understanding the evolution of plateau lakes. Compared with typical shallow lakes, Tibetan Plateau lakes are characterized by high elevation, strong radiation, pronounced inter-lake and inter-annual variability, and in some cases considerable basin depth, which limits the accuracy, stability, and generalization ability of existing bathymetric inversion methods based on single-source optical imagery. Meanwhile, although ICESat-2 can provide sparse but high-precision along-track bathymetric constraints, a unified framework suitable for plateau-lake scenarios is still lacking. To address this issue, this study proposes TabKAN, a bathymetric inversion framework for Tibetan Plateau lakes under joint constraints from hyperspectral imagery and ICESat-2 data. TabKAN constructs tabular input features from hyperspectral reflectance, water indices, imaging geometry, and environmental variables; employs TabNet for feature selection and encoding; and introduces a KAN regression head to enhance nonlinear bathymetric mapping. A joint-supervision and bias-correction mechanism is further designed to incorporate ICESat-2 samples, thereby improving model robustness across lakes and acquisition dates. To enhance the temporal coverage of training samples, multi-year sample expansion based on stereo-mapping data is introduced, and a stripe-aware self-supervised learning strategy is developed for hyperspectral image restoration and pretraining. Experiments on five Tibetan Plateau lakes, including Anglaren Co, Caiduo Chaka, Cuoe, Geren Co, and Qixiang Co, show that the proposed method outperforms benchmark methods in both overall accuracy and depth-stratified evaluation, while providing more stable recovery of basin morphology and depth

gradients. These results demonstrate that combining hyperspectral information, ICESat-2 laser constraints, and stripe-aware pretraining can effectively improve the accuracy and robustness of bathymetric inversion for Tibetan Plateau lakes and provide a new technical route for storage estimation and change monitoring of cold inland lakes.

Keywords: hyperspectral remote sensing; ICESat-2; lake bathymetric inversion; Tibetan Plateau

1. Introduction

Lake depth and basin morphology are fundamental descriptors of lake shape, storage capacity, and hydrological processes. They are also indispensable for lake-volume estimation, water-balance analysis, interpretation of ecological processes, and regional water-resource assessment. For inland lakes, bathymetry also determines how hydrological forcing is expressed at the lake surface. For example, the same water-level increase may cause extensive shoreline migration in a gently sloping basin, whereas it may mainly lead to vertical storage increase in a steep basin. Therefore, monitoring lake-bottom gradients and nearshore morphology is essential for interpreting lake-area fluctuations, water-level variations, and the storage-buffering role of lake basins. On the Tibetan Plateau, lakes are widely distributed in cold high-altitude environments and are among the most sensitive natural units responding to regional climate change and water-cycle variations. In recent years, remotely sensed changes in lake area and water level on the Tibetan Plateau have received considerable attention; however, compared with lake-surface area and water level, bathymetric information remains severely lacking, which greatly constrains a deeper understanding of lake-volume change, water-budget patterns, and basin evolution [1–3].

Traditional acquisition of lake bathymetry mainly relies on field surveys such as boat-borne sounding. Although these methods can provide high accuracy, they are expensive, inefficient, and spatially limited, making them unsuitable for repeated and large-scale surveys across the Tibetan Plateau, where lakes are numerous, scattered, and often located in harsh environments. With the development of remote sensing, optical bathymetric inversion has gradually become an important pathway for retrieving shallow-water depth. Since the pioneering work of Lyzenga [4], and the classical band-ratio model proposed by Stumpf et al. [5], remote sensing bathymetry has evolved from empirical formulations to semi-analytical models, machine learning, and deep learning. Reviews have shown that, although optical bathymetric inversion has the advantages of non-contact sensing, large-area coverage, and relatively low cost, its accuracy and applicability are still jointly constrained by water transparency, bottom type, imaging conditions, and regional differences [6].

Among various optical data sources, hyperspectral remote sensing is generally considered to have greater potential than traditional multispectral imagery for characterizing subtle differences in water properties and bottom conditions, owing to its continuous and narrow-band spectral sampling. Representative studies have already shown that hyperspectral bathymetric inversion has a solid theoretical and empirical foundation. The classical HOPE framework established a physical foundation for hyperspectral inversion in shallow water, and subsequent work further demonstrated the feasibility of using hyperspectral data in optically shallow bathymetric inversion [7,8]. Domestic studies based on Hyperion, PRISMA, and Chinese hyperspectral sensors have also indicated that hyperspectral imagery has both theoretical and experimental potential to outperform multispectral imagery in bathymetric applications [9–11]. Nevertheless, existing hyperspectral bathymetry studies

have mainly focused on clear shallow marine waters, and systematic research for inland plateau lakes remains notably insufficient.

It is worth noting that the value of hyperspectral imagery does not simply lie in "having more bands"; the real challenge is how to extract the most bathymetrically informative signals from high-dimensional, redundant observations that are susceptible to noise. Previous studies have pointed out that deep-learning-based hyperspectral bathymetric inversion is still at a relatively early stage. On the one hand, high-dimensional spectral information is not yet fully exploited; on the other hand, environmental variation, imaging-geometry differences, and dispersed key-band importance can substantially weaken model stability. This issue is even more pronounced for Tibetan Plateau lakes, because the same depth may correspond to different spectral responses across lakes, years, and observation conditions. Therefore, one major difficulty is whether high-dimensional spectral information, environmental variables, and cross-scene variability can be handled simultaneously in a unified framework. For point-sample-driven bathymetric inversion, the task is essentially closer to a continuous regression problem driven by multi-source heterogeneous tabular features rather than a simple image-spectrum fitting problem. TabNet and KAN provide useful ideas for model design in this context [12,13].

Complementary to passive optical sensing, ICESat-2 satellite laser altimetry provides sparse but accurate along-track observations. ICESat-2 carries a 532 nm green photon-counting laser altimeter, which has shown the potential to retrieve underwater information in clear lakes and shallow-water environments [14]. Previous studies have shown that ICESat-2 can be used not only for bathymetric reconstruction of Tibetan Plateau lakes [1], but also in combination with Sentinel-2 and other optical imagery to generate continuous bathymetric maps [11,15]. At the same time, active–passive fusion studies have pointed out that, although ICESat-2 provides valuable high-accuracy constraints, its sparse along-track geometry, environmentally dependent errors, and variable transferability across study areas limit its direct use as ground truth. Therefore, how to take advantage of the high accuracy of ICESat-2 while avoiding the direct injection of its systematic bias into the main optical inversion relationship remains a key gap in current optical–laser bathymetric inversion research.

In addition to label-source inconsistency, the quality of hyperspectral imagery itself can further complicate the modeling process. Push-broom hyperspectral sensors are often affected by stripe noise, detector-response inconsistency, and local outliers, which not only damage radiometric consistency across bands but also directly affect downstream feature learning and regression stability [16,17]. Recent advances in self-supervised hyperspectral denoising further indicate that, when clean paired samples are unavailable, jointly considering denoising and the downstream task is an effective way to improve the usability of real satellite hyperspectral data [18]. For Tibetan Plateau lake bathymetry, where samples are already limited and inter-annual variability is pronounced, data quality control should not be treated as an isolated preprocessing step; instead, it should be designed together with feature learning and downstream regression.

Motivated by the above observations, this study proposes a unified deep-learning framework, namely TabKAN, for hyperspectral lake bathymetric inversion over the Tibetan Plateau. Rather than simply using satellite observations as additional data sources, the proposed framework is designed to reduce the major error sources discussed above. Hyperspectral redundancy and stripe-related noise are addressed through adaptive feature selection and stripe-aware self-supervised pretraining; cross-scene spectral variability is alleviated by incorporating water indices, imaging geometry, and environmental variables; complex nonlinear relationships between reflectance and depth are represented by the KAN regression head; and potential ICESat-2 label bias is handled by an explicit environmental

bias branch. The framework formulates lake-depth inversion as a continuous regression problem driven by multi-source heterogeneous tabular features. At the input level, hyperspectral reflectance, water indices, imaging geometry, and environmental variables are fused in a unified representation. At the model level, TabNet is used to adaptively select informative variables, while KAN is introduced to enhance the expressive power for complex nonlinear relationships. At the supervision level, a bias subnet driven only by environmental features is introduced to explicitly characterize the environment-dependent bias of ICESat-2 labels relative to measured samples. In addition, multi-year sample expansion using stereo mapping data and stripe-aware self-supervised pretraining are incorporated to improve hyperspectral data quality and encoder stability. The main contributions of this study are summarized as follows:

1. A multi-source tabular modeling strategy is proposed for bathymetric inversion of Tibetan Plateau lakes, in which hyperspectral reflectance, water indices, imaging geometry, and environmental variables are organized within a unified regression framework.
2. A TabKAN backbone that integrates TabNet and KAN is developed for lake bathymetric inversion, enhancing nonlinear representation while retaining interpretable feature selection.
3. A joint-supervision and bias-correction strategy based on measured samples and ICESat-2 samples is proposed, turning ICESat-2 from a set of additional training points into an explicitly constrained high-precision auxiliary reference.
4. A combination of multi-year sample expansion, self-supervised pretraining, and stripe suppression is introduced to improve the stability and generalization ability of hyperspectral lake bathymetric inversion across multiple acquisition dates.

2. Related Work

Research on bathymetric inversion has generally evolved from empirical models and semi-analytical or physical models to machine learning, deep learning, and, more recently, active–passive remote sensing fusion. Early optical bathymetric inversion was mainly based on empirical relationships between multispectral bands and water depth. Lyzenga systematically discussed passive multispectral retrieval of water depth and bottom features, laying the foundation for subsequent empirical models [4]. The band-ratio model proposed by Stumpf et al. further improved robustness by reducing the influence of bottom-type differences and radiometric variations, and has since become one of the most representative classical baselines in satellite-derived bathymetry [5]. Subsequent studies developed a relatively complete technical spectrum of empirical, semi-empirical, semi-analytical, and physical models. Recent reviews have emphasized that, although these methods are effective in clear shallow-water environments, they still depend strongly on measured samples, parameter settings, and local scene conditions, and their cross-region and cross-temporal generalization ability remains limited [6,19].

Beyond empirical and semi-empirical methods, semi-analytical and physical models attempt to improve physical consistency by explicitly describing radiative transfer in the water column. The HOPE framework proposed by Lee et al. is a representative contribution in this direction. Its core idea is to jointly estimate water depth, water optical properties, and bottom reflectance so that hyperspectral reflectance can be interpreted within a unified model [7]. Later studies showed that these methods can balance bathymetric inversion and water-property estimation when water transparency is high and the model parameters are reasonably constrained. However, they also suffer from large parameter spaces, complex optimization, and sensitivity to initialization. For lake environments with strong variability, especially where bottom type, water transparency, and imaging conditions co-vary, it is difficult to stably describe the complex spectrum–depth relationship using a fixed radiative-

https://doi.org/10.3390/rs18121886

transfer structure. Thus, although physical models provide an important theoretical basis for bathymetric inversion, their practical usability in complex scenarios remains limited.

With the development of hyperspectral remote sensing, researchers began to exploit contiguous narrow-band spectral information to improve bathymetric inversion. Compared with conventional multispectral imagery, hyperspectral data are theoretically better suited to characterizing water absorption, scattering, and bottom differences, and are therefore widely considered promising for improving bathymetric accuracy. Earlier work verified the feasibility of hyperspectral imagery for depth retrieval in optically shallow water [8]. Subsequent studies summarized the major models and application conditions of satellite hyperspectral bathymetry in shallow seas and further demonstrated the practical value of domestic hyperspectral sensors and attention-based deep learning for this task [9]. At the same time, recent studies have emphasized that hyperspectral data are not simply "multispectral data with more bands"; rather, they bring higher redundancy, greater sensitivity to noise, and more demanding requirements for model design. In other words, the key issue is not whether hyperspectral data are available, but whether the most bathymetrically effective information can be extracted from high-dimensional and environmentally sensitive observations. Although progress has been made, most existing studies still focus on clear shallow marine waters, and systematic investigations for inland plateau lakes remain limited.

The introduction of machine learning and deep learning has brought a methodological shift to bathymetric inversion. Compared with empirical models and shallow statistical regressions, machine learning and deep learning are better at learning complex nonlinear mappings from multi-band observations, and therefore often achieve stronger fitting ability and better local accuracy in both multispectral and hyperspectral scenarios. Recent studies have begun to use random forests, support vector machines, convolutional neural networks, and hybrid physics-guided networks for bathymetric inversion. For example, a fast feature cascade learning model and a dual-physics-guided deep learning framework have both demonstrated that data-driven models can significantly improve bathymetric performance in complex shallow-water environments [20,21]. However, most existing methods still focus primarily on the imagery itself, typically treating bathymetric inversion as a single image-regression problem and placing emphasis on feature extraction from reflectance and a few handcrafted indices, while rarely accounting systematically for imaging geometry, environmental metadata, time information, and label-source differences. This is particularly important for Tibetan Plateau lakes because their spectral response is controlled not only by depth, but also by lake-level variation, transparency, acquisition time, and local environmental conditions. As a result, models based solely on image spectra or local spatial context may perform well in local tasks but still lack transferability across lakes, years, and sensors.

Complementary to passive optical sensing, ICESat-2 satellite laser altimetry has become an important source of auxiliary information. Since launch, ICESat-2 has been widely used in cryospheric, oceanic, and inland-water applications. Its 532 nm green laser can, under favorable water conditions, penetrate water and support underwater observations [14]. In the context of lakes, Han et al. demonstrated the potential of ICESat-2 for reconstructing bathymetry and estimating storage of clear Tibetan Plateau lakes [1]. This means that ICESat-2 can contribute not only water-surface elevation but also direct subsurface information under suitable conditions. However, repeated studies have shown that the usefulness of ICESat-2 depends strongly on water transparency, noise-photon removal, refraction correction, and track coverage. In deep lakes, turbid waters, or complex environments, the sparsity of along-track observations substantially limits continuous bathymetric reconstruction. Therefore, the real advantage of ICESat-2 is better understood as high-precision

https://doi.org/10.3390/rs18121886

but sparse constraint information rather than a continuous truth surface that can directly replace in situ measurements.

Under this background, active–passive fusion has become an important direction in bathymetric research. A number of studies have attempted to combine ICESat-2 with optical imagery to generate continuous bathymetric maps. ICESat-2 point-depth observations have been used together with Sentinel-2 imagery to provide new high-precision samples in the absence of large-scale field data [11], and cloud-native regional active–passive bathymetric frameworks have also been reported [15]. Later work incorporated time series, error reduction, and machine learning into such fusion frameworks to further improve the accuracy and stability of continuous bathymetric maps [22,23]. These studies clearly show that active–passive fusion is an effective route for improving optical bathymetry. Nevertheless, recent reviews have also emphasized that ICESat-2 bathymetry itself involves multiple stages, including water-surface detection, underwater-photon identification, refraction correction, and precision control, and that its error sources are strongly environment dependent [24]. This implies that, if ICESat-2 samples are directly treated as labels equivalent to measured samples in a deep-learning framework, the model may inadvertently mix laser-label bias into the true optical spectrum–depth mapping, thereby degrading cross-scene generalization. Although the value of ICESat-2 has been demonstrated, explicit modeling of such heterogeneous label bias remains insufficient.

Another often underestimated aspect of point-sample-based bathymetric inversion is that the task is fundamentally closer to multi-source heterogeneous tabular regression than to end-to-end image segmentation. Each training sample usually includes not only high-dimensional spectral reflectance, but also water indices, imaging geometry, temporal information, and environmental variables. Therefore, high-dimensional feature selection, heterogeneous variable fusion, and interpretable modeling within a unified framework are central to model performance. In the tabular deep learning literature, TabNet introduced explicit feature selection through sequential attention and provides a framework that balances predictive performance and interpretability for heterogeneous tabular data [12]. Subsequent work has shown that TabNet-like structures have considerable potential for hyperspectral tasks in terms of band selection and representation learning [25]. Meanwhile, KAN replaces conventional linear weights with learnable one-dimensional functions and has demonstrated strong nonlinear expressiveness and favorable interpretability in scientific regression problems [13]. Although applications of these two model families to bathymetric inversion are still rare, they directly address two critical gaps in current research: how to identify truly informative variables from high-dimensional heterogeneous inputs, and how to represent complex yet interpretable nonlinear relationships in a unified framework. For the sample structure considered here, namely "hyperspectral reflectance + water indices + imaging geometry + environmental variables + heterogeneous supervision", tabular deep learning is particularly well suited.

Finally, the data-quality issue of hyperspectral imagery itself leaves further room for improvement. Push-broom hyperspectral images often suffer from stripe noise, detector inconsistency, and local outliers, which can destroy radiometric consistency across bands and contaminate downstream feature representations [16,17]. Recent studies have emphasized that hyperspectral denoising should not be treated as an isolated preprocessing step, but rather be considered together with specific downstream applications. For example, recent work reinterpreted destriping and denoising from a task-decomposition perspective [26], while self-supervised reconstruction methods have shown that it is possible to support hyperspectral representation learning and denoising initialization without paired clean samples [27]. Even so, most existing studies on hyperspectral stripe suppression and

https://doi.org/10.3390/rs18121886

self-supervised pretraining still proceed independently. Systematically coupling the two with a specific bathymetric inversion task remains uncommon.

To make the methodological landscape clearer, existing bathymetric inversion methods can be summarized according to their core models and parameterization strategies. Empirical and semi-empirical methods, such as log-linear and band-ratio models, usually rely on a small number of fitted coefficients and are easy to implement, but their parameters are strongly scene dependent and their transferability is limited. Semi-analytical and physical methods explicitly parameterize water-column optical properties, bottom reflectance, and depth, providing stronger physical interpretability; however, they often require more prior information and are sensitive to initialization and water-optical variability. Traditional machine-learning methods, including instance-based and tree-ensemble models, learn nonlinear mappings from spectral or tabular features to depth and are effective for structured samples, but their extrapolation ability is still constrained by the distribution and representativeness of training samples. Deep-learning methods further enhance nonlinear representation and feature learning, but they usually require larger and more diverse training data and may be affected by heterogeneous label sources and sensor noise. Active–passive fusion methods introduce ICESat-2 or other laser-derived depth constraints to compensate for sparse field measurements, but they must handle track sparsity, water-clarity dependence, and possible laser-label bias. Compared with these approaches, the proposed TabKAN framework combines hyperspectral feature selection, heterogeneous tabular feature fusion, KAN-based nonlinear regression, ICESat-2 bias-aware auxiliary supervision, multi-year sample expansion, and stripe-aware self-supervised pretraining within a unified framework, aiming to improve both accuracy and robustness for Tibetan Plateau lake bathymetry.

3. Study Area and Data

3.1. Study Area

This study focuses on five representative lakes on the Tibetan Plateau: Anglaren Co, Caiduo Chaka, Cuoe, Geren Co, and Qixiang Co. The geographic locations of these lakes are presented in Figure 1. The Tibetan Plateau, characterized by high mean elevation, complex climatic conditions, and dense lake distribution, hosts the largest concentration of cold inland lakes in the world. It is also a key region for investigating lake morphology, water-storage changes, and their climatic responses. The selected lakes are all located in high-elevation cold-arid environments and share several common characteristics, including broad water surfaces, complex observation conditions, and pronounced inter-annual water-level fluctuations. At the same time, they differ in basin morphology, shoreline sinuosity, nearshore slope, and the potential existence of deep troughs, providing a favorable test bed for assessing model adaptability and robustness across different lake settings.

These environmental and geomorphological characteristics directly affect bathymetric inversion. Strong radiation and variable imaging geometry may change the observed water-leaving reflectance, while interannual water-level fluctuations alter the relationship between shoreline position and local water depth. Differences in basin slope, nearshore morphology, water clarity, and bottom conditions can also cause the same depth to show different spectral responses across lakes and acquisition dates. Therefore, bathymetric inversion for Tibetan Plateau lakes requires not only hyperspectral information, but also auxiliary variables and constraints describing acquisition conditions, environmental variability, and basin morphology. In this study, these factors are considered through water indices, imaging-geometry and environmental variables, multi-year lake-level-constrained sample augmentation, ICESat-2 joint supervision, and depth-stratified evaluation.

https://doi.org/10.3390/rs18121886

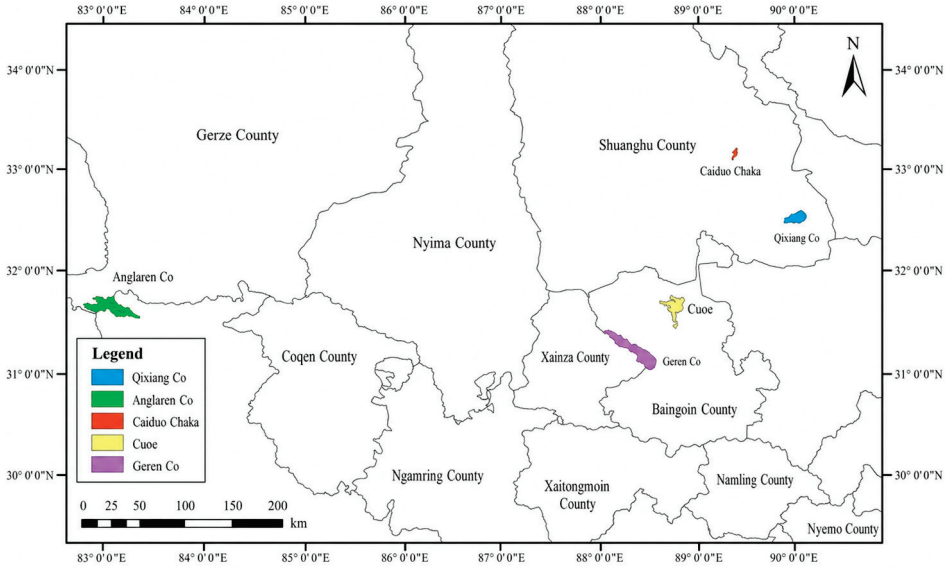


Figure 1. Geographical location of the study area on the Tibetan Plateau.

3.2. Datasets

The data used in this study span the period 2020–2025 and mainly comprise four categories: hyperspectral satellite imagery, ICESat-2 laser altimetry data, ZY-3 and GF-7 stereo mapping data, and in situ bathymetric measurements. These data sources serve different purposes in the proposed framework. Hyperspectral imagery provides continuous areal spectral observations, ICESat-2 supplies sparse but high-precision along-track bathymetric constraints, ZY-3 and GF-7 are used to estimate inter-annual water-level differences and support sample expansion, and in situ depth measurements serve as the main reference for model training and evaluation.

The hyperspectral data are mainly based on the Resource-1 02E AHSI sensor (developed by the Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China), and the dataset was sourced from the China Center for Resources Satellite Data and Application (Beijing, China). This sensor covers the visible, near-infrared, and shortwave infrared spectral ranges, provides 166 bands, has a spatial resolution of 30 m, and a swath width of approximately 60 km. Compared with conventional multispectral imagery, AHSI offers richer spectral information and is therefore more suitable for capturing subtle variations related to water depth, water properties, and bottom conditions.

ICESat-2 data were obtained from the ATL03 photon product. Compared with higher-level derived products, ATL03 preserves more complete original photon information and is therefore more suitable for lake applications such as water-surface detection, underwater-photon extraction, and refraction correction, which are necessary to construct along-track depth-constrained samples. ICESat-2 carries the ATLAS photon-counting green laser altimeter, featuring six beams, meter-scale footprints, and high pulse repetition, and has a certain capability for underwater penetration in clear lakes. In this study, ICESat-2 is not treated as a continuous truth surface; rather, it is used as a sparse, high-precision auxiliary supervision source to strengthen model adaptability across times and lakes.

To support sample augmentation, multi-year stereo-mapping data from ZY-3 and GF-7 were also introduced. ZY-3 provides three-line-array stereo imaging, and GF-7 offers higher spatial resolution and stronger mapping capability. Both are used here to estimate lake-level changes between different years.

The in situ bathymetric data were collected by boat-borne single-beam sounding surveys. The sounding device was an Airmar SS60 dual-frequency single-beam echosounder (Airmar Technology Corporation, Milford, NH, USA), which supports both 50 kHz and

200 kHz operating frequencies. The 200 kHz signal is suitable for shallow-water and local detail measurements, whereas the 50 kHz signal is more stable for deeper waters. The field surveys used as dense a line layout as possible, with additional cross-check lines in key areas to improve basin-shape constraints and quality control. These measured depth points serve as the primary training labels and the main benchmark for accuracy assessment. Figure 2 illustrates a schematic diagram of the in situ bathymetric points and ICESat-2 laser altimetry data over Anglaren Co.

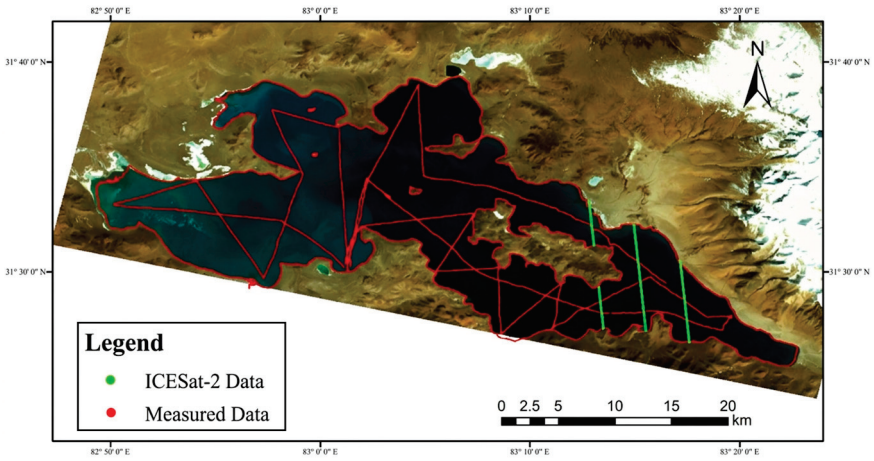


Figure 2. Illustration of the data sources used in this study.

3.3. Sample Augmentation

Because of the logistical challenges of field surveys in plateau environments, in situ bathymetric samples usually cover only a limited number of dates and a limited spatial extent, which directly constrains the sample size and temporal generalization ability of deep learning models. To alleviate this problem, we introduce a sample augmentation strategy constrained by multi-year lake-level differences. The strategy expands measured samples from a single survey date to multiple years, thereby increasing both the temporal span of the training set and the range of water-level conditions represented in the samples.

The basic idea is that, under the assumption that lake-bottom morphology remains relatively stable at the inter-annual scale for the high-altitude inland lakes considered in this study, depth changes at the same location between different years are primarily controlled by changes in lake-surface elevation. This assumption is physically equivalent to treating the lake basin as a stable container during the short observation period. For a fixed point, water depth can be expressed as $D_t = H_t - Z_{b,t}$ where H_t is the lake-surface elevation and $Z_{b,t}$ is the lake-bed elevation. The migrated label from a reference year can therefore be written as $D_{ref} + \left(H_t - H_{ref}\right) = H_t - Z_{b,ref}$, indicating that the systematic error introduced by the stability assumption is $Z_{b,t} - Z_{b,ref}$, namely the actual interannual change in lake-bed elevation. Available sedimentological evidence suggests that this term is very small for typical Tibetan Plateau lakes. Modern sedimentation rates derived from ^{137}Cs and ^{210}Pb dating have been reported to be only 0.45–2.05 mm yr^{-1} in Qinghai Lake [28], and sedimentary records over the past century indicate rates of approximately 1.0–1.59 mm yr^{-1} [29]. Similar bathymetric and sedimentation investigations of high-altitude Tibetan lakes also support the generally stable morphology of plateau lake basins under natural conditions [30]. Even using the upper-bound sedimentation rate, the cumulative bed-elevation change during 2020–2025 is only at the centimeter scale, approximately 1–2 cm, which is orders of magnitude smaller than the meter-level uncertainty of satellite bathymetric inversion. Therefore, the influence of natural lake-bed deformation on the migrated depth labels is expected to be negligible in this study. This treatment is also

consistent with common satellite-hydrology studies of Tibetan Plateau lake storage and volume changes, in which lake basins are commonly regarded as stable containers over multi-year to decadal periods [31–34]. Therefore, if lake-level differences between years can be estimated from ZY-3 and GF-7 stereo mapping data, the measured depth samples from a reference year can be vertically shifted to generate equivalent depth samples for other years. In other words, the method does not simply duplicate existing samples; instead, it uses independently derived stereo-elevation information to perform physically constrained temporal migration of the labels. The resulting samples preserve the spatial positions and hyperspectral feature expressions of the measured samples while explicitly introducing depth changes caused by multi-year water-level variability.

In practice, a year with in situ sounding data is first selected as the reference period, and hyperspectral reflectance, water indices, and environmental variables are extracted for the corresponding sample locations to form the original sample set. Next, lake-level differences between years are estimated for each lake using multi-temporal stereo mapping data from ZY-3 and GF-7. Finally, these water-level differences are added to the measured depths of the reference year to create cross-year equivalent labels, which are matched with the pixels at the same spatial locations in hyperspectral imagery from the corresponding years. This procedure substantially expands the temporal coverage and water-level range of the training data without requiring additional large-scale field sounding.

In addition to the cross-year expanded samples, along-track bathymetric samples derived from ICESat-2 are introduced as a second type of auxiliary sample. Compared with measured samples, ICESat-2 samples are spatially sparser but temporally more flexible and usually have high vertical precision. In this study, they are not treated as equivalent to measured truth; rather, they are used as auxiliary supervision samples carrying environment-dependent bias and are incorporated into the joint-supervision framework during training. The final sample system therefore includes three label sources: measured samples, ICESat-2 samples, and cross-year expanded samples. These play the roles of main supervision, auxiliary supervision, and temporal expansion, respectively, and together form the basis for multi-temporal hyperspectral bathymetric inversion.

After the above processing and quality control, the final point-sample dataset used in the experiments contained 104,616 valid tabular records. Here, a sample denotes a point/pixel-level record obtained by collocating a depth label with the corresponding hyperspectral reflectance, water indices, imaging geometry, and environmental variables, rather than an independent lake object. Compared with a dataset based solely on single-date field measurements, this sample system represents a substantial expansion not only in sample number but also in diversity with respect to acquisition year, water-level state, and lake condition, thereby providing a stronger foundation for building a more robust and generalizable model.

3.4. Data Preprocessing

To ensure that multi-source data could participate in the model within a unified framework, systematic preprocessing was conducted for hyperspectral imagery, ICESat-2 data, stereo-mapping data, and measured bathymetric data. The overall workflow included spatial-reference harmonization, image–point co-registration, outlier removal, auxiliary-variable calculation, and final construction of tabular samples.

For hyperspectral imagery, preprocessing mainly included image selection, radiometric correction, atmospheric correction, geometric correction, and cropping to the study area. Because hyperspectral imagery is susceptible to stripe noise and invalid bands, obviously corrupted bands, low-SNR bands, and local outliers were additionally filtered. At later stages, data stability was further enhanced through self-supervised reconstruction and

https://doi.org/10.3390/rs18121886

stripe suppression. For multiple scenes covering the same lake, a unified geometric reference and mosaicking procedure were also applied to ensure spatial consistency during sample extraction.

For ICESat-2 ATL03 data, preprocessing focused on track selection, noise-photon removal, water-surface detection, underwater-photon extraction, and refraction correction. Effective tracks crossing the study lakes were first extracted according to lake boundaries. The original photon clouds were then denoised to separate water-surface and underwater photons. Finally, underwater-photon elevations were corrected for refraction at the air–water interface, producing along-track bathymetric samples. To reduce the influence of noise photons and anomalous returns, obviously unreasonable outliers were removed before sample construction, and non-lake photons were filtered using lake-boundary constraints.

For ZY-3 and GF-7 stereo mapping data, preprocessing concentrated on stereo registration and estimation of lake-level change. Original stereo images were first processed by block adjustment and stereo mapping. Lake-surface elevations were then extracted from the resulting elevation products, and inter-annual lake-level differences were calculated. Because the purpose of the stereo data here was not to derive a lake-bottom DEM directly, but rather to provide lake-level constraints for cross-year sample expansion, preprocessing emphasized relative elevation consistency among dates rather than absolute terrain-detail recovery.

For in situ bathymetric data, preprocessing included the co-registration of sounding records with GNSS positions, coordinate-system unification, vertical-datum unification, and inspection of anomalous points. Field quality control had already been performed through dense survey lines and cross-check lines. During office processing, obviously anomalous points were further removed, and the bathymetric points were spatially overlaid with hyperspectral pixels and stereo-elevation results under a unified coordinate and vertical reference. The resulting measured depth data could then serve both as the main supervision for model training and as the reference standard for ICESat-2 bias analysis and overall accuracy evaluation.

Based on the above preprocessing, all multi-source observations were finally organized into a point-sample-oriented tabular dataset. Each sample corresponded to a spatial point and contained four types of information: (1) hyperspectral features, namely multi-band reflectance extracted from the corresponding pixel; (2) index features, such as water-related indices including NDWI and MNDWI; (3) environmental and geometric variables, such as acquisition date, coordinates, solar elevation angle, viewing geometry, and elevation-related information; and (4) labels, including measured depth, ICESat-2-derived depth, or cross-year expanded depth. Through this organization, multi-source observations with different origins and formats were unified into heterogeneous tabular samples that can be directly input into the model, thereby laying the data foundation for subsequent TabKAN modeling.

4. Methodology

4.1. Overall Data Flow of the Proposed Framework

To provide a clearer description of the proposed framework, Figure 3 illustrates the complete data flow of TabKAN from multi-source data preprocessing to full-scene bathymetric prediction. The workflow consists of four main stages: multi-source preprocessing, tabular sample construction, model pretraining and joint-supervised training, and final bathymetric inference.

https://doi.org/10.3390/rs18121886

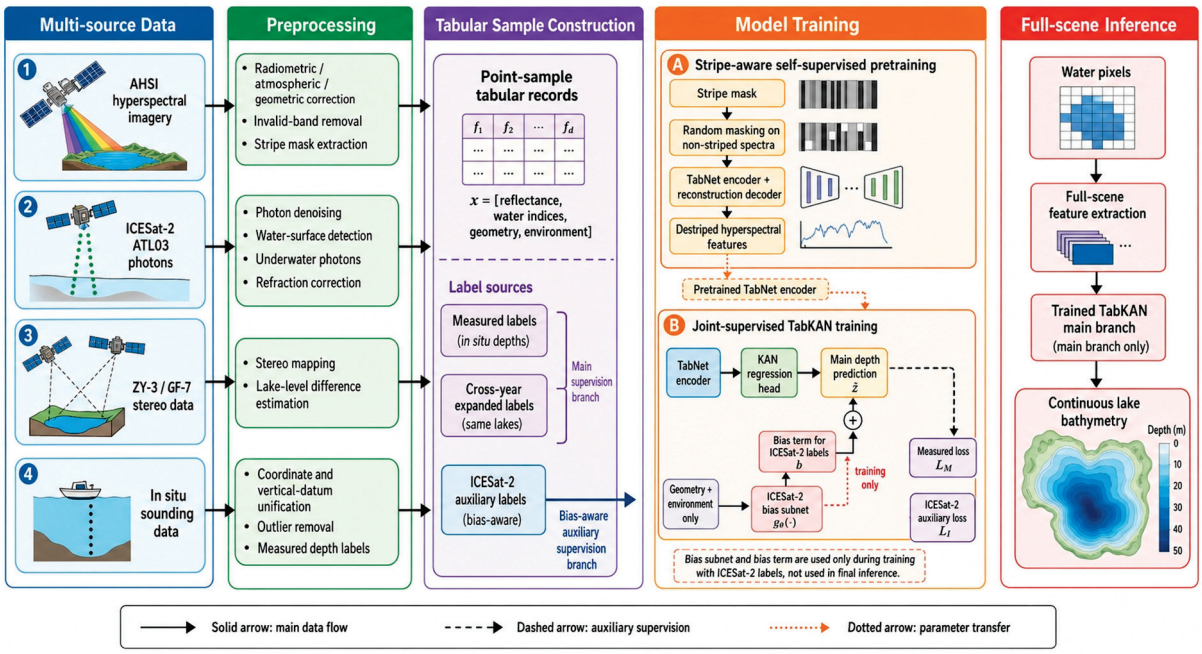


Figure 3. Overall data flow of the proposed TabKAN framework.

First, four types of data are preprocessed separately. For AHSI hyperspectral imagery, radiometric correction, atmospheric correction, geometric correction, invalid-band removal, and stripe-mask extraction are performed to obtain reliable spectral observations. For ICESat-2 ATL03 data, photon denoising, water-surface detection, underwater-photon extraction, and refraction correction are conducted to generate along-track auxiliary depth samples. For ZY-3 and GF-7 stereo mapping data, lake-surface elevations from different years are extracted to estimate interannual lake-level differences. For in situ sounding data, coordinate unification, vertical-datum correction, and outlier removal are performed to produce the primary measured depth labels.

Second, the preprocessed multi-source observations are converted into point-sample-oriented tabular records. Each sample contains four groups of input features: hyperspectral reflectance, water indices, imaging-geometry variables, and environmental variables. The corresponding labels come from three sources: original measured sounding samples, cross-year expanded samples derived from lake-level migration, and ICESat-2 auxiliary samples. The measured and cross-year expanded samples are used to constrain the main bathymetric mapping, whereas ICESat-2 samples are introduced as auxiliary supervision through the bias-aware learning mechanism.

Third, model training contains two closely related parts. The stripe-aware self-supervised pretraining module learns spectral representations from non-striped hyperspectral regions and restores stripe-contaminated spectra. The pretrained TabNet encoder is then transferred to the downstream TabKAN model. In the supervised training stage, the TabNet encoder and KAN regression head form the main depth prediction branch. Meanwhile, ICESat-2 samples are used together with an environmental bias subnet to provide additional along-track constraints without forcing the main branch to directly absorb ICESat-2-specific label bias.

Finally, during full-scene inference, each water pixel is converted into the same tabular feature format as the training samples and then fed into the trained TabKAN model. The final bathymetric map is generated using only the main depth prediction branch. The ICESat-2 bias subnet is used to improve training through auxiliary supervision, but it is not added to the final prediction because the target of inference is the true water depth rather than the reproduction of ICESat-2 label bias.

https://doi.org/10.3390/rs18121886

4.2. TabKAN Architecture

To improve the organization of high-dimensional heterogeneous features and the nonlinear modeling capacity in hyperspectral bathymetric inversion, we construct a backbone model composed of a TabNet encoder and a KAN regression head, referred to as TabKAN. The model is designed for point-sample-based tabular data. Its input is a feature vector composed of hyperspectral reflectance, water indices, imaging geometry, and environmental variables, and the output is the predicted depth at the corresponding sample point. The architecture of TabKAN is illustrated in Figure 4, where Figure 4a shows the TabNet encoder and Figure 4b shows the KAN regression head.

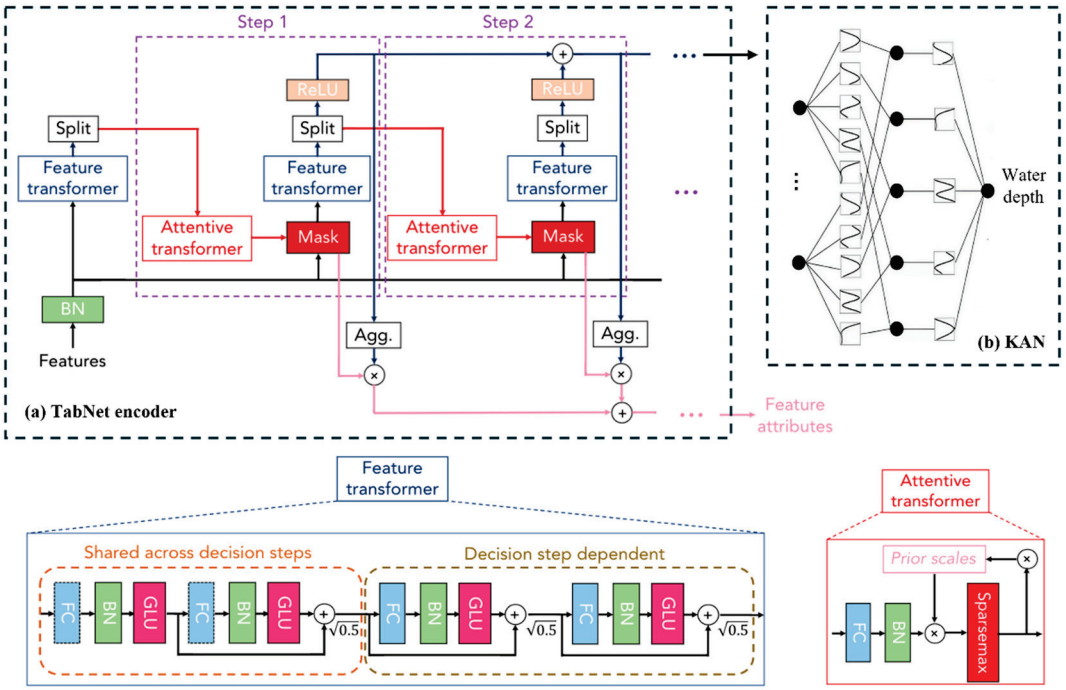


Figure 4. Architecture of the proposed TabKAN model.

As shown in Figure 4a, the TabNet encoder adopts a sequential feature-selection mechanism with multiple decision steps. The raw input features are first batch-normalized and then processed through multiple decision steps. In each step, the feature transformer performs nonlinear mapping on the current input and extracts an intermediate representation. A split operation then separates the features into the part used for decision output and the part used for attention assignment. The attentive transformer generates a sparse feature mask according to the prior information from the previous step, thereby determining which input variables should be emphasized at the next step. Some feature-transformer modules share parameters across steps, whereas others are specific to the current decision step. This design allows the model to maintain a consistent representation capability while progressively learning more refined discriminative information. The outputs of different decision steps are aggregated to form the final encoded representation.

This TabNet structure is particularly suitable for the task considered in this study. On the one hand, the input features include hyperspectral reflectance, water indices, imaging geometry, and environmental variables, which are high-dimensional, heterogeneous, and strongly correlated. On the other hand, the contributions of different bands and auxiliary variables to bathymetric inversion are highly unbalanced. Through the stepwise feature-selection mechanism implemented by the attentive transformer and the masks, the model can automatically highlight the key bands and environmental information most relevant to

https://doi.org/10.3390/rs18121886

water depth while suppressing redundant variables and noisy features, thereby improving the stability of downstream regression.

After obtaining the encoded representation from TabNet, the final depth prediction is carried out by the KAN regression head. As shown in Figure 4b, the key characteristic of KAN is that, unlike conventional multilayer perceptrons that mainly rely on learnable weights between nodes, KAN defines learnable nonlinear functions on the edges of the network and performs inter-layer aggregation through node summation. For a complex regression task such as lake bathymetry, which is influenced jointly by spectral response, water conditions, and environmental information, this design is theoretically better suited to expressing complicated nonlinear mappings.

4.3. Joint Supervision and Bias-Aware Learning with ICESat-2

Although ICESat-2 can provide sparse and high-precision along-track bathymetric constraints, its derived depths cannot be treated as identical to true depths. On the one hand, ICESat-2 depth estimation is affected by multiple factors, including water transparency, noise photons, air–water refraction correction, observation geometry, and bottom-return identification errors. On the other hand, these errors usually do not manifest as a constant offset but instead show clear environmental dependence. If ICESat-2 samples were treated as labels equivalent to measured samples and used directly in a unified training set, the model might erroneously absorb systematic differences in the laser-derived labels into the main optical mapping, thereby weakening the physical consistency and cross-scene robustness of the primary regression relationship. Based on this observation, we introduce a joint-supervision and bias-aware learning mechanism on top of the TabKAN backbone, so that measured samples and ICESat-2 samples can participate jointly in training while their systematic discrepancy is explicitly modeled.

As illustrated in Figure 5, the model output is decomposed into two parts. One is the main depth term produced by the shared backbone network, which can be understood as the shared depth estimate determined by hyperspectral and related features under ideal conditions. The other is an environmental bias term produced by a bias subnet, which is used to describe the systematic discrepancy of ICESat-2 labels relative to the shared depth estimate under specific environmental conditions. For samples with measured labels, the full input feature vector is denoted by x and the corresponding label by y. The sample is fed into the shared TabNet encoder and KAN regression head to produce the main depth prediction:

$$p = f_{\text{main}}(x), \tag{1}$$

where $f_{\text{main}}(\cdot)$ denotes the main mapping composed of the shared encoder and the main regression head. Because measured samples most directly reflect true water depth, the error between p and y is used directly as the supervision signal to constrain the shared backbone to learn the common mapping between hyperspectral/environmental features and true depth.

For ICESat-2 samples, the full input feature vector is denoted by x_{ICESat} and the corresponding ICESat-2 label by y_{ICESat}. As with measured samples, x_{ICESat} is passed through the shared backbone to produce the main depth term $f_{\text{main}}(x_{\text{ICESat}})$. Meanwhile, a feature subset containing only imaging-geometry and environmental variables is extracted from the ICESat-2 sample and denoted by x'_{ICESat}, which is then fed into the KAN bias subnet to estimate the environment-driven systematic bias:

$$b_{\text{env}} = f_{\text{bias}}\left(x'_{\text{ICESat}}\right), \tag{2}$$

https://doi.org/10.3390/rs18121886

where $f_{\text{bias}}(\cdot)$ denotes the bias-modeling function. The final prediction for the ICESat-2 label is then written as

$$p_{\text{ICESat}} = f_{\text{main}}(x_{\text{ICESat}}) + b_{\text{env}}. \tag{3}$$

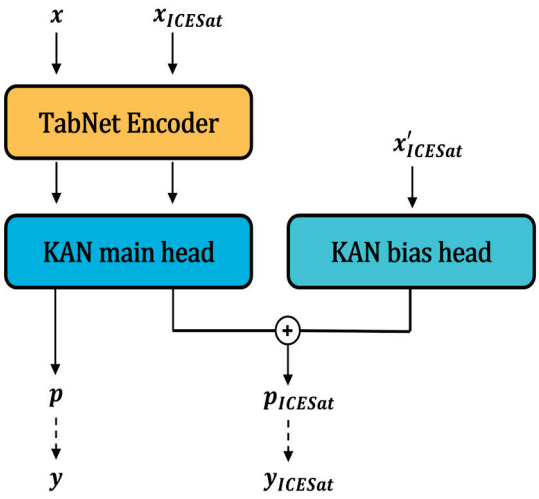


Figure 5. Overall structure of TabKAN under multi-source supervision.

The key point is that the bias subnet does not use hyperspectral reflectance directly, but only geometry and environmental variables. This design aims to constrain the systematic discrepancy between ICESat-2 and measured samples to an "environment-driven bias term" rather than letting the bias branch learn an independent spectrum–depth relationship. In other words, the main regression head learns the shared bathymetric mapping, whereas the bias subnet absorbs the additional deviation caused by observation conditions, water environment, and label-source differences.

Two regression losses are defined separately for measured samples and ICESat-2 samples. For the set of measured samples $\mathcal{D}_m$, a mean-squared-error loss is used to constrain the main regression term:

$$L_m = \frac{1}{|\mathcal{D}_m|} \sum_{(x,y)\in\mathcal{D}_m} (f_{\text{main}}(x) - y)^2. \tag{4}$$

Here, the measured samples include both the original field bathymetry and the cross-year expanded samples that are treated as field-equivalent samples in this study, and together they provide the main supervision for learning the primary mapping.

For the ICESat-2 sample set $\mathcal{D}_i$, the auxiliary supervision loss is defined as

$$L_i = \frac{1}{|\mathcal{D}_i|} \sum_{(x_{\text{ICESat}}, x'_{\text{ICESat}}, y_{\text{ICESat}})\in\mathcal{D}_i} \left(f_{\text{main}}(x_{\text{ICESat}}) + f_{\text{bias}}(x'_{\text{ICESat}}) - y_{\text{ICESat}}\right)^2. \tag{5}$$

The purpose of this loss is not to replace true depth with ICESat-2 labels, but to let ICESat-2 provide additional high-precision constraints through the combination of a main depth term and an environmental bias term without destroying the physical meaning of the main mapping.

The total objective of the model is therefore

$$L = L_m + L_i. \tag{6}$$

This formulation is concise and physically meaningful. L_m constrains the main regression branch to learn the shared mapping between hyperspectral/environmental fea-

https://doi.org/10.3390/rs18121886

tures and true depth, whereas L_i guides the model to explicitly perceive and absorb the environment-dependent bias in ICESat-2 labels during joint training. Through this combined loss, the model can simultaneously benefit from the authenticity of measured samples and the high precision of ICESat-2 samples, while maintaining a reasonable balance between shared mapping and bias correction.

Training is conducted in a staged manner. First, the TabNet encoder and KAN main regression head are pre-trained using only measured samples so that the main mapping can converge to a stable solution under measured supervision. At this stage, only L_m is optimized, and the goal is to let the model first learn the basic depth relationship supported by hyperspectral and environmental information. Then, while keeping the overall backbone stable, ICESat-2 samples are introduced, the bias subnet is activated, and the shared encoder and main head are jointly fine-tuned with a small learning rate while the bias subnet is trained. In this way, the model gradually learns the systematic offset of ICESat-2 labels relative to the shared mapping under different environmental conditions. Finally, all samples are used in joint optimization until the main depth term and the environmental bias term converge under the unified loss.

This staged strategy is motivated by two considerations. First, if measured and ICESat-2 samples are introduced simultaneously from the very beginning, the main regression head may be disturbed by heterogeneous label differences before a stable mapping is formed. Second, learning the main mapping from measured samples first and then performing bias-aware fine-tuning with ICESat-2 samples is more consistent with the logic of "first learn the true depth relationship, then incorporate auxiliary constraints". This allows the model to fully exploit both sample types while keeping a clear division between the main branch and the bias branch.

It should also be emphasized that, although ICESat-2 samples participate in joint supervision through the bias branch during training, the final full-scene prediction targeting true depth mainly uses the output of the main regression branch $f_{\text{main}}(x)$ as the final bathymetric result. In other words, the bias subnet mainly plays its role during training and fine-tuning, helping the model identify and absorb environment-dependent systematic differences in ICESat-2 labels, rather than directly being added to the full-scene prediction during inference. In this way, the auxiliary information provided by ICESat-2 can be fully utilized while preserving the physical interpretability of the final output to the greatest extent.

Overall, the joint-supervision and bias-aware learning mechanism proposed in this section does not merely address the question of "how to use one more type of sample". Rather, it addresses how to preserve the physical consistency of the main mapping under multi-source labels while making full use of the high precision of ICESat-2. By explicitly decomposing the environment-dependent error of ICESat-2 into a bias term, TabKAN becomes not just a hyperspectral regression model, but a multi-source supervision framework capable of simultaneously learning the shared depth mapping and absorbing heterogeneous-label bias. This also lays the foundation for further improving model robustness through stripe suppression and self-supervised pretraining.

4.4. Stripe-Aware Self-Supervised Pretraining and Destriping

Hyperspectral satellite imagery is often affected by the push-broom imaging mechanism, detector-response inconsistency, and instrument instability, which may produce obvious stripe noise in some bands or local regions. For the Resource-1 02E AHSI imagery used in this study, such stripe noise not only degrades visual quality, but more importantly damages radiometric consistency across bands and spectral-curve shapes, thereby directly affecting sample feature extraction, model training, and the stability of full-scene depth

https://doi.org/10.3390/rs18121886

prediction. In point-sample-driven hyperspectral bathymetry, if training samples happen to fall within stripe-contaminated regions, the model may incorrectly learn stripe artifacts as bathymetrically informative features, which in turn weakens cross-temporal and cross-lake generalization. Therefore, stripe processing is not treated here as a simple preprocessing step isolated from the downstream task. Instead, it is jointly designed together with self-supervised representation learning and bathymetric regression in a stripe-aware self-supervised pretraining and destriping strategy.

4.4.1. Stripe Mask Extraction

Stripe noise usually appears as line-shaped radiometric anomalies elongated along the scanning direction, and its response varies across bands. To automatically identify such anomalous regions as much as possible, each AHSI band is first analyzed statistically, and a stripe mask is constructed by combining bandwise statistics and local spatial consistency. Let the image of the b-th band be denoted by I_b. For each band, the column-wise mean, column-wise standard deviation, and the difference response with neighboring columns are calculated to obtain a stripe-strength indicator $s_b(j)$. When a column deviates significantly from its local background in radiometric statistics and the anomaly shows strong consistency along the row direction, it is marked as a potential stripe region. The detection results of all bands are then aggregated spectrally to generate a stripe confidence map, which is thresholded to obtain the final stripe mask:

$$M_{\text{stripe}} \in \left\{0,1\right\}^{H\times W\times B}, \tag{7}$$

where $M_{\text{stripe}} = 1$ indicates that the corresponding spatial location and band are judged to be stripe-contaminated, whereas $M_{\text{stripe}} = 0$ indicates that the position is considered relatively reliable.

It should be emphasized that the purpose of constructing M_{stripe} is not to obtain absolutely precise noise labels, but rather to identify regions that should not be directly used as supervision targets. In other words, the stripe mask primarily serves to mark regions without reliable ground-truth spectra rather than acting as a direct supervision mask in the self-supervised loss.

4.4.2. Self-Supervised Pretraining on Non-Striped Regions

Because stripe-contaminated regions lack clean and reliable reference spectra, self-supervised pretraining is performed only on non-striped regions. Specifically, let the original hyperspectral vector of a pixel be

$$z \in \mathbb{R}^B, \tag{8}$$

where B is the number of input bands. The corresponding stripe-indicator vector is denoted by

$$m_{\text{stripe}} \in \left\{0,1\right\}^B. \tag{9}$$

During self-supervised pretraining, reconstruction supervision is not imposed on stripe positions. Instead, a random mask m_{rand} is generated only on the non-striped dimensions satisfying $m_{\text{stripe}} = 0$. Therefore, the random mask and stripe mask satisfy

$$m_{\text{rand}} \odot m_{\text{stripe}} = 0, \tag{10}$$

https://doi.org/10.3390/rs18121886

which means that random masking is applied only to non-striped positions. Stripe positions and randomly masked positions are both treated as unavailable features, resulting in the pretraining input vector

$$\widetilde{z} = \left(1 - m_{\text{stripe}} - m_{\text{rand}}\right) \odot z. \tag{11}$$

In the network structure, we adopt the TabNet self-supervised pretraining framework, in which the TabNet encoder is followed by the corresponding reconstruction decoder, as shown in Figure 6.

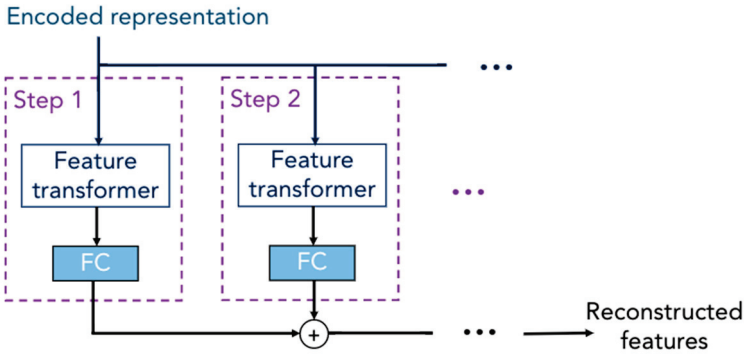


Figure 6. Schematic diagram of the TabNet decoder used in self-supervised reconstruction.

Let the encoder be denoted by $f_{\text{enc}}(\cdot)$ and the decoder by $g_{\text{dec}}(\cdot)$. The reconstructed spectrum is then

$$\hat{z} = g_{\text{dec}}\left(f_{\text{enc}}\left(\widetilde{z}\right)\right). \tag{12}$$

Since only the randomly masked positions have reliable true values, the self-supervised reconstruction loss is computed only on positions where $m_{\text{rand}} = 1$:

$$L_{ssl} = \frac{1}{\| m_{rand} \|_1} \| m_{rand} \odot (\hat{z} - z) \|_1. \tag{13}$$

The core meaning of this design is that the model learns how to reconstruct randomly masked normal spectra from uncorrupted and unmasked spectral information, rather than being supervised directly on stripe-contaminated positions. This avoids using unreliable stripe values as targets and allows the network to learn the internal spectral dependencies of normal hyperspectral signals.

4.4.3. Stripe Restoration and Parameter Transfer

After self-supervised pretraining is completed, the trained encoder–decoder structure is used to restore stripe-contaminated regions. For any spectral vector z affected by stripes, only its non-striped part is retained as input:

$$\widetilde{z}_{stripe} = \left(1 - m_{stripe}\right) \odot z, \tag{14}$$

and the pretrained TabNet encoder and decoder are used to obtain

$$\hat{z}_{stripe} = g_{dec}\left(f_{enc}\left(\widetilde{z}_{stripe}\right)\right). \tag{15}$$

The reconstructed values are then used to replace the original contaminated values only at the stripe positions, yielding the corrected spectral vector

$$z_{clean} = \left(1 - m_{stripe}\right) \odot z + m_{stripe} \odot \hat{z}_{stripe}. \tag{16}$$

https://doi.org/10.3390/rs18121886

This process can be interpreted as a targeted restoration strategy based on the statistical regularities of normal spectra: the model does not rewrite the entire spectrum, but only replaces positions identified as stripe-contaminated, thereby preserving valid original observations to the greatest extent while reducing the influence of stripe noise on downstream tasks.

During the downstream bathymetric inversion stage, the pretrained TabNet encoder parameters are retained and used to initialize the encoder in the TabKAN backbone described in Section 4.1, whereas the self-supervised decoder $g_{\mathrm{dec}}(\cdot)$ is discarded after pretraining and does not participate in the subsequent regression training. In this way, the spectral-dependency structure and anti-noise representation ability learned during self-supervised pretraining can be naturally transferred to the downstream bathymetric inversion task.

5. Experiments and Results

5.1. Experimental Settings

After multi-source data preprocessing and sample organization, the final experimental dataset contained 104,616 valid point-sample records covering five lakes and the period 2020–2025. Because point samples exhibit obvious spatial correlation, directly splitting all samples at random could place spatially adjacent samples in both the training and test sets, thereby overestimating the model's generalization ability. To reduce this effect as much as possible, samples were first grouped by lake and year and then stratified further by spatial position. The data were split into training, validation, and test sets in a ratio of 7:1:2. Auxiliary samples derived from ICESat-2 were likewise used only during training and were not involved in validation or testing. Consequently, model selection and final accuracy evaluation were all based on independent measured test samples, ensuring that the reported performance genuinely reflects the ability of each model to predict reference depth.

To evaluate the stability of the results with respect to data partitioning, the above lake–year–spatial context-stratified splitting procedure was repeated five times using different random seeds. For each repeated split, all ablation settings, benchmark methods, and the proposed TabKAN model were retrained under the same input feature system and evaluation protocol. The final results are reported as mean ± standard deviation.

In addition, statistical significance tests were conducted to verify whether the improvement of TabKAN over the reference methods was significant. Since all methods were evaluated on the same measured test samples, paired Wilcoxon signed-rank tests were used. Because RMSE is the primary ranking metric in this study, the tests were performed on paired squared prediction errors between each benchmark method and TabKAN. The resulting p-values were adjusted using the Holm–Bonferroni correction for multiple comparisons. A difference was considered statistically significant when $p_{adj} < 0.05$.

To make the experimental results more physically interpretable, this study reports not only overall metrics but also depth-stratified evaluation. Previous studies have shown that optical- or ICESat-based depth inversion is generally more sensitive in shallow water. As depth increases, spectral responses tend to become more homogeneous, while laser penetration and bottom identification also become increasingly uncertain. Therefore, the test samples were divided into a shallow-water interval of 0–30 m and a deep-water interval greater than 30 m, and the accuracy within each interval was evaluated separately. The purpose of this setting is not to artificially complicate the interpretation, but to avoid drawing conclusions from a single metric that mixes errors from physically different sensitivity regimes. For the depth-stratified evaluation, the sample size of each depth interval

https://doi.org/10.3390/rs18121886

was also reported. The measured test set contained 20,924 samples in total, including 15,770 samples in the 0–30 m interval and 5154 samples in the >30 m interval.

Three common indicators were used to comprehensively evaluate the inversion accuracy: root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). Let the number of test samples be N, the true depth be y_i, and the predicted depth be $\hat{y}_i$. Then,

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}, \tag{17}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|, \tag{18}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \tag{19}$$

where RMSE is more sensitive to large errors and reflects overall predictive stability, MAE directly describes the average absolute deviation, and R^2 measures the goodness of fit to the variation in the true bathymetry. Subsequent analysis is based mainly on these three indicators.

Regarding comparative methods, and considering the need to balance representativeness, comparability, and space, eight benchmark methods were selected: the classical empirical Stumpf band-ratio model [5], the instance-learning-based KNN model [35], two Bagging tree methods, Random Forest [36] and ExtraTrees [37], two Boosting tree methods, XGBoost [38] and HistGradientBoosting [39], the generic deep neural regression model MLP [40], and the deep tabular learning model TabNet [12]. Stumpf represents the traditional empirical route in remote sensing bathymetry; KNN represents local sample-similarity learning; Random Forest and ExtraTrees represent two typical Bagging tree ensembles; XGBoost and HistGradientBoosting represent mainstream Boosting tree models; MLP serves as a generic deep-learning baseline; and TabNet is the deep tabular model most closely related to the proposed method, making it suitable for evaluating the actual gain brought by the KAN head, joint supervision, and stripe-aware self-supervised pretraining. All data-driven baselines used the same input feature system and the same training, validation, and test splits to ensure fair comparison.

In implementation, all data-driven methods took as input the same set of hyperspectral reflectance, water indices, imaging geometry, and environmental variables. The validation set was used for hyperparameter selection, training control, and early stopping, whereas the test set was used only once after training for final evaluation. This experimental design allows not only a fair comparison of the overall predictive ability of different methods under the same data conditions, but also an explicit analysis of whether the proposed method can more effectively exploit high-dimensional hyperspectral features together with sparse laser-derived constraints.

5.2. Ablation Experiments and Analysis

To systematically evaluate the contribution of each component in TabKAN, a series of ablation experiments were designed under the same data split, input feature system, and evaluation metrics. Four aspects were analyzed: (1) the incorporation of auxiliary features such as water indices, imaging geometry, and environmental variables; (2) the improvement brought by the KAN regression head over the main regression structure; (3) the enhancement provided by multi-year sample augmentation; and (4) the gains brought by ICESat-2 joint supervision and stripe-aware self-supervised pretraining. Because many

https://doi.org/10.3390/rs18121886

shoreline samples have 0 m or near-zero depth, the main text reports RMSE, MAE, and R^2 as the primary metrics. The ablation results are listed in Table 1.

Table 1. Ablation results for different module combinations. Values are reported as mean ± standard deviation over five repeated data splits. The best performance in each column is highlighted in bold.

ID	Model Setting	RMSE (m)	MAE (m)	R^2
A1	TabNet (hyperspectral reflectance only)	6.73 ± 0.48	3.64 ± 0.39	0.90 ± 0.01
A2	A1 + water indices + environment variables	6.35 ± 0.35	3.29 ± 0.26	0.91 ± 0.02
A3	A2 + KAN main regression head	5.42 ± 0.45	3.06 ± 0.34	0.93 ± 0.03
A4	A3 + multi-year sample augmentation	4.41 ± 0.20	1.62 ± 0.12	0.96 ± 0.01
A5	A4 + ICESat-2 joint supervision	3.80 ± 0.14	**1.53** ± 0.08	0.97 ± 0.01
A6	A5 + stripe-aware self-supervised pretraining	**3.05** ± 0.10	1.57 ± 0.06	**0.98** ± 0.01

Table 1 shows a clear and stable trend of performance improvement as more modules are introduced. All values are reported as mean ± standard deviation over five repeated data splits; therefore, the following discussion mainly focuses on the mean values, while the standard deviations are used to assess the robustness of the observed trends.

The baseline model A1, which uses a standard TabNet structure and only hyperspectral reflectance as input, yields a mean RMSE of 6.73 m, an MAE of 3.64 m, and an R^2 of 0.90. This result indicates that TabNet already has a certain capability to organize high-dimensional spectral features, but hyperspectral reflectance alone is still insufficient for fully capturing depth variation in complex plateau-lake environments.

After introducing water indices, imaging geometry, and environmental variables, A2 reduces the RMSE to 6.35 m and the MAE to 3.29 m. Although the improvement is moderate, it is consistent across repeated data splits. This suggests that lake bathymetric inversion is not merely a problem of spectral fitting. The same water depth may correspond to different spectral responses under different acquisition times, viewing geometries, lake conditions, and environmental backgrounds. Therefore, incorporating auxiliary variables helps the model better describe cross-temporal and cross-scene spectral differences.

Further replacing the conventional output layer with a KAN main regression head leads to A3, whose RMSE decreases to 5.42 m and R^2 increases to 0.93. Compared with A2, this corresponds to an RMSE reduction of approximately 14.6%. This improvement demonstrates that the KAN head is beneficial for representing complex nonlinear feature–depth relationships. Compared with a conventional MLP-style regression output, KAN introduces learnable nonlinear functions and is therefore better suited to describing continuous depth variations across different basin regions.

After introducing multi-year sample augmentation, A4 further improves the performance, reducing RMSE to 4.41 m and MAE to 1.62 m. The decrease in MAE is particularly evident, indicating that multi-year sample expansion does not merely increase the number of training samples, but also improves the temporal coverage of water-level states and lake conditions. This enables the model to learn a more stable cross-temporal bathymetric relationship and reduces local point-wise errors.

The comparison between A5 and A4 shows that ICESat-2 joint supervision is another important source of improvement. After introducing ICESat-2 constraints, RMSE decreases from 4.41 m to 3.80 m, and MAE further decreases from 1.62 m to 1.53 m. This indicates that sparse but high-precision satellite-laser observations can effectively complement the spatial distribution of measured samples and strengthen constraints on the overall basin morphology, especially in medium- and deep-water regions where purely optical signals become less sensitive.

https://doi.org/10.3390/rs18121886

On this basis, A6 further introduces stripe-aware self-supervised pretraining and achieves the best overall performance in terms of RMSE and R^2, with an RMSE of 3.05 m and an R^2 of 0.98. Compared with A1, the full model reduces RMSE and MAE by approximately 54.7% and 56.9%, respectively. It should be noted that the MAE of A6 is slightly higher than that of A5. However, both RMSE and R^2 are further improved, suggesting that the main contribution of stripe-aware self-supervised pretraining lies in suppressing larger errors, improving radiometric consistency, and enhancing the stability of full-scene bathymetric reconstruction rather than uniformly reducing all local absolute errors. The relatively small standard deviation of A6 also indicates that this improvement is stable across repeated data splits.

The visual comparison before and after destriping is shown in Figure 7, where the stripe-aware self-supervised strategy clearly reduces stripe interference, improves radiometric consistency and spatial continuity, and thus provides more stable input features for downstream bathymetric inversion.

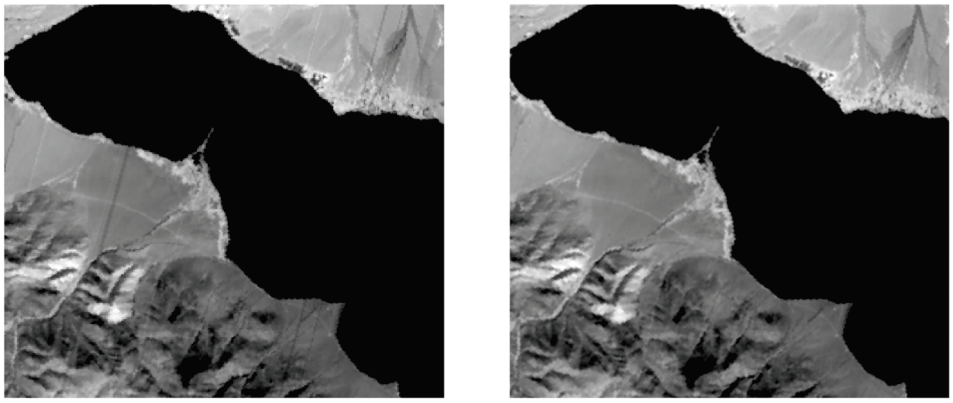


Figure 7. Comparison of hyperspectral imagery before and after destriping.

To further evaluate the reconstruction quality quantitatively, a pseudo-clean reconstruction experiment was conducted on non-striped hyperspectral regions. Because paired clean reference images are unavailable for real stripe-contaminated satellite hyperspectral data, pixels from visually and statistically non-striped regions were used as pseudo-clean references. A subset of pixels and spectral bands was randomly masked, and the masked spectral entries were restored using both QRNN3D [41] and the proposed stripe-aware self-supervised reconstruction method. The restored results were then compared with the original pseudo-clean observations using spectral angle mapper (SAM), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM). To ensure the robustness and statistical reliability of the evaluation, the experiment was independently repeated five times with different random masking patterns, and the average results were reported. The quantitative results are reported in Table 2.

Table 2. Quantitative assessment of hyperspectral reconstruction on non-striped pseudo-clean regions. The best performance in each column is highlighted in bold.

Model	SAM (°)	PSNR (dB)	SSIM
QRNN3D	2.25 ± 0.01	51.36 ± 0.19	0.92 ± 0.01
Proposed stripe-aware SSL	**1.02** ± 0.01	**60.25** ± 0.18	**0.94** ± 0.02

As shown in Table 2, the proposed method achieves better reconstruction quality than QRNN3D. The SAM decreases from 2.25° for QRNN3D to 1.02° for the proposed method,

https://doi.org/10.3390/rs18121886

indicating better preservation of spectral shape. The PSNR increases from 51.36 dB to 60.25 dB, and the SSIM increases from 0.92 to 0.94, suggesting that the proposed method more effectively restores the masked spectral information while maintaining spatial structural consistency. These quantitative results are consistent with the visual comparison and further confirm that the stripe-aware self-supervised reconstruction provides more reliable hyperspectral inputs for bathymetric inversion.

Figure 8 further shows the local bathymetric inversion results for Geren Co obtained using the original image, the QRNN3D-restored image, and the image restored by the proposed method. The visual comparison indicates that the proposed stripe-aware self-supervised restoration achieves a favorable denoising effect and supports more stable bathymetric mapping, which is consistent with the quantitative results in Table 2.

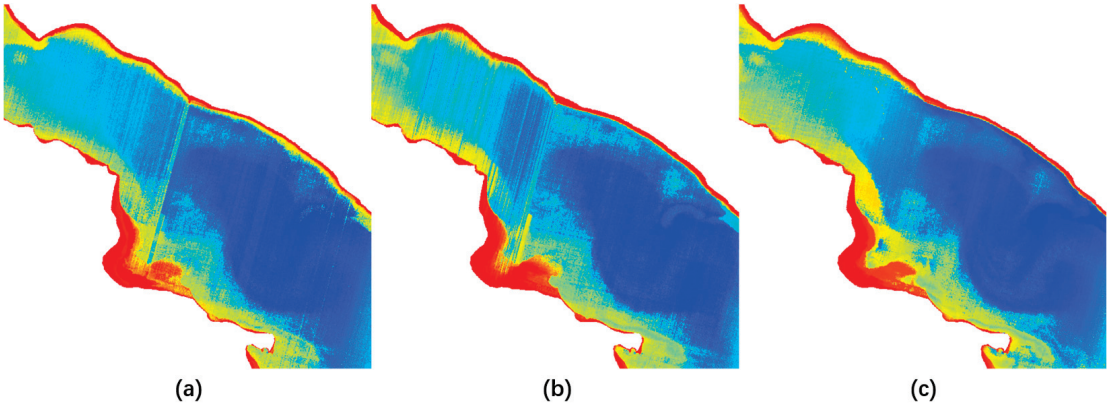


Figure 8. Local bathymetric inversion results for Geren Co using different hyperspectral inputs: (**a**) original image without destriping, (**b**) QRNN3D-restored image, and (**c**) image restored by the proposed stripe-aware self-supervised method.

To further analyze the behavior of the proposed method in different depth intervals, the depth-stratified performance of the TabNet baseline and the full TabKAN model was compared. Considering that the 20–40 m interval contains relatively few samples and tends to exhibit strong statistical fluctuation, a two-interval evaluation scheme was finally adopted, namely 0–30 m and greater than 30 m. The results are shown in Table 3.

Table 3. Depth-stratified performance of TabNet and TabKAN. N denotes the number of measured test samples in each depth interval. Accuracy values are reported as mean ± standard deviation over five repeated data splits.

Model	Depth Interval	N	RMSE (m)	MAE (m)	R^2
TabNet	All	20,924	6.73 ± 0.48	3.64 ± 0.39	0.90 ± 0.01
	0–30 m	15,770	5.39 ± 0.50	2.41 ± 0.30	0.59 ± 0.08
	>30 m	5154	9.96 ± 0.57	7.18 ± 0.66	0.41 ± 0.06
TabKAN	All	20,924	3.05 ± 0.10	1.57 ± 0.06	0.98 ± 0.01
	0–30 m	15,770	2.37 ± 0.13	0.92 ± 0.07	0.90 ± 0.02
	>30 m	5154	4.46 ± 0.18	3.19 ± 0.12	0.88 ± 0.05

To further analyze the behavior of the proposed method in different depth intervals, the depth-stratified performance of the TabNet baseline and the full TabKAN model was compared. As shown in Table 3, the test samples include 20,924 samples in total, with 15,770 samples in the 0–30 m interval and 5154 samples in the >30 m interval. Reporting the sample size for each interval helps clarify the statistical basis of the depth-stratified evaluation.

TabKAN significantly outperforms TabNet in both depth intervals. In the 0–30 m interval, the RMSE decreases from 5.39 m to 2.37 m, corresponding to a reduction of approximately 56.0%, while R^2 increases from 0.59 to 0.90. This confirms that the proposed framework improves shallow-water bathymetric inversion, where optical signals still contain relatively strong bottom information.

In the >30 m interval, the improvement is also substantial. The RMSE decreases from 9.96 m to 4.46 m, and the MAE decreases from 7.18 m to 3.19 m. Although the absolute error remains higher in deep water than in shallow water, which is expected because of stronger optical attenuation and weaker bottom-reflected signals, the performance degradation of TabKAN is much smaller than that of TabNet. The increase in R^2 from 0.41 to 0.88 further indicates that TabKAN better preserves the overall depth variation and basin morphology in deeper regions.

These depth-stratified results suggest that the proposed framework is not only effective in optically sensitive shallow waters, but also provides stronger constraints for deeper-water reconstruction through the combined use of multi-source features, ICESat-2 supervision, and stripe-aware pretraining.

5.3. Comparative Results

Table 4 reports the overall comparison between TabKAN and eight representative benchmark methods over five repeated data splits. In addition to RMSE, MAE, and R^2, the last column provides the Holm-adjusted p-values obtained from paired Wilcoxon signed-rank tests between each benchmark method and TabKAN. These statistical tests are used to examine whether the observed performance gains are significant rather than being caused by a particular data partition.

Table 4. Overall performance comparison over five repeated data splits. Values are reported as mean ± standard deviation. p_{adj} denotes the Holm-adjusted p-value from paired Wilcoxon signed-rank tests comparing squared prediction errors between each reference method and TabKAN. The best performance in each column is highlighted in bold.

Method	RMSE (m)	MAE (m)	R^2	p_{adj} vs. TabKAN
Stumpf	10.35 ± 0.08	7.59 ± 0.05	0.77 ± 0.01	<0.001
MLP	7.52 ± 0.43	4.60 ± 0.28	0.86 ± 0.01	<0.001
XGBoost	7.28 ± 0.09	3.87 ± 0.03	0.89 ± 0.01	<0.001
HistGradientBoosting	6.91 ± 0.08	3.78 ± 0.04	0.89 ± 0.01	<0.001
TabNet	6.73 ± 0.48	3.64 ± 0.39	0.90 ± 0.01	<0.001
KNN	6.06 ± 0.11	5.01 ± 0.04	0.92 ± 0.01	<0.001
Random Forest	5.79 ± 0.08	3.04 ± 0.03	0.93 ± 0.01	<0.001
ExtraTrees	4.02 ± 0.07	1.73 ± 0.03	0.96 ± 0.01	<0.001
TabKAN	**3.05** ± 0.10	**1.57** ± 0.06	**0.98** ± 0.01	-

The traditional empirical Stumpf model performs the worst among all methods, with an RMSE of 10.35 m and an R^2 of 0.77. This indicates that, for Tibetan Plateau lakes characterized by complex water conditions, variable basin morphology, and strong cross-temporal differences, a fixed empirical band-ratio relationship is insufficient to accurately capture the relationship between hyperspectral features and water depth. Although the generic MLP regression model substantially improves upon Stumpf, its RMSE remains relatively high, suggesting that a simple fully connected neural network is still limited in organizing high-dimensional heterogeneous tabular features.

The two Boosting tree models, XGBoost and HistGradientBoosting, achieve better performance than MLP and Stumpf, indicating that ensemble learning is effective for

structured hyperspectral samples. TabNet further improves the results by introducing explicit feature selection and sequential decision mechanisms, which are useful for high-dimensional hyperspectral and environmental variables. KNN obtains a lower RMSE than TabNet, but its MAE remains relatively high, implying that local sample-similarity learning may be sensitive to the distribution of neighboring samples and less stable for representing continuous basin-scale depth gradients.

Among the benchmark methods, Random Forest and ExtraTrees are the most competitive. In particular, ExtraTrees achieves the best baseline performance, with an RMSE of 4.02 m, an MAE of 1.73 m, and an R^2 of 0.96. This highlights the strong capability of randomized tree ensembles in handling nonlinear tabular regression tasks. However, some limitations remain because tree-based ensembles mainly rely on piecewise partitioning of the feature space and may still be insufficient for representing smooth and continuous bathymetric variations across complex lake basins.

In contrast, the proposed TabKAN achieves the best performance in all three accuracy metrics, with an RMSE of 3.05 m, an MAE of 1.57 m, and an R^2 of 0.98. Compared with the strongest benchmark, ExtraTrees, TabKAN reduces RMSE by approximately 24.1% and MAE by approximately 9.2%. Compared with TabNet, which is the benchmark most closely related to the proposed backbone, TabKAN reduces RMSE by approximately 54.7% and MAE by approximately 56.9%. These improvements indicate that the proposed method benefits not only from TabNet-based feature organization, but also from the additional nonlinear representation capacity of the KAN head, the geometric constraints provided by ICESat-2 supervision, the temporal coverage introduced by multi-year sample augmentation, and the improved input quality obtained from stripe-aware self-supervised pretraining.

The statistical tests further support this conclusion. All benchmark methods show $p_{adj} < 0.001$ when compared with TabKAN, indicating that the improvement of TabKAN is statistically significant after multiple-comparison correction. Moreover, the performance gap between TabKAN and the strongest benchmark is clearly larger than the standard deviations across repeated splits, suggesting that the advantage of TabKAN is stable rather than dependent on a particular train–validation–test partition.

From the perspective of spatial reconstruction, the comparative results also suggest different characteristics among methods. Empirical and simple neural models are more likely to be affected by regional spectral variability. Tree-based ensemble methods are highly competitive in point-wise accuracy, but may still have limitations in representing smooth basin-scale bathymetric gradients. TabKAN, by combining adaptive feature selection, nonlinear regression, auxiliary laser constraints, and stripe-aware spectral restoration, provides a more stable framework for reconstructing continuous lake-bottom morphology. Overall, the comparative experiments demonstrate that TabKAN outperforms empirical models, conventional machine-learning models, and deep tabular learning baselines in both statistical accuracy and robustness across repeated data splits.

6. Conclusions and Future Work

This study addressed several challenges in bathymetric inversion of Tibetan Plateau lakes, including limited samples, obvious stripe noise in hyperspectral imagery, pronounced cross-temporal variability, and inconsistent label sources between ICESat-2 and measured samples, by proposing a lake bathymetric reconstruction method, TabKAN, that combines hyperspectral imagery, ICESat-2 laser constraints, and multi-year sample expansion. The method uses TabNet as the encoder to organize high-dimensional heterogeneous tabular features, KAN as the main regression head to enhance complex nonlinear mapping, and a bias-aware joint-supervision mechanism to incorporate measured sam-

https://doi.org/10.3390/rs18121886

ples and ICESat-2 samples into a unified training framework. In addition, a stripe-aware self-supervised pretraining and restoration pipeline was developed to improve hyperspectral input quality and provide a more robust initialization for downstream bathymetric inversion. Experiments based on multi-source data from five lakes, namely Anglaren Co, Caiduo Chaka, Cuoe, Geren Co, and Qixiang Co, during 2020–2025 show that the proposed method outperforms traditional empirical models and common data-driven methods in overall accuracy, depth-stratified performance, and spatial continuity of lake-bottom reconstruction.

The main conclusions can be summarized in three aspects. First, compared with models relying only on hyperspectral reflectance, the joint introduction of water indices, imaging geometry, and environmental variables substantially improves the adaptability of the model to spectral differences across times and lakes, indicating that hyperspectral bathymetric inversion is fundamentally a regression problem driven by the interaction of multiple heterogeneous features rather than a simple spectrum-fitting problem. Second, the combination of TabNet as feature encoder and KAN as regression head more effectively organizes high-dimensional tabular features and expresses complex nonlinear relationships. On this basis, ICESat-2 joint supervision further strengthens constraints on the overall basin shape, whereas stripe-aware self-supervised pretraining improves spectral quality and training stability. Third, the depth-stratified evaluation shows that the proposed method is most reliable in the shallow-water regime, where optical sensitivity is strongest, and still maintains relatively good continuity and overall reconstruction capability in deeper water, although the error increases. This suggests that the method is better understood as a multi-source-constrained bathymetric reconstruction framework rather than a purely optical direct depth-retrieval method. Beyond methodological improvement, the proposed framework also has practical significance for lake monitoring and hydrological applications. Continuous bathymetric maps generated by TabKAN can support the extraction of lake-basin morphology, identification of nearshore slopes and deep troughs, construction of area–level–volume relationships, and estimation of lake water storage. For Tibetan Plateau lakes, where field surveys are difficult and bathymetric information remains scarce, such products can provide useful support for water-balance analysis, regional water-resource assessment, and studies of climate-driven lake change.

Despite the encouraging results, several issues deserve further investigation. Taken together, the results suggest that Tibetan Plateau lake bathymetric inversion is moving from single-source optical retrieval toward multi-source constrained reconstruction; however, further work is still needed to optimize how different information sources are fused and how uncertainty is controlled, especially in deep-water and optically complex conditions. First, bathymetric reconstruction in deep-water areas still faces considerable physical uncertainty, and future work should incorporate stronger basin-shape constraints, more measured data, more reliable underwater-photon extraction algorithms, and uncertainty-aware or adaptive weighting strategies for fusing hyperspectral observations, ICESat-2 constraints, stereo-derived lake-level information, and environmental variables to improve the credibility of deep-water predictions. Second, the multi-year sample expansion proposed in this study is based on the assumption that lake-bottom morphology remains relatively stable at the inter-annual scale. For lakes undergoing rapid bottom changes or pronounced shoreline adjustments, this assumption still needs to be examined more carefully. Third, although the current model, being a point-sample-driven tabular learning framework, is advantageous for organizing high-dimensional heterogeneous features, its use of local spatial context remains limited. Future work could explore integrating tabular models with convolutional networks, graph networks, or Transformer architectures to enhance the representation of fine-scale local bathymetric structures. Fourth, the general-

ization of the trained model to water bodies beyond high-altitude Tibetan Plateau lakes should be treated with caution. The five lakes used in this study provide a useful test bed for evaluating cross-lake and cross-temporal robustness within the plateau-lake domain, but they do not fully represent artificial reservoirs, highly eutrophic lakes, strongly turbid lakes, or shallow lakes with dense aquatic vegetation. In such water bodies, chlorophyll-a, suspended sediments, colored dissolved organic matter, aquatic vegetation, and different trophic states may substantially alter the spectrum–depth relationship and weaken the bottom-reflected signal. For artificial reservoirs or strongly regulated lakes, sediment accumulation near inflows, dredging, engineered shorelines, and rapid water-level regulation may also change local bottom morphology and weaken the lake-bed stability assumption used in interannual sample augmentation. Therefore, TabKAN should be regarded as a transferable modeling framework rather than a universal pretrained model. Applications to other regions or water-body types should incorporate local calibration samples, water-quality variables, transfer learning or domain adaptation, and uncertainty screening to reduce the risk of out-of-domain prediction. Overall, TabKAN provides a new technical route for multi-source remote-sensing-based bathymetric reconstruction of plateau lakes and lays a foundation for subsequent applications to larger regions and a wider range of lake types.

Author Contributions: Conceptualization, C.Z. and Y.Z.; methodology, Y.Z.; software, Y.Z.; validation, Y.Z., M.P., Q.Z. and X.S.; formal analysis, Y.Z.; investigation, Y.Z.; resources, C.Z., L.C., N.W. and F.B.; data curation, Y.Z. and M.P.; writing—original draft preparation, Y.Z.; writing—review and editing, C.Z., Y.Z., M.P., Q.Z., X.S., L.C., N.W. and F.B.; visualization, Y.Z. and M.P.; supervision, C.Z.; project administration, C.Z.; funding acquisition, C.Z. and F.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the China Geological Survey under the project "Basic Geological Survey by Remote Sensing" (Grant No. DD202301016) and by the Ministry of Natural Resources of the People's Republic of China under the project "Research on Geological High-Resolution Hyperspectral Satellite Imaging Technology System" (Grant No. 2025ZD1008601).

Data Availability Statement: The data presented in this study are not publicly available due to confidentiality restrictions associated with the source geospatial datasets. Access to the data may be considered by the corresponding author upon reasonable request and with permission from the data-owning institutions, subject to applicable confidentiality regulations.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Han, X.; Zhang, G.; Wang, J.; Tseng, K.-H.; Li, J.; Woolway, R.I.; Shum, C.K.; Xu, F. Reconstructing Tibetan Plateau Lake Bathymetry Using ICESat-2 Photon-Counting Laser Altimetry. *Remote Sens. Environ.* **2024**, *315*, 114458. [CrossRef]
2. Khazaei, B.; Read, L.K.; Casali, M.; Sampson, K.M.; Yates, D.N. GLOBathy, the Global Lakes Bathymetry Dataset. *Sci. Data* **2022**, *9*, 36. [CrossRef] [PubMed]
3. Jiang, L.; Nielsen, K.; Andersen, O.B. Improvements in Mountain Lake Monitoring from Satellite Altimetry over the Past 30 Years–Lessons Learned from Tibetan Lakes. *Remote Sens. Environ.* **2023**, *295*, 113702.
4. Lyzenga, D.R. Passive Remote Sensing Techniques for Mapping Water Depth and Bottom Features. *Appl. Opt.* **1978**, *17*, 379–383. [CrossRef]
5. Stumpf, R.P.; Holderied, K.; Sinclair, M. Determination of Water Depth with High-resolution Satellite Imagery over Variable Bottom Types. *Limnol. Oceanogr.* **2003**, *48*, 547–556. [CrossRef]
6. Ashphaq, M.; Srivastava, P.K.; Mitra, D. Review of Near-Shore Satellite Derived Bathymetry: Classification and Account of Five Decades of Coastal Bathymetry Research. *J. Ocean Eng. Sci.* **2021**, *6*, 340–359. [CrossRef]
7. Lee, Z.; Carder, K.L.; Mobley, C.D.; Steward, R.G.; Patch, J.S. Hyperspectral Remote Sensing for Shallow Waters: 2. Deriving Bottom Depths and Water Properties by Optimization. *Appl. Opt.* **1999**, *38*, 3831–3843. [CrossRef]

8. Ma, S.; Tao, Z.; Yang, X.; Yu, Y.; Zhou, X.; Li, Z. Bathymetry Retrieval from Hyperspectral Remote Sensing Data in Optical-Shallow Water. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 1205–1212. [CrossRef]
9. Zhang, Y.; Huang, R.; Yu, K.; Fan, M.; Zhou, G. Estimation of Shallow Water Depth based on Satellite Hyperspectral Images. *J. Geo-Inf. Sci.* **2020**, *22*, 1567. (In Chinese) [CrossRef]
10. Junbo, W.; Ping, P.; Qingfeng, M.; Liping, Z. Investigation of Water Depth, Water Quality and Modern Sedimentation Rate in Mapam Yumco and La'ang Co, Tibet. *J. Lake Sci.* **2013**, *25*, 609–616. [CrossRef]
11. Ma, Y.; Xu, N.; Liu, Z.; Yang, B.; Yang, F.; Wang, X.H.; Li, S. Satellite-Derived Bathymetry Using the ICESat-2 Lidar and Sentinel-2 Imagery Datasets. *Remote Sens. Environ.* **2020**, *250*, 112047. [CrossRef]
12. Arik, S.Ö.; Pfister, T. Tabnet: Attentive Interpretable Tabular Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 6679–6687.
13. Liu, Z.; Wang, Y.; Vaidya, S.; Ruehle, F.; Halverson, J.; Soljačić, M.; Hou, T.Y.; Tegmark, M. KAN: Kolmogorov-Arnold Networks. In Proceedings of the International Conference on Learning Representations 2025 (ICLR 2025), Singapore, 24–28 April 2025.
14. Markus, T.; Neumann, T.; Martino, A.; Abdalati, W.; Brunt, K.; Csatho, B.; Farrell, S.; Fricker, H.; Gardner, A.; Harding, D. The Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2): Science Requirements, Concept, and Implementation. *Remote Sens. Environ.* **2017**, *190*, 260–273. [CrossRef]
15. Thomas, N.; Pertiwi, A.P.; Traganos, D.; Lagomasino, D.; Poursanidis, D.; Moreno, S.; Fatoyinbo, L. Space-Borne Cloud-Native Satellite-Derived Bathymetry (SDB) Models Using ICESat-2 And Sentinel-2. *Geophys. Res. Lett.* **2021**, *48*, e2020GL092170. [CrossRef]
16. Gómez-Chova, L.; Alonso, L.; Guanter, L.; Camps-Valls, G.; Calpe, J.; Moreno, J. Correction of Systematic Spatial Noise in Push-Broom Hyperspectral Sensors: Application to CHRIS/PROBA Images. *Appl. Opt.* **2008**, *47*, F46–F60. [CrossRef]
17. Rasti, B.; Scheunders, P.; Ghamisi, P.; Licciardi, G.; Chanussot, J. Noise Reduction in Hyperspectral Imagery: Overview and Application. *Remote Sens.* **2018**, *10*, 482. [CrossRef]
18. Qin, J.; Zhao, H.; Liu, B. Self-Supervised Denoising for Real Satellite Hyperspectral Imagery. *Remote Sens.* **2022**, *14*, 3083. [CrossRef]
19. He, J.; Zhang, S.; Cui, X.; Feng, W. Remote Sensing for Shallow Bathymetry: A Systematic Review. *Earth-Sci. Rev.* **2024**, *258*, 104957. [CrossRef]
20. Wu, Z.; Zhao, Y.; Wu, S.; Chen, H.; Song, C.; Mao, Z.; Shen, W. Satellite-Derived Bathymetry Using a Fast Feature Cascade Learning Model in Turbid Coastal Waters. *J. Remote Sens.* **2024**, *4*, 0272. [CrossRef]
21. Ye, L.; Chen, H.; Xiao, J.; Lu, C.; Zhang, J.; Chu, S.; Cheng, L. PHY-SDB: A Dual-Physics-Guided Deep Learning Framework for Hyperspectral Satellite-Derived Bathymetry. *GIScience Remote Sens.* **2025**, *62*, 2594809. [CrossRef]
22. Le, Y.; Sun, X.; Chen, Y.; Zhang, D.; Wu, L.; Liu, H.; Hu, M. High-Accuracy Shallow-Water Bathymetric Method Including Reliability Evaluation Based on Sentinel-2 Time-Series Images and ICESat-2 Data. *Front. Mar. Sci.* **2024**, *11*, 1470859. [CrossRef]
23. Ye, M.; Yang, C.; Zhang, X.; Li, S.; Peng, X.; Li, Y.; Chen, T. Shallow Water Bathymetry Inversion Based on Machine Learning Using ICESat-2 and Sentinel-2 Data. *Remote Sens.* **2024**, *16*, 4603. [CrossRef]
24. Jung, J.; Parrish, C.E.; Magruder, L.A.; Herrmann, J.; Yoo, S.; Perry, J.S. ICESat-2 Bathymetry Algorithms: A Review of the Current State-of-the-Art and Future Outlook. *ISPRS J. Photogramm. Remote Sens.* **2025**, *223*, 413–439. [CrossRef]
25. Shah, C.; Du, Q.; Xu, Y. Enhanced TabNet: Attentive Interpretable Tabular Learning for Hyperspectral Image Classification. *Remote Sens.* **2022**, *14*, 716.
26. Pan, E.; Ma, Y.; Mei, X.; Huang, J.; Chen, Q.; Ma, J. Hyperspectral Image Destriping and Denoising from a Task Decomposition View. *Pattern Recognit.* **2023**, *144*, 109832. [CrossRef]
27. Zhu, L.; Wu, J.; Biao, W.; Liao, Y.; Gu, D. Spectralmae: Spectral Masked Autoencoder for Hyperspectral Remote Sensing Image Reconstruction. *Sensors* **2023**, *23*, 3728. [CrossRef] [PubMed]
28. Xu, H.; Liu, X.; An, Z.; Hou, Z.; Dong, J.; Liu, B. Spatial Pattern of Modern Sedimentation Rate of Qinghai Lake and a Preliminary Estimate of the Sediment Flux. *Chin. Sci. Bull.* **2010**, *55*, 621–627. [CrossRef]
29. Sha, Z.; Wang, Q.; Wang, J.; Du, J.; Hu, J.; Ma, Y.; Kong, F.; Wang, Z. Regional Environmental Change and Human Activity over the Past Hundred Years Recorded in the Sedimentary Record of Lake Qinghai, China. *Environ. Sci. Pollut. Res.* **2017**, *24*, 9662–9674. [CrossRef]
30. Wang, J.; Zhu, L.; Daut, G.; Ju, J.; Lin, X.; Wang, Y.; Zhen, X. Investigation of Bathymetry and Water Quality of Lake Nam Co, the Largest Lake on the Central Tibetan Plateau, China. *Limnology* **2009**, *10*, 149–158. [CrossRef]
31. Song, C.; Huang, B.; Ke, L. Modeling and Analysis of Lake Water Storage Changes on the Tibetan Plateau Using Multi-Mission Satellite Data. *Remote Sens. Environ.* **2013**, *135*, 25–35. [CrossRef]
32. Qiao, B.; Zhu, L.; Yang, R. Temporal-Spatial Differences in Lake Water Storage Changes and Their Links to Climate Change throughout the Tibetan Plateau. *Remote Sens. Environ.* **2019**, *222*, 232–243. [CrossRef]
33. Zhang, G.; Yao, T.; Shum, C.K.; Yi, S.; Yang, K.; Xie, H.; Feng, W.; Bolch, T.; Wang, L.; Behrangi, A.; et al. Lake Volume and Groundwater Storage Variations in Tibetan Plateau's Endorheic Basin. *Geophys. Res. Lett.* **2017**, *44*, 5550–5560. [CrossRef]

34. Crétaux, J.-F.; Abarca-del-Río, R.; Bergé-Nguyen, M.; Arsen, A.; Drolon, V.; Clos, G.; Maisongrande, P. Lake Volume Monitoring from Space. *Surv. Geophys.* **2016**, *37*, 269–305. [CrossRef]
35. Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [CrossRef]
36. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
37. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely Randomized Trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
38. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
39. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
40. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning Representations by Back-Propagating Errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
41. Wei, K.; Fu, Y.; Huang, H. 3-D Quasi-Recurrent Neural Network for Hyperspectral Image Denoising. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 363–375.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

remote sensing

MDPI

Article

Improving Satellite-Based Estimation and Mapping of Soil Lead by Using an Enhanced Spectral Feature Set and XGBoost Model

Xibo Xu [1,2], Ying Wang [3], Xinrui Dai [4], Qi Shen [4], Quanyuan Wu [1], Zeqiang Wang [2,4] and Jianfei Cao [1,5,*]

[1] College of Geography and Environment, Shandong Normal University, Jinan 250358, China; xuxibo@stu.sdnu.edu.cn (X.X.); wqy6420582@163.com (Q.W.)
[2] Key Laboratory of Environmental Change and Natural Disaster of Ministry of Education, Beijing Normal University, Beijing 100875, China; 202531099010@mail.bnu.edu.cn
[3] Ningxia Institute of Basic Geological Survey (Ningxia Central Laboratory of Geology and Mineral Resources), Yinchuan 750000, China; wangying_win@163.com
[4] College of Geographical Sciences, Harbin Normal University, Harbin 150025, China; daixinrui@stu.hrbnu.edu.cn (X.D.)
[5] Satellite Application Center for Ecology and Environment, Ministry of Ecology and Environment, Beijing 100094, China
* Correspondence: caojianfei@sdnu.edu.cn; Tel.: +86-15628842807

Highlights

What are the main findings?

- An enhanced spectral feature set (ESFS) was constructed, and two new spectral indices for Pb-contaminated soils, named SPPI-2 and SPPI-3, were developed and incorporated into the ESFS.
- The XGBoost-based satellite hyperspectral mapping model achieved satisfactory performance (RPD = 2.06).

What are the implications of the main findings?

- This study presents new spectral indices for effectively characterizing Pb-contaminated soils and supporting regional-scale soil management.
- This study provides a solution framework for estimating and mapping soil heavy metals, which assists in contamination hotspot identification and promotes environmental sustainability.

Abstract

Satellite hyperspectral remote sensing offers an efficient and cost-effective approach for estimating and mapping soil lead (Pb), thereby supporting pollution identification and environmental sustainability. However, the development of satellite-based spectral estimation models depends on the availability of a robust spectral feature set for soil Pb as input, which is difficult to obtain under field conditions due to interference from moisture, particle size, and light scattering. To address this issue, controlled spectral experiments were conducted on laboratory-prepared soil samples with varying Pb contamination levels. The spectral characteristics associated with Pb contamination were analyzed, and an enhanced spectral feature set (ESFS) was constructed using the successive projections algorithm–Shapley additive explanations method. Two new spectral indices for Pb-contaminated soils, named SPPI-2 and SPPI-3, were developed and incorporated into the ESFS. The ESFS was then applied to satellite hyperspectral data calibrated via direct standardization, with its spectral parameters used as input variables and measured Pb concentrations from field soil samples as the dependent variable. Finally, a satellite-based spectral model for soil Pb estimation was developed using the XGBoost (eXtreme Gradient Boosting) algorithm.

Results showed that the spectral parameters in the ESFS included four characteristic bands (R_{840}, R_{1013}, R_{1215}, and R_{2239}) and two newly developed spectral indices (SPPI-2 and SPPI-3). The satellite-based spectral estimation model based on the ESFS and XGBoost algorithm achieved the best performance, with R^2 (coefficient of determination) and RPD (ratio of performance to deviation) values of 0.78 and 2.10, respectively, representing a maximum improvement of 164.10% and a minimum improvement of 12.86% (in terms of RPD values) compared to common methods. Hotspot areas of Pb-contaminated soils were mainly found in the eastern part of the abandoned coal mining area, which is associated with improper coal mining and transportation activities. This study presents a satellite hyperspectral framework for effectively estimating the distribution pattern of soil Pb and supporting the regional-scale soil management and environmental sustainability.

Keywords: soil lead; satellite-based estimation; enhanced spectral indices set; XGBoost model; coal mining areas

1. Introduction

Soil lead (Pb) contamination represents one of the most prevalent forms of soil pollution and a driver of ecological degradation worldwide [1,2]. Such contamination is commonly observed in regions associated with mining activities and intensive industrialization [3,4]. Although Pb occurs naturally in soil parent material as an inherent constituent, anthropogenic activities, including mining, smelting, and transportation, can generate elevated Pb concentrations. These Pb inputs are transferred into the soil through direct discharge and atmospheric deposition, thereby leading to Pb contamination of the soil [5,6]. People living in these areas are exposed to soil Pb, which can enter the body through ingestion, inhalation, and skin contact. Long-term exposure can harm neurocognitive development, especially in children, and can also affect red blood cell production, posing serious health risks [7,8]. Therefore, effective monitoring of the distribution patterns and contamination hotspots of soil Pb is crucial for protecting human health and promoting environmental sustainability.

The satellite hyperspectral technique is effective for estimating the distribution patterns and detecting contamination hotspots of soil Pb in coal mining areas [9–11]. Traditionally, the estimation and mapping of soil Pb in coal mining areas relied on an integrated approach involving field soil sampling, laboratory analysis, and geostatistical interpolation [12,13]. Although the method is simple, the labor and time costs are high, and soil Pb monitoring at a regional scale is expensive. Hyperspectral sensors equipped on satellite platforms can collect spectral data over large areas in a short time. By combining machine learning methods, Pb information in the soil can be extracted, enabling large-scale monitoring of soil Pb status and contamination hotspots [9].

A robust set of spectral features used as input variables is essential for developing reliable satellite-based spectral estimation models for soil Pb [14,15]. Usually, spectral features such as characteristic bands are selected from satellite hyperspectral data and then used to explain soil Pb status in spectral models. Satellite hyperspectral data, with its high spectral resolution, provides soil spectral information across many bands, which helps identify absorption features related to soil Pb [16,17]. However, hyperspectral data also contain noise and irrelevant spectral information from environmental factors and other soil properties, not just the spectra of the target soil components [18,19]. Therefore, accurately extracting the spectral features of soil Pb and building an effective spectral feature set are

https://doi.org/10.3390/rs18091446

key challenges for developing reliable satellite-based hyperspectral estimation models for soil Pb.

To address this issue, supervised searching methods (such as stepwise regression, correlation coefficient analysis, successive projections algorithm, and competitive adaptive reweighted sampling) were used to select spectral features for soil Pb from high-dimensional hyperspectral data. These selected features were then used as input variables to improve the computational efficiency and accuracy of satellite-based spectral estimation models [20–24]. Furthermore, Pb exists in soil mainly as compounds and is closely associated with adsorptive soil materials, such as organic matter, clay minerals, and iron oxides. Therefore, the spectral features extracted by supervised searching methods indirectly reflect soil Pb, as they are based on the spectral signals of Pb-associated soil components [22,24,25].

Although significant progress has been made in constructing the spectral feature set for soil Pb, two key issues still remain. On one hand, the spectral features of soil Pb and related soil components are often affected or masked by environmental factors such as moisture, particle size, and scattered light. This problem is especially pronounced in satellite hyperspectral data from the complex soil environments found in coal mining areas. As a result, accurately extracting the spectral features of soil Pb and building an efficient spectral feature set remain challenging. On the other hand, spectral indices have clear advantages for characterizing target soil components and improving the accuracy of satellite-based spectral estimation models. They should be an important part of the spectral feature set for soil Pb. However, the current spectral feature sets show clear limitations in this aspect [22,26].

Therefore, in this study, a controlled spectral experiment of laboratory-prepared Pb-contaminated soil was conducted to analyze the response mechanisms between Pb contamination and soil spectra. We then constructed an enhanced spectral feature set (ESFS) by combining characteristic bands related to soil Pb with two newly developed spectral indices for Pb-contaminated soils (SPPI). This feature set can more effectively indicate soil Pb levels. In addition, an advanced machine learning algorithm, eXtreme Gradient Boosting (XGBoost), was used to model the nonlinear relationship between soil Pb and the spectral parameters in the ESFS, thereby building a robust satellite-based spectral estimation model for soil Pb [27].

Spectral indices are powerful tools for indicating soil components, as they enhance spectral information by strategically combining multiple spectral bands and revealing the relationships between them [28,29]. These indices can also serve as covariates during model calibration, thereby improving model accuracy and reliability [30]. For example, Chen et al. [22], Zhang et al. [31], and Santos et al. [32] suggested that the performance of soil spectral estimation models can be improved by including spectral indices as input variables, as these indices are less sensitive to environmental noise. Thus, in this study, an SPPI was developed based on a controlled spectral experiment of laboratory-prepared Pb-contaminated soils and subsequently integrated as a critical component of the ESFS, supporting the development of satellite-based spectral estimation models for soil Pb in coal mining areas.

XGBoost is a machine learning algorithm that can efficiently determine the relationship between spectral parameters within the ESFS and soil Pb concentrations, thus supporting the construction of an accurate satellite-based spectral estimation model [27,33]. This regularized gradient boosting method combines multiple weak predictors iteratively to achieve high accuracy. Its loss function includes regularization terms, which help reduce overfitting and improve model performance [34]. Previous studies have demonstrated its effectiveness in developing spectral estimation models for monitoring soil components such as organic matter, salt, arsenic, pH, and total nitrogen [35–38]. Overall, combining

ESFS with the XGBoost algorithm has great potential to improve the satellite-based spectral estimation of soil Pb.

The study's objectives were to: (1) develop new spectral indices for Pb-contaminated soils (i.e., SPPI); (2) create an ESFS for characterizing soil Pb status; and (3) build a satellite-based spectral estimation model of soil Pb in coal mining bare-soil areas.

2. Materials and Methods

2.1. Flowchart for Building the Satellite-Based Spectral Estimation Model of Soil Pb

The process of building the satellite-based spectral estimation model of soil Pb is illustrated in Figure 1. First, soil samples with varying Pb contamination levels were prepared in the laboratory, and their spectral data were collected. Next, we identified the characteristic bands of soil Pb and developed an SPPI. These spectral features were combined to create an ESFS for characterizing soil Pb status. Then, an ESFS was constructed from the transformed ZY1-02D satellite hyperspectral data using the direct standardization (DS) algorithm. Subsequently, we used the spectral parameters in the ESFS as independent variables and the measured Pb values of field soil samples as the dependent variable. Finally, a satellite-based spectral estimation model for soil Pb in coal mining bare-soil areas was developed using the XGBoost algorithm.

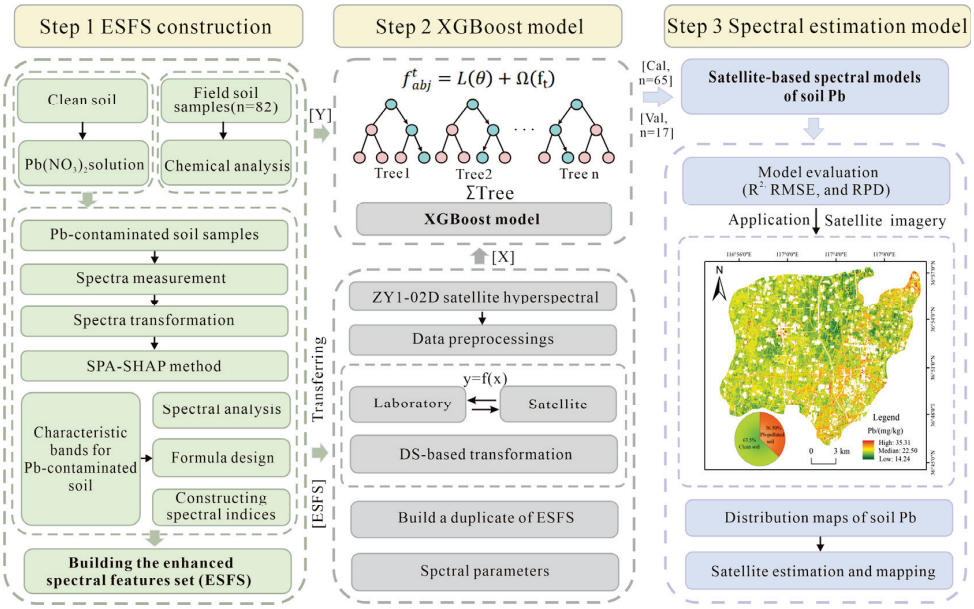


Figure 1. Flowchart for building the satellite-based spectral estimation model of soil Pb in coal mining bare-soil areas.

2.2. Study Area and Field Soil Sampling

(1) Study area

An abandoned coal mining area in eastern China was selected as the study area (Figure 2). The study area has a maximum length of 24.2 km and a maximum width of 20.1 km, covering a total area of approximately 335.2 km^2, with geographical coordinates ranging from 116°54′26″E to 117°11′2″E and 36°46′3″N to 36°57′21″N. The study area is characterized by the warm temperate semi-humid monsoon climate, with mean annual temperature of 12.8 °C and precipitation of 583.3 mm. The main soil type is Fluvo-aquic soil. Intensive coal mining began in 2007, producing 750,000 metric tons annually until operations ceased completely in 2019 [39].

https://doi.org/10.3390/rs18091446

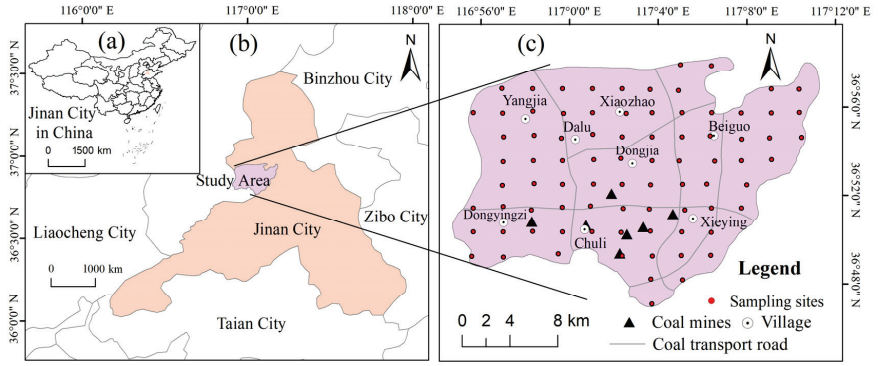


Figure 2. Geographic location of the coal mining area and field soil samplings: (**a**) Location of Jinan City in China; (**b**) the study area within Jinan City; (**c**) locations of soil sampling sites and coal mines.

Improper coal mining and coal transport practices have resulted in the dispersion of coal slag and fly ash on soil surfaces [40]. Heavy metal contaminants have entered the soil through hydrologic washout and aeolian transport, leading to the accumulation of soil Pb [41].

(2) Field soil sampling and chemical analysis

Considering factors such as land use, mine distribution, and road conditions, a total of 82 sampling locations were predetermined based on digital maps generated in ArcGIS 10.2 software. Field soil samples were collected during 25–26 February 2023. A handheld GPS was used to navigate to each predetermined sampling location. The distance between any two soil sampling sites is greater than 1.5 km. A multi-point sampling method was employed within a 30-m radius to collect approximately 0.5 kg of soil at a depth of 0–20 cm. The collected soil samples were then sealed in bags and transported to the laboratory for further analysis. Finally, the actual geographic coordinates of each sampling location were recorded again using the handheld GPS device, and environmental conditions, such as temperature and land use, were documented with a smart camera.

In the laboratory, all soil samples were air-dried under natural conditions. We removed small stones, roots, and other impurities, then ground the soil and sieved it through a 0.2 mm mesh. Next, we weighed 0.2 g of soil and transferred the sample to a crucible. After moistening the sample with water, we sequentially added 4 mL HNO_3, 1 mL $HClO_4$, and 5 mL HF. The mixture was then digested on a sand bath. After cooling, 5 mL of supernatant was collected into a test tube and diluted with a 1:1 HNO_3 solution. Finally, soil Pb concentration was measured using inductively coupled plasma mass spectrometry (ICP-MS) [42].

2.3. Hyperspectral Imagery Acquisition and Preprocessing

2.3.1. Satellite Hyperspectral Imagery Acquisition

ZY1-02D, launched in 2019, is a hyperspectral satellite designed for environmental monitoring and disaster mitigation [43]. The satellite is equipped with two sensors: a panchromatic/multispectral imager and a hyperspectral sensor [44]. The hyperspectral sensor covers the 400–2500 nm range with 166 spectral bands, including 76 bands with the spectral resolution of 10 nm in the 400–1000 nm range and 90 bands with a spectral resolution of 20 nm in the 1000–2500 nm range [45]. The spatial resolution of the hyperspectral imagery is 30 m. ZY1-02D satellite hyperspectral imagery was downloaded from the China Center for Resources Satellite Data and Application (https://data.cresda.cn). The acquisition date of the satellite hyperspectral imagery was 1 March 2023, which is close to the time when the field soil samples were collected.

https://doi.org/10.3390/rs18091446

2.3.2. Extracting the Bare Soil Areas from the Hyperspectral Imagery

The preprocessing for ZY1-02D hyperspectral imagery included three steps [44,46]: First, radiometric calibration and atmospheric correction were conducted using standard modules in ENVI 5.6 software. Next, geometric correction and orthorectification were performed using Landsat 8 reference images and 250 m DEM data. Finally, noisy bands in the spectral ranges of 400–422 nm and 2451–2500 nm were removed, leaving 148 spectral bands for spectral estimation modeling.

Additionally, the non-bare soil regions in the ZY1-02D hyperspectral images were excluded [47,48]. Based on the results of preliminary experiments and previous studies [15,28], regions with NDVI values between 0 and 0.5 were classified as bare soil. A binary image approach was used to mask the non-bare soil regions and extract the bare soil areas. The accuracy of the extracted bare soil areas was assessed using Google Maps (Version 10.15) through random sampling methods, achieving an overall accuracy of 95%. The masking of bare soil areas in satellite hyperspectral imagery is reliable.

2.4. Building the Enhanced Spectral Features Set (ESFS)

2.4.1. Laboratory Preparation of Pb-Contaminated Soil Samples

(1) Preparation of the Pb-contaminated soil samples

A total of 80 soil samples with different Pb contamination levels were prepared in the laboratory. The soil materials associated with the Pb such as soil organic matter, clay, and iron oxides, were set as control variables. This controlled laboratory experiment strongly supports the analysis of the spectral mechanisms of Pb-contaminated soils and the development of spectral feature sets [47,49].

The procedures for preparing Pb-contaminated soil samples are as follows (as shown in Figure 3): First, 10 kg of clean soil was collected from the coal mining area and air-dried under natural conditions (approximately 25 °C). Stones, twigs, and roots were removed, and the soil was sieved through a 1 mm mesh. Next, 50 g of soil was weighed into beakers (labeled D01–D80), and $Pb(NO_3)_2$ solutions (50 μg/mL or 1 mg/mL) were added. The volume of the $Pb(NO_3)_2$ solution used is provided in Table S1. Ultrapure water was added to cover the soil, and the mixture was thoroughly stirred with glass rods. The prepared Pb-contaminated soil samples were air-dried in a fume hood, ground again, and sieved through a 0.2 mm mesh. Each sample was divided into two parts: one for hyperspectral measurement and the other for soil Pb determination. All steps were conducted in accordance with strict safety protocols, including proper ventilation, the use of personal protective equipment (respirators, gloves, and goggles), and appropriate waste disposal [50].

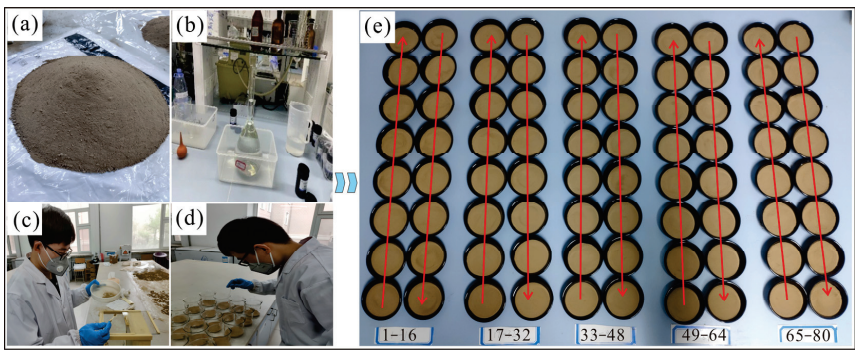


Figure 3. Illustrations of the laboratory preparation of Pb-contaminated soil samples: (**a**) clean soils collected from the study area; (**b**) $Pb(NO_3)_2$ solutions with a mass fraction of 50 μg/mL or 1 mg/mL; (**c**,**d**) steps in the laboratory preparation of Pb-contaminated soil samples; (**e**) laboratory-prepared soil samples with different levels of Pb contamination.

(2) Statistical characteristics of soil Pb concentrations in laboratory-prepared soil samples

The statistical characteristics of Pb concentrations in the laboratory-prepared soil samples are shown in Table 1. Pb concentrations ranged from 16.20 to 474.98 mg/kg, with both the mean and median values being 105.55 mg/kg, which indicates that the dataset follows a normal distribution. The coefficient of variation for the laboratory-prepared soil sample dataset is 1.14, indicating high variability [20]. The kurtosis and skewness values were 2.08 and 1.81, respectively, indicating a right-skewed distribution. The background value of soil Pb in the coal mining area is 21.40 mg/kg, which is significantly lower than the range of Pb concentrations observed in the laboratory-prepared soil samples [51].

Table 1. Descriptive statistics of Pb concentrations in laboratory-prepared soil samples/(mg/kg).

Items	Mean	Minimum	Maximum	Median	Standard Deviation	Coefficient of Variation	Kurtosis	Skewness	Background Value
Pb	105.55	16.20	474.98	105.55	120.31	1.14	2.08	1.81	21.40

Note: The background values of Pb concentrations in the soil were referred to by Lv and Yu [51].

2.4.2. Spectral Measurement

(1) Imaging spectral measurement

The HySpex hyperspectral imaging system (Norsk Elektro Optikk, NEO, Oslo, Norway) was used to measure the spectral reflectance of the laboratory-prepared Pb-contaminated soil samples [52]. The HySpex system provided spectral measurements in the range of 400–2500 nm across 472 spectral bands. The spectral range of 400–1000 nm contains 186 bands with a resolution of 3 nm, while the spectral range of 1000–2500 nm includes 286 bands with a resolution of 5 nm.

Interfering factors such as scattered light, moisture, and particle size were excluded during the laboratory spectral measurements of soil samples. 15 g of Pb-contaminated soil was evenly spread onto black trays, which were then placed on the stage. In addition, a white reference panel was placed on the stage. Twelve 50 W halogen lamps were arranged strategically in the darkroom as the light source. Finally, the HySpex system was activated, and each soil sample was measured three times.

(2) Spectral pretreatment

Spectral pretreatment was conducted to obtain representative spectra of the laboratory-prepared Pb-contaminated soil samples [53]. First, the raw DN values of the imaging spectral data were converted into spectral reflectance. Next, noisy data, such as that at the image edges, was removed. The spectral range of 400–447 nm was excluded due to its low signal-to-noise ratio. In addition, the spectral regions affected by water absorption (1300–1500 nm and 1800–2000 nm) were eliminated. After preprocessing, a total of 374 spectral bands were retained. Finally, 20 random pixels were selected, and their average values were calculated to represent the spectra of each soil sample.

$$R_s = (DN_s / DN_w) \times R_w \quad (1)$$

where R_s represents the spectral reflectance of the soil samples, DN_s and DN_w represent the digital values of the soil samples and the white reference panel, respectively. R_w represents the calibration coefficient of the spectral reflectance provided by NEO Company.

2.4.3. Spectral Transformation

Spectral transformations were performed to enhance the spectral features of Pb-contaminated soil and reduce random noise [54]. Eleven methods were applied to transform

the soil spectral data: mean centering (CT), first derivative (D1), second derivative (D2), detrending (DT), max-min normalization (MMS), Savitzky–Golay smoothing (SG) (The window size and polynomial order were defined as 5 and 2, respectively), standard normal variate (SNV), square root transformation (SQ), standardization (SS), reciprocal transformation (RE), and moving average smoothing (MA). Detailed descriptions of these methods are available in references [55–58].

2.4.4. Characteristic Bands Selection

Identifying the characteristic bands of soil Pb is crucial for constructing SPPI and for the development of ESFS. First, the optimal spectral transformation method was determined based on the lowest $RMSE_{CV}$ (Root mean square error after five-fold cross-validation), using the successive projections algorithm (SPA). SPA is a forward cycling feature selection algorithm where feature subsets with minimal collinearity are selected using projection methods in vector space [59]. Based on the 11 transformed spectral data, vectors with higher projections in the vector space become new starting vectors in each iteration of the SPA method. Through regression calculations, the $RMSE_{CV}$ can be obtained. The spectral data with the minimum $RMSE_{CV}$ value is regarded as the optimal spectral transformation method. Next, the importance of spectral bands was analyzed using the Shapley Additive Explanations (SHAP) algorithm, and the characteristic bands of soil Pb were identified. SHAP is a method used to quantitatively assess the explanatory contribution of spectral bands to target soil properties. The contribution of each spectral band to the estimation model is computed to evaluate the importance of individual bands and to identify the characteristic bands for soil Pb [60].

The characteristic bands of soil Pb were selected and are shown in Figure 4. The $RMSE_{CV}$ values of the transformed spectral data followed the following increasing order: MMS (lowest), SQ, SS, SNV, MA, CT, D1, DT, SG, RE, and D2 (highest). Among these methods, MMS was identified as the most effective approach for spectral transformation. Four characteristic bands were identified: R_{840}, R_{1013}, R_{1215}, and R_{2239}.

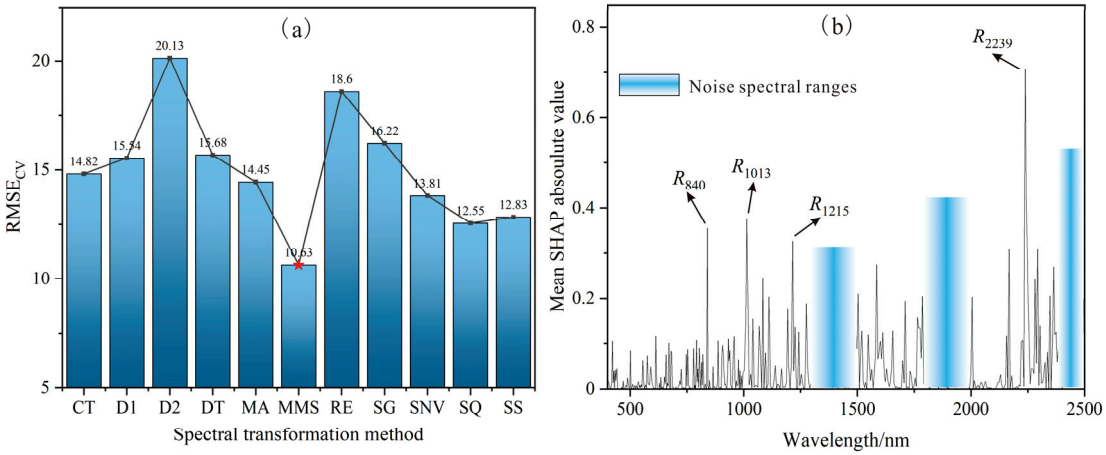


Figure 4. $RMSE_{CV}$ values of various transformed spectra (**a**) and SHAP values of spectral bands in MMS-transformed data (**b**). $RMSE_{CV}$ stands for root mean square error, SHAP stands for Shapley additive explanations, and MMS denotes the max-min normalization transformation method.

2.4.5. SPPI Construction

(1) Key parameters and formulas

Three characteristic bands were defined as spectral points, namely point A (λ_{840}, R_{840}), point B (λ_{1013}, R_{1013}), and point C (λ_{1215}, R_{1215}), which are located in the increasing part of the spectral curve (as indicated in Figure 5). In contrast, spectral point D (λ_{2239}, R_{2239})

occurs at a distinct reflectance peak. Together, these four spectral points characterize the reflectance and absorption features of the soil MMS-transformed spectral curve.

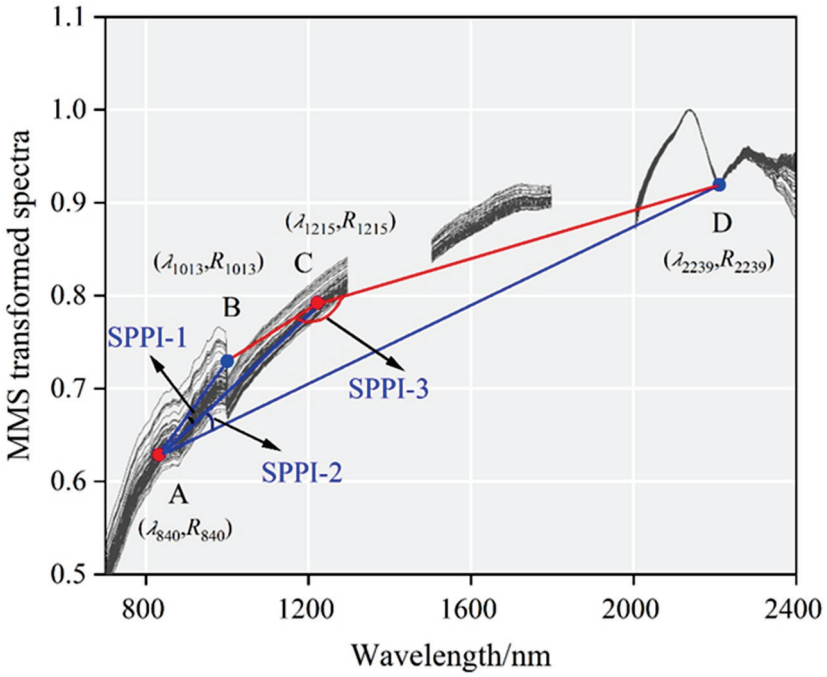


Figure 5. Illustrations of the new spectral indices for Pb-contaminated soil (SPPI-1, SPPI-2, and SPPI-3).

The changes in these four spectral points reflect the shapes and variations in the spectral curves of Pb-contaminated soils, and also indicate the contamination levels of the soil samples. Therefore, points A, B, C, and D form the foundation for developing the SPPI. Additionally, the triangular spectral index effectively captures the morphology and variations in spectral curves by combining multiple spectral bands [61]. Here, we developed the SPPI as follows:

$$\text{SPPI-1} = \frac{(R_{840} - R_{1013})(\lambda_{840} - \lambda_{1215}) - (R_{840} - R_{1215})(\lambda_{840} - \lambda_{1013})}{(\lambda_{840} - \lambda_{1013})(\lambda_{840} - \lambda_{1215}) - (R_{840} - R_{1013})(R_{840} - R_{1215})} \quad (2)$$

$$\text{SPPI-2} = \frac{(R_{840} - R_{1013})(\lambda_{840} - \lambda_{2239}) - (R_{840} - R_{2239})(\lambda_{840} - \lambda_{1013})}{(\lambda_{840} - \lambda_{1013})(\lambda_{840} - \lambda_{2239}) - (R_{840} - R_{1013})(R_{840} - R_{2239})} \quad (3)$$

$$\text{SPPI-3} = \frac{(R_{1013} - R_{1215})(\lambda_{1013} - \lambda_{2239}) - (R_{1013} - R_{2239})(\lambda_{1013} - \lambda_{1215})}{(\lambda_{1013} - \lambda_{1215})(\lambda_{1013} - \lambda_{2239}) - (R_{1013} - R_{1215})(R_{1013} - R_{2239})} \quad (4)$$

where λ represents the wavelengths of the spectral bands, and R indicates the soil reflectance values derived from the MMS-transformed spectral data.

(2) Optimal SPPIs

The Pearson correlation coefficients (CC) of the spectral parameters (including SPPI-1, SPPI-2, SPPI-3, and all single bands) are shown in Figure 6. Within the spectral range of 560–1020 nm, the CC values of the spectral bands increase progressively with increasing wavelength, reaching a maximum CC value of 0.46 at R_{1020}. The CC values of the characteristic bands R_{840} and R_{1013} are 0.41 and 0.44, respectively. In addition, the CC value of spectral band R_{2100} (0.30) was found to be the lowest within the spectral range of 2000–2450 nm, while the CC values of the neighboring spectral bands are relatively high. The CC value of the characteristic band R_{2239} is 0.42. In contrast, the CC values for SPPI-1, SPPI-2, and SPPI-3 are 0.61, 0.94, and 0.98, respectively, indicating a strong correlation with

https://doi.org/10.3390/rs18091446

soil Pb concentrations. Consequently, SPPI-2 and SPPI-3 were identified as the optimal spectral indices for Pb-contaminated soil and were selected to constitute the ESFS.

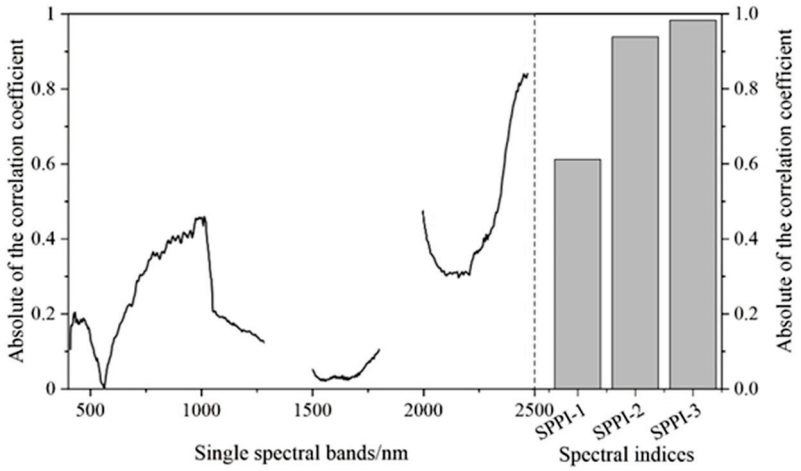


Figure 6. Pearson correlation coefficients (CC) of spectral bands (**left**) and spectral indices for Pb-contaminated soil (SPPI-1, SPPI-2, and SPPI-3) (**right**).

2.5. Transferring the ESFS from Laboratory to Satellite

Narrowing the differences between laboratory and satellite spectra is essential for transferring the ESFS from laboratory to satellite data. First, band removal and resampling methods were integrated to achieve consistency in the number of spectral bands (n = 148) and wavelength positions between the satellite and laboratory data. Noted that the characteristic bands need to be retained during the dimensionality reduction of the laboratory spectral data. Next, the direct standardization (DS) algorithm was applied to establish the relationship between laboratory and satellite spectra, and to minimize their differences through spectral transformation [47,62,63]. DS was originally developed for spectral model transfer between instruments with varying operating conditions [64]. Ji et al. [65] further expanded the application of the DS method to link laboratory and field soil spectra, effectively removing the influence of moisture from field soil spectra. In this study, the DS-based transformation method was employed to construct the ESFS using ZY1-02D hyperspectral data. The equation for the DS algorithm is given below:

$$X_{\text{lab}} = X_{\text{sat}} \times B + E \tag{5}$$

where X_{sat} is ZY1-02D satellite hyperspectral data, X_{lab} represents the laboratory hyperspectral data, and B and E are the transformation matrix and residual matrix, respectively. Detailed information can be found in references [47,62,64,66]. The DS method was performed using Python 3.8 software.

2.6. XGBoost Algorithm

XGBoost is an advanced machine learning algorithm that combines gradient boosting with regularization to reduce overfitting while improving the generalization ability of models [27,33]. The XGBoost model works by iteratively building multiple decision trees (serving as weak predictors), each trained to correct the errors of the previous ones. During the training process, XGBoost adds regularization terms to the loss function to control model complexity and enhance robustness [34]. The target model consists of k decision trees, each with optimized weights at its leaf nodes, and makes predictions by summing all

https://doi.org/10.3390/rs18091446

trees' weighted outputs to create a complete gradient-boosted decision tree [35,37,38]. The XGBoost model can be expressed as follows:

$$f_{\text{abj}}^{t} = \mathcal{L}(\theta) + \Omega(f_t) \tag{6}$$

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{7}$$

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{8}$$

where $L(\theta)$ represents the loss function, $\Omega(f_t)$ denotes the regularization term, y_i and $\hat{y}_i$ represent the measured and predicted values of the ith sample, respectively, γ is the penalty cost for new leaf nodes, T is the total number of leaf nodes, λ is the regularization parameter, and ω_j is the weight of the jth leaf node.

The hyperparameter configuration was optimized to obtain a robust XGBoost model. The number of estimators was chosen from a range of 50 to 300. The maximum tree depth values ranged from 3 to 10. The value ranges for learning rate, subsample ratio, column sample by tree ratio, and gamma values were (0.01 to 0.2), (0.6 to 0.8), (0.6 to 0.8), and (0 to 0.6) respectively. The optimal configuration of hyperparameters was obtained using cross-validation and grid search methods, indicating that the model achieved an optimal balance between accuracy, computational cost, and time efficiency [67].

Additionally, an early stopping method was implemented to prevent overfitting in the XGBoost model. If the validation error did not decrease over 10 epochs, training would be terminated early, thereby avoiding model overfitting. The early stopping strategy monitors model performance during the training process and automatically stops training to prevent the model from overfitting on the training set, thus improving the model's generalization ability on unknown sample data and ensuring model accuracy and robustness [68].

2.7. Model Calibration and Evaluation

A total of 82 field soil samples were randomly divided into a calibration set (n = 65) and a validation set (n = 17). The satellite-based spectral estimation model of soil Pb was calibrated using the calibration set [20]. A 10-fold cross-validation method was used to improve the model's robustness. One-tenth of the stratified samples were randomly selected to validate the model accuracy. The model calculation ended only after all samples had been selected at least once for model calibration and validation. The class proportions of samples were considered in each fold, with representative samples provided in each fold, thereby reducing model bias [67]. The model with the minimum error was considered the optimal model. In addition, the validation set was used only for model evaluation and was kept completely separate from the model calibration [69].

The coefficient of determination (R^2), root mean square error (RMSE), and ratio of performance to deviation (RPD) are commonly used metrics to evaluate model performance. R^2 indicates how well the predicted values match the measured values, with values closer to 1 representing better accuracy. RMSE measures the magnitude of prediction errors, with smaller values indicating higher precision. RPD is calculated as the ratio of the standard deviation of the validation data to the RMSE. According to Xu et al. [70] and Wang et al. [71], model performance can be classified as poor (RPD < 1.4), moderate (1.4 $\leq$ RPD < 2), or excellent (RPD $\geq$ 2). The best satellite-based spectral models are characterized by high R^2 and RPD values, along with a low RMSE.

Spectral feature sets composed of different spectral parameters were constructed to evaluate the performance of satellite-based spectral estimation models under varying

https://doi.org/10.3390/rs18091446

inputs. Four characteristic bands (R_{840}, R_{1013}, R_{1215}, and R_{2239}) constitute a common spectral feature set (CSFS). The optimal SPPIs were also individually incorporated into the CSFS, generating multiple model inputs (I: R_{840}, R_{1013}, R_{1215}, R_{2239}, and SPPI-1; II: R_{840}, R_{1013}, R_{1215}, R_{2239}, and SPPI-2). The enhanced spectral feature set (ESFS) comprised four characteristic bands and two novel spectral indices (SPPI-1 and SPPI-2). These comparative experiments were conducted to examine the role of SPPIs and ESFS in improving the performance of the XGBoost model.

Additionally, several modeling methods were used to build a satellite-based spectral estimation model for soil Pb, including partial least squares regression (PLSR), artificial neural networks (ANN), support vector machines (SVM), random forests (RF), and extreme learning machine (ELM). Details about these models are available in references [72–75]. These models were compared with the XGBoost models proposed in this study to evaluate their accuracy and robustness. Furthermore, the tuning of hyperparameters and the integration of grid search method to achieve optimal model performance. The configuration of hyperparameter combinations and the search space are shown in Table 2.

Table 2. Configuration of hyperparameter combinations and search space of the compared models.

ID	Model	Key Hyperparameters	Values or Settings
1	ANN	Hidden layers Activation Early stopping Optimizer	2, 3, 4, 5, 6, 7, 8 Relu True (5 epochs) Adam
2	SVM	Kernel Cost Gamma	Poly, RBF, Linear (1, 10) 0.001, 0.01, 0.1, 1
3	RF	N estimators Max depth Min samples split Min samples leaf Max features	50, 100, 150, 200 3, 5, 7, 9 2, 3, 4, 5, 6 2, 3, 4, 5, 6 ‘sqrt’, ‘log2’, None
4	ELM	Hidden neurons Activation Regularization	20, 30, 50, 70, 90 RBF, Relu, Sigmoid 0.001, 0.01, 0.1, 1
5	PLSR	N components	2, 3, 4, 5

3. Results and Analysis

3.1. Descriptive Statistical Characteristics of Pb Concentrations in Field Soil Samples

The descriptive statistics of soil Pb concentrations in the total sample set, calibration set, and validation set are shown in Figure 7. The Pb concentrations in the total set range from 17.40 to 29.40 mg/kg, with an average of 22.58 mg/kg. Both the calibration and validation sets have similar ranges of soil Pb concentrations compared to the total sample set. In addition, the mean and median values in the total, calibration, and validation sets are in close agreement, indicating that the sample datasets are approximately normally distributed [76]. The descriptive statistics of the three sample sets demonstrate that the sample data are suitable for spectral model calibration and validation.

Additionally, the maximum Pb concentration in soils in the total sample set was 29.40 mg/kg, which is about 1.37 times higher than the background value [51] of 21.40 mg/kg. Furthermore, approximately 68.29% of the samples exceeded the background value, indicating moderate Pb contamination in soils in the coal mining area. Therefore,

monitoring the soil Pb status and contamination of hotspot areas is essential to promote environmental sustainability and protect human health in the coal mining area.

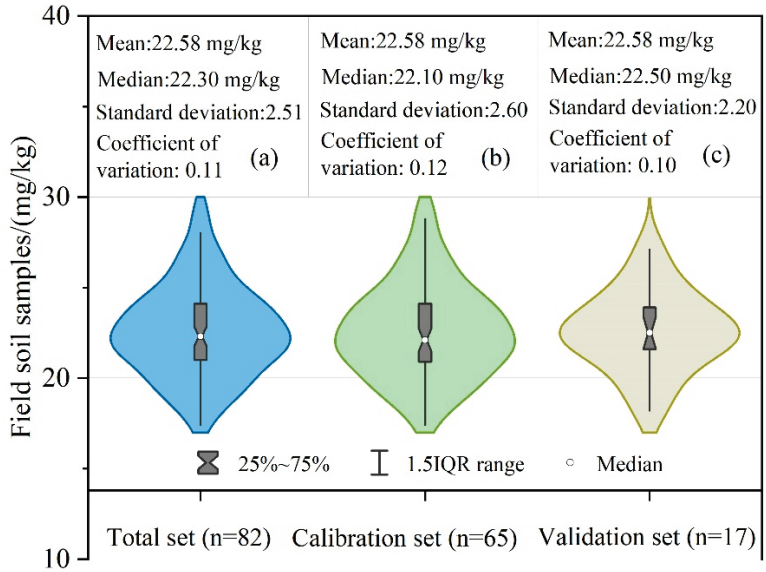


Figure 7. Descriptive statistical characteristics of Pb concentrations in the total sample set, calibration sample set, and validation set.

3.2. Spectral Characteristics of Laboratory, Satellite, and DS-Transformed Satellite Spectra

The laboratory spectra, raw satellite spectra, and DS-transformed satellite spectra are shown in Figure 8a. The reflectance values of the satellite spectra (around 0.2) are much lower than those of the laboratory spectra (around 0.4), indicating that the environmental effects on the satellite spectra are significant. After DS transformation, the satellite spectra show a significant increase in reflectance, closely matching the laboratory spectra. The environmental effect on the satellite spectra is reduced, and the difference between the laboratory and satellite spectra is narrowed, which supports the construction of an ESFS based on ZY1-02D satellite hyperspectral data.

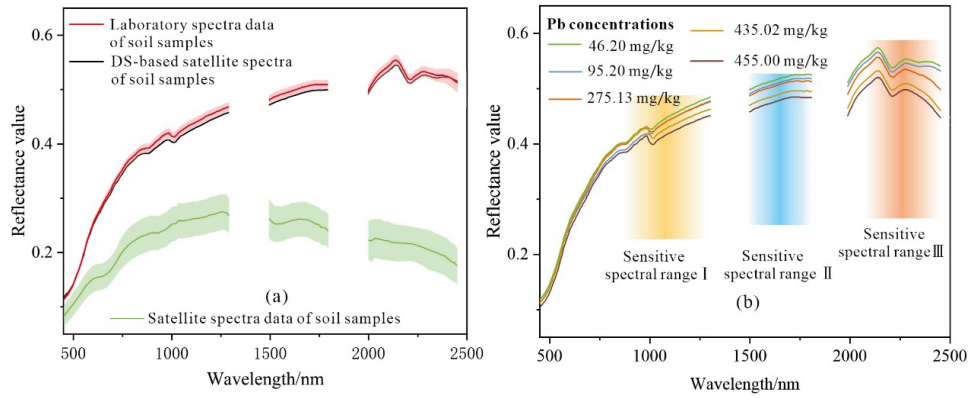


Figure 8. Soil spectral curves from laboratory, satellite, and DS-based transformed satellite data (**a**). Spectral curves of laboratory-prepared soil samples with different Pb concentrations (**b**).

Figure 8b shows the spectral curves of laboratory-prepared soil samples with different Pb concentrations. The reflectance values decrease as the Pb concentration increases, indicating a negative correlation. The soil sample with a Pb concentration of 46.20 mg/kg has significantly higher reflectance than samples with concentrations of 95.20 mg/kg, 275.13 mg/kg, 435.02 mg/kg, and 455.00 mg/kg. Additionally, the spectral curves of soil samples with different Pb levels show an overall increase in reflectance across the

https://doi.org/10.3390/rs18091446

400–2100 nm range. This increase is most pronounced between 400 and 800 nm, after which the rate of increase gradually slows. Two absorption valleys are removed near 1400 nm and 1900 nm, which may be due to soil moisture [77]. Reflectance peaks appear near 1000 nm and 2200 nm, possibly related to the presence of clay minerals, organic matter, and iron oxides in the soil [23,78,79].

Additionally, three sensitive spectral ranges for soil Pb are highlighted in Figure 8b, specifically at 800–1300 nm, 1500–1800 nm, and 2100–2450 nm. Within these ranges, there is a clear negative correlation between soil Pb and spectral reflectance. Overall, data analysis of laboratory spectra, satellite spectra, and DS-transformed satellite spectra supports the identification of characteristic bands and the development of the SPPI. The observed negative relationship between Pb concentrations and spectral reflectance provides a solid foundation for constructing satellite-based spectral estimation models of soil Pb.

3.3. Comparison of Satellite-Based Spectral Estimation Models for Soil Pb

Satellite-based spectral estimation models of soil Pb were developed using the XGBoost algorithm with two spectral feature sets: CSFS (R_{840}, R_{1013}, R_{1215}, R_{2239}) and ESFS (R_{840}, R_{1013}, R_{1215}, R_{2239}, SPPI-2, and SPPI-3). The validation results of the models are shown in Figure 9. Model IV (using ESFS as input variables) (Figure 9d) shows the best performance, with R^2, RMSE, and RPD values of 0.78, 1.11 mg/kg, and 2.10, respectively. The deviation between measured and estimated Pb values from Model IV was minimal. Model II (Figure 9b) was calibrated using a combination of CSFS and SPPI-2 as input variables, and it also showed good performance (R^2 = 0.73, RPD = 1.90). Furthermore, Models III (Figure 9c) and I (Figure 9a) showed moderate accuracy, with RPD values of 1.77 and 1.61, respectively. Overall, the XGBoost algorithm combined with ESFS produced an accurate satellite-based spectral model for soil Pb.

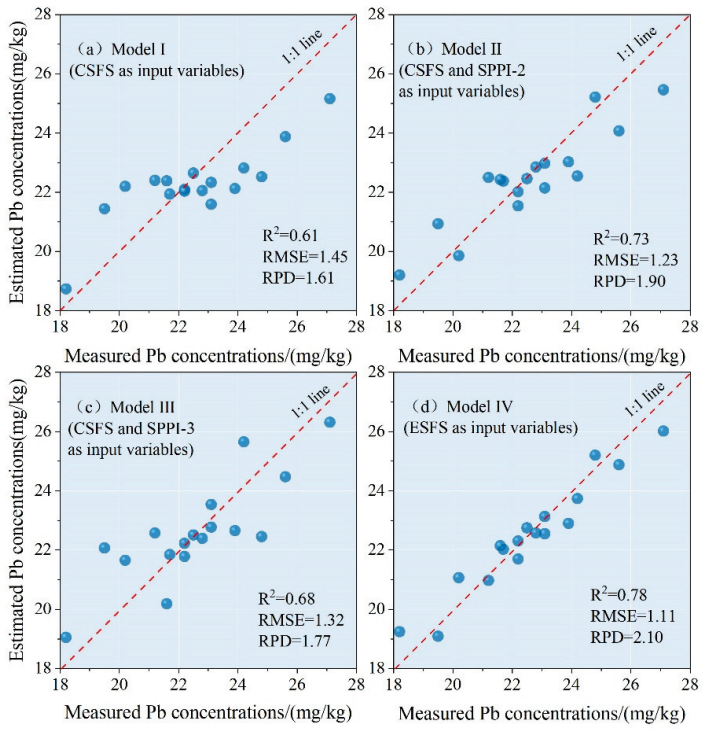


Figure 9. Scatter plots of measured vs. estimated Pb values in soils using satellite-based spectral estimation models based on the common spectral feature set (CSFS) and enhanced spectral feature set (ESFS). SPPI-2 and SPPI-3 represent the newly developed spectral indices for Pb-contaminated soils. The unit of RMSE is mg/kg.

https://doi.org/10.3390/rs18091446

3.4. Soil Pb Maps

To assess the applicability of the satellite-based spectral estimation model of soil Pb (Model IV, as indicated in Figure 9d), the model was applied to coal mining bare-soil areas using the DS-transformed satellite hyperspectral data. The estimated soil Pb values and their distribution map are shown in Figure 10. Contamination hotspots were mainly found in the eastern part of the coal mining area, while soils in the central and western regions were relatively clean. Most abandoned mines are located in the southeast of the coal mining area (as indicated in Figure 2), which may contribute to soil Pb contamination through the transfer of coal slag and coal ash from the historical mining activities [80,81]. In addition, contamination hotspots were also observed in roadside soils, likely due to inappropriate coal transportation activities.

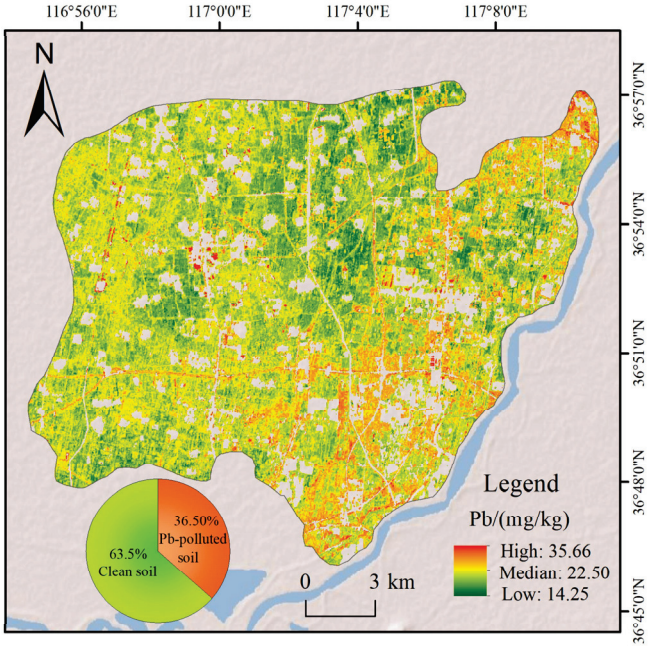


Figure 10. Distribution map of soil Pb concentration in a coal mining bare-soil area in China.

Based on the reference background value of soil Pb (21.40 mg/kg) [51], approximately 36.5% of soils in the coal mining bare-soil area are classified as Pb-contaminated, while the remaining 63.5% are regarded as clean. Pb-contaminated soils are mainly distributed in the east and south of the coal mining area, while clean soils are primarily found in the west and north. The estimated Pb concentrations range from 14.25 to 35.66 mg/kg, which are basically consistent with the measured soil Pb concentrations (17.40–29.40 mg/kg). Mapping results indicate that the satellite-based spectral estimation model for soil Pb (Model IV) developed in this study is reliable.

3.5. Uncertainty in the Soil Pb Mapping

The satellite-based spectral estimation model of soil Pb (Model IV, as indicated in Figure 9d) was iterated 50 times and its standard deviation was calculated to assess the uncertainty of soil Pb mapping. The uncertainty of soil Pb mapping using the satellite-based spectral estimation model is shown in Figure 11. The uncertainty of soil Pb mapping ranges from 0 to 0.5 mg/kg, with an average of 0.44 mg/kg. The eastern region of the study area exhibits higher uncertainty than the western region, indicating that soils in the eastern area have stronger spatial heterogeneity and more complex environmental conditions, thereby increasing the uncertainty of soil Pb mapping. In the future, incorporating appropriate

https://doi.org/10.3390/rs18091446

environmental covariates into satellite-based spectral modeling or increasing the number of field samples collected in specific areas may reduce the uncertainty of soil Pb mapping [70].

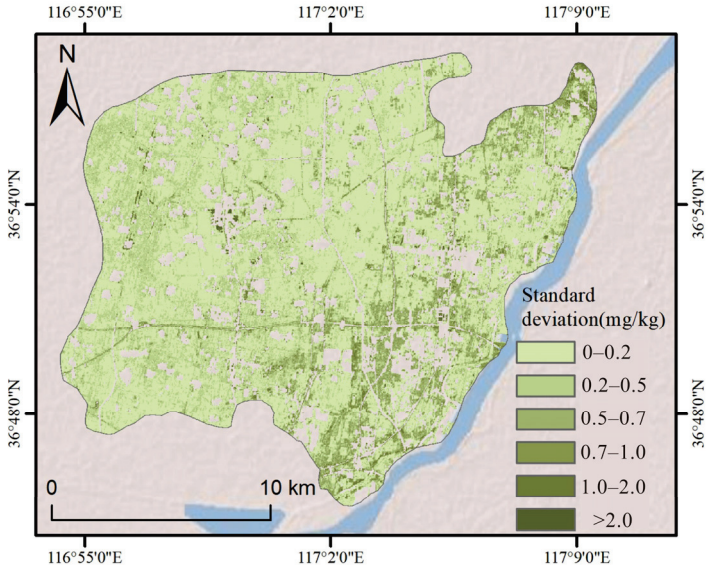


Figure 11. Uncertainty in the soil Pb mapping (the standard deviation represents the average results of estimated Pb values from 50 iterations of the model).

4. Discussion

In this study, we developed an enhanced spectral feature set (i.e., ESFS) derived from a controlled spectral experiment with laboratory-prepared Pb-contaminated soil samples. This ESFS comprises four characteristic bands (R_{840}, R_{1013}, R_{1215}, and R_{2239}) and two newly developed spectral indices (SPPI-2, SPPI-3), which effectively characterize soil Pb status in coal mining areas. Then, the ESFS was adapted for satellite application by applying a DS-based transformation to satellite hyperspectral data. The spectral parameters from the ESFS were used as independent variables, and the measured Pb concentrations from field soil samples were used as dependent variables. The satellite-based spectral estimation model for soil Pb was constructed using the XGBoost algorithm.

A controlled spectral measurement experiment using laboratory-prepared Pb-contaminated soil samples is recommended for understanding the spectral mechanisms of Pb-contaminated soils and for identifying the spectral features of soil Pb. Common methods for extracting spectral features, such as supervised searching methods (e.g., competitive adaptive reweighted sampling, genetic algorithms, and successive projections algorithm), were applied based on a repeated iterative fitting approach [20,44,82]. A statistical result of the selected bands was used in CSFS. However, environmental factors such as moisture, particle size, and scattered light were not considered during feature extraction [65,83]. As a result, the CSFS may only be suitable under a given environmental condition. The CSFS faces challenges in accurately characterizing soil Pb in areas where soil environments change significantly, especially in coal mining regions. Therefore, soil Pb contamination was prepared and a controlled spectral measurement experiment was conducted, with interfering factors such as moisture, particle size, scattered light, and soil background properties excluded. The experimental data supported understanding the spectral response of Pb-contaminated soils, and the new spectral indices for Pb-contaminated soils (SPPI-2 and SPPI-3) were developed, leading to the establishment of the ESFS.

The new spectral indices for Pb-contaminated soil, SPPI-2 and SPPI-3, are a crucial part of the ESFS, which can effectively indicate the Pb contamination status in soils. The spectral indices SPPI-2 and SPPI-3 were calculated using tangent angles from four charac-

teristic bands (R_{840}, R_{1013}, R_{1215}, and R_{2239}). The spectral features of Pb-contaminated soils, identified using the tangent method, were further enhanced by strategically combining these features and by analyzing the interrelationships among key spectral bands [84]. In soil, Pb exists in forms bound to active substances (such as organic matter, iron oxides, and clay) adsorbed in the soil. Therefore, the spectral features displayed by Pb are closely linked to these spectrally active soil components [85,86]. Sun et al. [86] suggested that clay in soil exhibits spectral features around 2200 nm due to the influence of soil clay minerals. Li et al. [87] indicated that spectral bands near 1000 nm, 1050 nm, 1300 nm, 2200 nm, and 2260 nm correlate with certain functional groups in soil organic matter. The spectral ranges of 870 nm and 1000–1100 nm are sensitive to iron oxides in soil, but are also influenced by soil organic matter [23,88,89]. In addition, Wang et al. [9] reported that the spectral ranges of 375–800, 900–1200, 1300–1450, 1700–2080, 2150, and 2200–2400 nm are sensitive spectral regions for heavy metals in soil. The characteristic bands selected in this study and the newly developed spectral indices (SPPI-2 and SPPI-3) further confirm existing spectral mechanisms of lead-contaminated soils. In addition, SPPI-2 and SPPI-3 highlight the global changes in the spectral curves of Pb-contaminated soils, while single spectral bands represent localized spectral features that may be masked or affected by environmental factors or other soil components.

Additionally, common spectral indices of soils were constructed in the comparative experiment, and showed unsatisfactory performance in characterizing the soil Pb status. Common spectral indices such as the difference spectral index (DI), ratio spectral index (RI), and normalized difference spectral index (NDI) were developed [31,90], and their CC values are shown in Figure 12. The absolute CC values for the common spectral indices range from 0 to 0.36, which are lower than those of SPPI-2 (0.53) and SPPI-3 (0.62). These differences suggest that common spectral indices are ineffective for indicating soil Pb status in coal mining bare-soil areas, likely due to the complex soil environment and influences from other soil components.

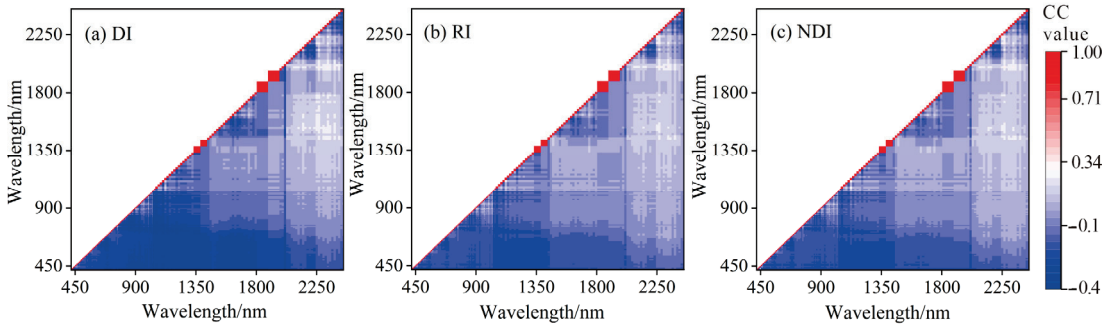


Figure 12. Pearson correlation coefficients (CC) of common spectral indices: (**a**) Difference spectral index (DI); (**b**) Ratio spectral index (RI); (**c**) Normalized difference spectral index (NDI).

The satellite-based spectral estimation model of soil Pb based on the ESFS and XGBoost (Model IV, shown in Figure 9) performed better than other commonly used models. Several models, such as PLSR, ANN, ELM, SVM, and RF, were constructed [64,69,72–74], and their performance metrics are listed in Table 3. The RF model achieved the highest accuracy among the common models, with an R^2 of 0.70 and an RPD of 1.83. The RPD value of the RF model was still about 12.86% lower than that of model IV. The ELM model ranked second for soil Pb spectral estimation (RPD = 1.70), while PLSR, ANN, and SVM had relatively lower accuracy (RPD ranging from 0.78 to 1.62). The Model IV, based on the ESFS and XGBoost algorithm, demonstrated the best accuracy and reliability, improving the minimum accuracy for soil Pb spectral estimation by 12.86% compared to the common models.

https://doi.org/10.3390/rs18091446

Table 3. Comparison of satellite-based spectral estimation models of soil Pb constructed using common methods.

Model	R^2	RMSE	RPD	Key Hyperparameters
PLSR	0.32	2.99	0.78	n_components = 3
ANN	0.61	1.44	1.62	hidden_layers = 5; activation = relu;
SVM	0.54	1.65	1.41	kernel = linear; gamma = 0.01; cost = 1
RF	0.70	1.27	1.83	n_estimators = 50; max_depth = 3; max_features = log2; min_samples_split = 4; min_samples_leaf = 2;
ELM	0.63	1.37	1.70	n_hidden = 50; activation = sigmoid; Regularization = 0.1
This study (Model IV)	0.78	1.11	2.10	n_estimators = 60; learning_rate = 0.1; max_depth = 5; colsample_bytree = 0.6; subsample = 0.8; gamma = 0

In addition, we acknowledge that improvements in satellite-based spectral estimation of soil Pb rely on the construction of a robust spectral feature set, the use of advanced machine learning models, and the optimization of the methodological framework. Clearly, the above improvements and optimizations are developed within the existing theoretical framework of soil hyperspectral estimation [9]. In future work, further efforts could focus on developing new theoretical frameworks to achieve greater improvements in the satellite-based spectral estimation and mapping of soil heavy metals.

The DS-based transformation method plays an important role in establishing the ESFS using satellite hyperspectral data. Differences between satellite and laboratory spectra, such as spectral resolution and reflectance values, are commonly observed. These differences are mainly caused by environmental factors and differences in spectrometer sensors. It should be noted that the ESFS was originally developed using laboratory spectra. DS is commonly used for transferring data or models between different spectrometers under varying measurement conditions, and its reliability has been demonstrated in studies by Ji et al. [62] and Xu et al. [45]. Therefore, in this study, the DS method was used to transform the satellite hyperspectral data, allowing the ESFS to be transferred from laboratory to satellite data and ensuring its applicability to the satellite data.

The sample size and representativeness of field-collected soil samples play an important role in the performance of satellite-based spectral estimation models. In this study, before field sampling, land use, soil type, mine distribution, and road conditions in the study area were considered to select suitable sampling locations and ensure spatial representativeness of the samples. Additionally, the XGBoost model, which is suitable for different sizes of sample datasets, was used for spectral modeling [37,38]. Results show that when the number of soil samples is less than 50, the XGBoost-based satellite estimation model has low accuracy ($R^2 < 0.60$ and RPD < 1.40) and cannot provide acceptable estimates of soil Pb. When the training sample size increases to more than 60, the model accuracy shows no obvious change and remains stable. These efforts help achieve a balance between model performance and sampling cost. In future work, more soil samples will be collected, and advanced deep learning models will be introduced to further improve the accuracy and reliability of satellite-based spectral estimation models.

To enhance the transferability of the proposed framework (integrating ESFS with the XGBoost model), two aspects merit further consideration. First, noise introduced by factors such as soil moisture in satellite hyperspectral imagery should be carefully assessed and minimized, as this would improve the reliability and accuracy of ESFS in characterizing soil Pb [91]. Second, the development of a reliable XGBoost model relies on representative field soil samples, appropriate hyperparameter tuning, and comprehensive evaluation using independent validation datasets [92]. These efforts are expected to improve the

applicability of the proposed framework in other environments, such as agricultural soils or post-conflict areas.

Some limitations exist that will need to be addressed as priorities in future research. We recognize that laboratory spectra of field-collected Pb-contaminated soil samples and laboratory-prepared Pb-contaminated soil samples may have slight differences in shape and absorption features, due to the differences in organic matter, clay, and other soil components between these two types of samples [22]. Although the DS method has been proven effective in reducing differences between the two types of spectra, the spectra of field-collected and laboratory-prepared Pb-contaminated soil samples are quasi-consistent. Moving forward, the mechanisms and sources of spectral differences between field-collected and laboratory-prepared Pb-contaminated soil samples should be investigated and eliminated to improve the applicability of the ESFS. In addition, the combination of cross-validation and grid search achieved an optimal configuration of model hyperparameters. The calibrated model achieved a balance between accuracy and computational cost in this study. In the next step, the integration of automatic hyperparameter search strategies (such as Bayesian optimization) with advanced artificial intelligence models (such as convolutional neural networks and transfer learning) will significantly improve the accuracy and robustness of satellite-based spectral estimation models [68,93].

5. Conclusions

This study proposes a robust satellite-based framework for the estimation and mapping of soil Pb based on the ESFS and the XGBoost model, and its effectiveness has been demonstrated through a case study in a coal mining area. The ESFS was developed using controlled spectral experiment of laboratory-prepared Pb-contaminated soils employing the SPA-SHAP method. Two new spectral indices for Pb-contaminated soils (SPPI-2 and SPPI-3) were constructed. Next, the ESFS was established using DS-transformed satellite hyperspectral data, where the spectral parameters in the ESFS were used as independent variables, and the measured Pb concentrations in field soil samples served as the dependent variable. Finally, a satellite-based spectral estimation model of soil Pb based on the XGBoost algorithm was developed. Results showed that: (1) the spectral parameters in the ESFS included R_{840}, R_{1013}, R_{1215}, R_{2239}, SPPI-2, and SPPI-3; (2) the optimal satellite-based estimation model of soil Pb, calibrated with the XGBoost algorithm and ESFS, achieved R^2 and RPD values of 0.78 and 2.10, respectively, which are at least 12.86% higher than those of common models; (3) Pb-contaminated soils were mainly distributed in the eastern part of the coal mining area, which is associated with improper coal mining and transportation activities. The results presented in this study provide an effective approach based on the ESFS and XGBoost model for the satellite estimation and mapping of soil Pb, providing strong support for the monitoring of soil Pb status and Pb-contaminated hotspot areas, and are valuable for soil management and environmental sustainability worldwide.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18091446/s1, Table S1: Volumes of $Pb(NO_3)_2$ solution added for laboratory preparation of Pb-contaminated soil samples.

Author Contributions: Conceptualization, X.X. and Y.W.; methodology, X.D.; software, Z.W.; formal analysis, Q.W.; investigation, X.X.; resources, Q.S.; data curation, X.X. and Y.W.; writing—original draft preparation, X.X.; writing—review and editing, X.D. and J.C.; supervision, J.C.; project administration, J.C.; funding acquisition, J.C. and Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of Shandong Province, China (ZR2025QC1023), National Natural Science Foundation of China (42501074), Youth Innovation Team

Program of Colleges and Universities in Shandong Province (2022KJ248), and Key Laboratory of Environmental Change and Natural Disaster of the Ministry of Education, Beijing Normal University (2023-KF-12).

Data Availability Statement: Data is available upon request.

Acknowledgments: We thank the Analysis and Test Center, State Key Laboratory of Earth Surface Processes and Hazard Risk Governance, Beijing Normal University, for their support.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang, M.; Wang, J.; Li, S. Tempo-spatial changes and main anthropogenic influence factors of vegetation fractional coverage in a large-scale opencast coal mine area from 1992 to 2015. *J. Clean. Prod.* **2019**, *232*, 940–952. [CrossRef]
2. Zerizghi, T.; Guo, Q.; Tian, L.; Wei, R.; Zhao, C. An integrated approach to quantify ecological and human health risks of soil heavy metal contamination around coal mining area. *Sci. Total Environ.* **2022**, *814*, 152653. [CrossRef]
3. Liu, X.; Bai, Z.; Shi, H.; Zhou, W.; Liu, X. Heavy metal pollution of soils from coal mines in China. *Nat. Hazards* **2019**, *99*, 1163–1177. [CrossRef]
4. Islam, A.; Jion, M.; Jannat, J.; Varol, M.; Islam, M.; Khan, R.; Idris, A.; Malafaia, G.; Habib, M. Perception and legacy of soil chromium and lead contamination in an operational small-scale coal mining community. *Environ. Geochem. Health* **2023**, *45*, 7237–7253. [CrossRef] [PubMed]
5. Zhu, L.; Zhang, L.; Wang, J.; Lv, J. Combining finite mixture distribution, receptor model, and geostatistical simulation to evaluate heavy metals pollution in soils: Source and spatial pattern. *Land Degrad. Dev.* **2021**, *32*, 2105–2115. [CrossRef]
6. Parzentny, H.R.; Róg, L. Distribution and mode of occurrence of Co, Ni, Cu, Zn, As, Ag, Cd, Sb, Pb in the feed coal, fly ash, slag, in the topsoil and in the roots of trees and undergrowth downwind of three power stations in Poland. *Minerals* **2021**, *11*, 133. [CrossRef]
7. Li, J.; Xu, X.; Lv, J.; Wu, Q.; Ren, M.; Cao, J.; Liu, P. Source apportionment and health risk quantification for heavy metal sources in soils near aluminum-plastic manufacturing facilities in northeast China. *Hum. Ecol. Risk Assess.* **2020**, *26*, 2225–2244. [CrossRef]
8. Collin, M.S.; Venkatraman, S.K.; Vijayakumar, N.; Kanimozhi, V.; Arbaaz, S.M.; Stacey, R.S.; Anusha, J.; Choudhary, R.; Lvov, V.; Tovar, G.I. Bioaccumulation of lead (Pb) and its effects on human: A review. *J. Hazard. Mater. Adv.* **2022**, *7*, 100094. [CrossRef]
9. Wang, Y.; Zou, B.; Chai, L.; Lin, Z.; Feng, H.; Tang, Y.; Tian, R.; Tu, Y.; Zhang, B.; Zou, H. Monitoring of soil heavy metals based on hyperspectral remote sensing: A review. *Earth-Sci. Rev.* **2024**, *254*, 104814. [CrossRef]
10. Mirzaei, M.; Verrelst, J.; Marofi, S.; Abbasi, M.; Azadi, H. Eco-friendly estimation of heavy metal contents in grapevine foliage using in-field hyperspectral data and multivariate analysis. *Remote Sens.* **2019**, *11*, 2731. [CrossRef]
11. Tan, K.; Ma, W.; Chen, L.; Wang, H.; Du, Q.; Du, P.; Yan, B.; Liu, R.; Li, H. Estimating the distribution trend of soil heavy metals in mining area from HyMap airborne hyperspectral imagery based on ensemble learning. *J. Hazard. Mater.* **2021**, *401*, 123288. [CrossRef] [PubMed]
12. Wang, Z.; Chen, X.; Yu, D.; Zhang, L.; Wang, J.; Lv, J. Source apportionment and spatial distribution of potentially toxic elements in soils: A new exploration on receptor and geostatistical models. *Sci. Total Environ.* **2021**, *759*, 143428. [CrossRef]
13. Reza, S.; Baruah, U.; Singh, S.; Das, T. Geostatistical and multivariate analysis of soil heavy metal contamination near coal mining area, Northeastern India. *Environ. Earth Sci.* **2015**, *73*, 5425–5433. [CrossRef]
14. Zhou, M.; Zou, B.; Tu, Y.; Feng, H.; He, C.; Ma, X.; Ning, J. Spectral response feature bands extracted from near standard soil samples for estimating soil Pb in a mining area. *Geocarto Int.* **2022**, *37*, 13248–13267. [CrossRef]
15. Mzid, N.; Castaldi, F.; Tolomio, M.; Pascucci, S.; Casa, R.; Pignatti, S. Evaluation of agricultural bare soil properties retrieval from Landsat 8, Sentinel-2 and PRISMA satellite data. *Remote Sens.* **2022**, *14*, 714. [CrossRef]
16. Canero, F.M.; Rodriguez-Galiano, V.; Chabrillat, S. Evaluating Airborne Hyperspectral Scanner (AHS) for the mapping of soil organic matter and clay in a Mediterranean forest ecosystem. *Catena* **2025**, *252*, 108889. [CrossRef]
17. Zhang, J.; Ding, J.; Tan, J.; Wang, J.; Zhang, Z.; Wang, Z.; Ge, X. Monitoring soil salinization in arid cotton fields using unmanned aerial vehicle hyperspectral imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *140*, 104584. [CrossRef]
18. Ma, X.; Wang, J.; Zhou, K.; Zhang, W.; Chen, A. Uncertainty in soil elemental prediction using machine learning and hyperspectral remote sensing. *J. Hazard. Mater.* **2025**, *494*, 138502. [CrossRef] [PubMed]
19. Shen, Q.; Shang, K.; Xiao, C.; Tang, H.; Wu, T.; Wang, C. A novel hyperspectral remote sensing estimation model for surface soil texture using AHSI/ZY1-02D satellite image. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *138*, 104453. [CrossRef]
20. Xu, X.; Ren, M.; Cao, J.; Wu, Q.; Liu, P.; Lv, J. Spectroscopic diagnosis of zinc contaminated soils based on competitive adaptive reweighted sampling algorithm and an improved support vector machine. *Spectrosc. Lett.* **2019**, *53*, 86–99. [CrossRef]

21. Liu, Z.; Lu, Y.; Peng, Y.; Zhao, L.; Wang, G.; Hu, Y. Estimation of soil heavy metal content using hyperspectral data. *Remote Sens.* **2019**, *11*, 1464. [CrossRef]
22. Chen, L.; Lai, J.; Tan, K.; Wang, X.; Chen, Y.; Ding, J. Development of a soil heavy metal estimation method based on a spectral index: Combining fractional-order derivative pretreatment and the absorption mechanism. *Sci. Total Environ.* **2022**, *813*, 151882. [CrossRef]
23. Rossel, R.V.; Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **2010**, *158*, 46–54. [CrossRef]
24. Sun, W.; Liu, S.; Zhang, X.; Zhu, H. Performance of hyperspectral data in predicting and mapping zinc concentration in soil. *Sci. Total Environ.* **2022**, *824*, 153766. [CrossRef] [PubMed]
25. Nawar, S.; Cipullo, S.; Douglas, R.; Coulon, F.; Mouazen, A. The applicability of spectroscopy methods for estimating potentially toxic elements in soils: State-of-the-art and future trends. *Appl. Spectrosc. Rev.* **2020**, *55*, 525–557. [CrossRef]
26. Lei, H.; Bao, N.; Yu, M.; Cao, Y. Estimating and mapping tailings properties of the largest iron cluster in China for resource potential and reuse: A new perspective from interpretable CNN model and proposed spectral index based on hyperspectral satellite imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *139*, 104512. [CrossRef]
27. Prakash, A.; Thangaraj, J.; Roy, S.; Srivastav, S.; Mishra, J.K. Model-aware XGBoost method towards optimum performance of flexible distributed Raman amplifier. *IEEE Photonics J.* **2023**, *15*, 1–10. [CrossRef]
28. Xu, X.; Wang, X.; Yang, P.; Meng, Y.; Yu, D.; Li, C. Strategy for mapping soil salt contents during the bare soil period through a satellite image: Optimal calibration set combined with random forest. *Catena* **2023**, *223*, 106900. [CrossRef]
29. Dharumarajan, S.; Lalitha, M.; Kalaiselvi, B.; Kaliraj, S.; Adhikari, K.; Vasundhara, R.; Niranjana, K.; Hegde, R.; Pradeep, C.; Hittanagi, P. Remote sensing of soils: Spectral signatures and spectral indices. In *Remote Sensing of Soils*; Elsevier: Amsterdam, The Netherlands, 2024; pp. 13–23.
30. Wang, J.; Ding, J.; Yu, D.; Ma, X.; Zhang, Z.; Ge, X.; Teng, D.; Li, X.; Liang, J.; Lizaga, I.; et al. Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China. *Geoderma* **2019**, *353*, 172–187. [CrossRef]
31. Zhang, Z.; Ding, J.; Wang, J.; Ge, X. Prediction of soil organic matter in northwestern China using fractional-order derivative spectroscopy and modified normalized difference indices. *Catena* **2020**, *185*, 104257. [CrossRef]
32. dos Santos, E.P.; Moreira, M.C.; Fernandes-Filho, E.I.; Demattê, J.A.; dos Santos, U.J.; Moura-Bueno, J.M.; Cruz, R.R.P.; da Silva, D.D.; Sampaio, E.V.d.S.B. Integrating satellite radar vegetation indices and environmental descriptors with visible-infrared soil spectroscopy improved organic carbon prediction in soils of semi-arid Brazil. *Geoderma* **2025**, *457*, 117288. [CrossRef]
33. Wang, S.; Zhou, Y.; You, X.; Wang, B.; Du, L. Quantification of the antagonistic and synergistic effects of Pb^{2+}, Cu^{2+}, and Zn^{2+} bioaccumulation by living Bacillus subtilis biomass using XGBoost and SHAP. *J. Hazard. Mater.* **2023**, *446*, 130635. [CrossRef]
34. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T. *R Package*, version 0.4-2. XGBoost: Extreme Gradient Boosting. CRAN (Comprehensive R Archive Network): Vienna, Austria, 2015.
35. Sun, M.; Liu, H.; Li, P.; Gong, P.; Yu, X.; Ye, F.; Guo, Y.; Wu, Z. Effects of salt content and particle size on spectral reflectance and model accuracy: Estimating soil salt content in arid, saline-alkali lands. *Microchem. J.* **2024**, *207*, 111666. [CrossRef]
36. Luo, Y.; Li, C.; Huang, J.; Dong, C.; Wang, J. Integrating fractional-order derivatives of soil and leaf hyperspectral reflectance for improved estimation of mangrove soil organic carbon. *Geoderma* **2025**, *458*, 117324. [CrossRef]
37. Liang, H.; Song, Y.; Dai, Z.; Liu, H.; Zhong, K.; Feng, H.; Xu, L. Soil total nitrogen content and pH value estimation method considering spatial heterogeneity: Based on GNNW-XGBoost model. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2025**, *330*, 125716. [PubMed]
38. Ye, M.; Zhu, L.; Li, X.; Ke, Y.; Huang, Y.; Chen, B.; Yu, H.; Li, H.; Feng, H. Estimation of the soil arsenic concentration using a geographically weighted XGBoost model based on hyperspectral data. *Sci. Total Environ.* **2023**, *858*, 159798.
39. Sui, J.; Hao, Q. Evaluation of geological environmental carrying capacity in the Huanghebei coal mining area, Shandong Province. *Shandong Land Resour.* **2017**, *33*, 34–40.
40. Zhang, B.; Yang, K.; Zhang, K.; Wang, Q.; Wu, N. Migration transformation, prevention, and control of typical heavy metal lead in coal gangue: A review. *Int. J. Coal Sci. Technol.* **2023**, *10*, 85. [CrossRef]
41. Wang, W.; Xu, X.; Zhou, Z.; Dong, X.; Tian, T. A joint method to assess pollution status and source-specific human health risks of potential toxic elements in soils. *Environ. Monit. Assess.* **2022**, *194*, 685. [CrossRef] [PubMed]
42. Anaman, R.; Peng, C.; Jiang, Z.; Liu, X.; Zhou, Z.; Guo, Z.; Xiao, X. Identifying sources and transport routes of heavy metals in soil with different land uses around a smelting site by GIS-based PCA and PMF. *Sci. Total Environ.* **2022**, *823*, 153759. [PubMed]
43. Cheng, S.; Yang, X.; Yang, G.; Chen, B.; Chen, D.; Wang, J.; Ren, K.; Sun, W. Using ZY1-02D satellite hyperspectral remote sensing to monitor landscape diversity and its spatial scaling change in the Yellow River Estuary. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *128*, 103716. [CrossRef]

44. Sun, W.; Liu, K.; Ren, G.; Liu, W.; Yang, G.; Meng, X.; Peng, J. A simple and effective spectral-spatial method for mapping large-scale coastal wetlands using China ZY1-02D satellite hyperspectral images. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102572. [CrossRef]
45. Xu, X.; Chen, Y.; Yang, S. Improving soil organic matter mapping using transfer learning and satellite-simulated samples from bare soil hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 1706–1717. [CrossRef]
46. Ahmad, W.; Liu, L.; Guo, Z.; Khalil, Y.S.; Islam, N.U.; Islam, F. Lithological classification using ZY1-02D hyperspectral data by means of machine learning and deep learning methods in the Kohat–Pothohar Plateau, Khyber Pakhtunkhwa, Pakistan. *Remote Sens.* **2025**, *17*, 1356. [CrossRef]
47. Zou, B.; Jiang, X.; Feng, H.; Tu, Y.; Tao, C. Multisource spectral-integrated estimation of cadmium concentrations in soil using a direct standardization and spiking algorithm. *Sci. Total Environ.* **2020**, *701*, 134890. [CrossRef]
48. Davis, E.; Wang, C.; Dow, K. Comparing Sentinel-2 MSI and Landsat 8 OLI in soil salinity detection: A case study of agricultural lands in coastal North Carolina. *Int. J. Remote Sens.* **2019**, *40*, 6134–6153. [CrossRef]
49. Xu, X.; Chen, Y.; Dai, X.; Lei, T.; Wang, S.; Li, K. An improved Vis-NIR estimation model of soil organic matter through the artificial samples enhanced calibration set. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 4626–4637. [CrossRef]
50. Zuo, T.; Zhang, L.; Wang, Y.; Nie, L.; Shen, M.; Liu, L.; Yu, J.; Jin, H.; Wei, F.; Ma, S. Technical guidelines for risk assessment of heavy metals in traditional Chinese medicines. *Chin. Med.* **2023**, *18*, 69. [CrossRef] [PubMed]
51. Lv, J.; Yu, Y. Source identification and spatial distribution of metals in soils in a typical area of the lower Yellow River, eastern China. *Environ. Sci. Pollut. Res.* **2018**, *25*, 21106–21117. [CrossRef]
52. Cristóbal, J.; Graham, P.; Prakash, A.; Buchhorn, M.; Gens, R.; Guldager, N.; Bertram, M. Airborne hyperspectral data acquisition and processing in the arctic: A pilot study using the hyspex imaging spectrometer for wetland mapping. *Remote Sens.* **2021**, *13*, 1178. [CrossRef]
53. Jiang, G.; Zhou, S.; Cui, S.; Chen, T.; Wang, J.; Chen, X.; Liao, S.; Zhou, K. Exploring the potential of HySpex hyperspectral imagery for extraction of copper content. *Sensors* **2020**, *20*, 6325. [CrossRef]
54. Yang, S.; Li, J.; Li, J.; Zhang, X.; Ma, C.; Liu, Z.; Ren, M. Estimating the canopy nitrogen content in maize by using the transform-based dynamic spectral indices and random forest. *Sustainability* **2024**, *16*, 8011. [CrossRef]
55. Hong, Y.; Liu, Y.; Chen, Y.; Liu, Y.; Yu, L.; Liu, Y.; Cheng, H. Application of fractional-order derivative in the quantitative estimation of soil organic matter content through visible and near-infrared spectroscopy. *Geoderma* **2019**, *337*, 758–769. [CrossRef]
56. Verboven, S.; Hubert, M.; Goos, P. Robust preprocessing and model selection for spectral data. *J. Chemom.* **2012**, *26*, 282–289. [CrossRef]
57. Gao, Y.; Cui, L.; Lei, B.; Zhai, Y.; Shi, T.; Wang, J.; Chen, Y.; He, H.; Wu, G. Estimating soil organic carbon content with visible–near-infrared (Vis-NIR) spectroscopy. *Appl. Spectrosc.* **2014**, *68*, 712–722. [CrossRef] [PubMed]
58. Jiao, Y.; Li, Z.; Chen, X.; Fei, S. Preprocessing methods for near-infrared spectrum calibration. *J. Chemom.* **2020**, *34*, e3306. [CrossRef]
59. Kamruzzaman, M.; Kalita, D.; Ahmed, M.T.; ElMasry, G.; Makino, Y. Effect of variable selection algorithms on model performance for predicting moisture content in biological materials using spectral data. *Anal. Chim. Acta* **2022**, *1202*, 339390. [CrossRef]
60. Zhong, L.; Guo, X.; Ding, M.; Ye, Y.; Jiang, Y.; Zhu, Q.; Li, J. SHAP values accurately explain the difference in modeling accuracy of convolution neural network between soil full-spectrum and feature-spectrum. *Comput. Electron. Agric.* **2024**, *217*, 108627. [CrossRef]
61. Xuan, F.; Liu, H.; Xue, J.; Li, Y.; Liu, J.; Huang, X.; Tan, Z.; Abd Elbasit, M.A.; Gu, X.; Su, W. The novel triangular spectral indices for characterizing winter wheat drought. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *134*, 104151. [CrossRef]
62. Ji, W.; Li, S.; Chen, S.; Shi, Z.; Rossel, R.A.V.; Mouazen, A.M. Prediction of soil attributes using the Chinese soil spectral library and standardized spectra recorded at field conditions. *Soil Tillage Res.* **2016**, *155*, 492–500. [CrossRef]
63. Pittaki-Chrysodonta, Z.; Hartemink, A.E.; Sanderman, J.; Ge, Y.; Huang, J. Evaluating three calibration transfer methods for predictions of soil properties using mid-infrared spectroscopy. *Soil Sci. Soc. Am. J.* **2021**, *85*, 501–519. [CrossRef]
64. Bouveresse, E.; Massart, D. Standardisation of near-infrared spectrometric instruments: A review. *Vib. Spectrosc.* **1996**, *11*, 3–15. [CrossRef]
65. Ji, W.; Viscarra Rossel, R.; Shi, Z. Improved estimates of organic carbon using proximally sensed vis–NIR spectra corrected by piecewise direct standardization. *Eur. J. Soil Sci.* **2015**, *66*, 670–678. [CrossRef]
66. Wang, J.; Zou, Q.; Xu, B.; Feng, Z.; Yuan, H. Enhancing Soil Organic Matter Prediction via Deep Transfer Learning from Mid-Infrared Soil Spectral Library to In-Situ Field. *Comput. Electron. Agric.* **2026**, *241*, 111243. [CrossRef]
67. Vitale, A.; Lamonaca, F. Enhancing GeoAI land cover classification via hyperparameter tuning and cross-validation: A case study in Ravenna, Italy. *Measurement* **2025**, *257*, 118662. [CrossRef]
68. Buzpinar, M.A.; Gunaydin, S.; Kavuncuoglu, E.; Cetin, N.; Sacilik, K.; Cheein, F.A. Model comparison and hyperparameter optimization for visible and near-infrared (Vis-NIR) spectral classification of dehydrated banana slices. *Expert Syst. Appl.* **2025**, *283*, 127858. [CrossRef]

69. Xu, X.; Zhai, X. Mapping soil organic matter content during the bare soil period by using satellite data and an improved deep learning network. *Sustainability* **2022**, *15*, 323. [CrossRef]
70. Xu, X.; Wang, Z.; Song, X.; Zhan, W.; Yang, S. A remote sensing-based strategy for mapping potentially toxic elements of soils: Temporal-spatial-spectral covariates combined with random forest. *Environ. Res.* **2024**, *240*, 117570. [CrossRef]
71. Wang, J.; Hu, X.; Shi, T.; He, L.; Hu, W.; Wu, G. Assessing toxic metal chromium in the soil in coal mining areas via proximal sensing: Prerequisites for land rehabilitation and sustainable development. *Geoderma* **2022**, *405*, 115399. [CrossRef]
72. Carvalho, M.; Cardoso-Fernandes, J.; Lima, A.; Teodoro, A.C. Convolutional neural networks applied to antimony quantification via soil laboratory reflectance spectroscopy in northern portugal: Opportunities and challenges. *Remote Sens.* **2024**, *16*, 1964.
73. Li, J.; Zheng, K.; Gao, L.; Ni, L.; Huang, M.; Chanussot, J. Model-informed multistage unsupervised network for hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–17. [CrossRef]
74. Hong, Y.; Shen, R.; Cheng, H.; Chen, Y.; Zhang, Y.; Liu, Y.; Zhou, M.; Yu, L.; Liu, Y.; Liu, Y. Estimating lead and zinc concentrations in peri-urban agricultural soils through reflectance spectroscopy: Effects of fractional-order derivative and random forest. *Sci. Total Environ.* **2019**, *651*, 1969–1982. [CrossRef]
75. Subi, X.; Eziz, M.; Zhong, Q.; Li, X. Estimating the chromium concentration of farmland soils in an arid zone from hyperspectral reflectance by using partial least squares regression methods. *Ecol. Indic.* **2024**, *161*, 111987. [CrossRef]
76. Xu, X.; Chen, Y.; Wang, M.; Wang, S.; Li, K.; Li, Y. Improving estimates of soil salt content by using two-date image spectral changes in Yinbei, China. *Remote Sens.* **2021**, *13*, 4165. [CrossRef]
77. Shi, Z.; Wang, Q.; Peng, J.; Ji, W.; Liu, H.; Li, X.; Viscarra Rossel, R.A. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations. *Sci. China Earth Sci.* **2014**, *57*, 1671–1680. [CrossRef]
78. Rossel, R.V.; Behrens, T.; Ben-Dor, E.; Brown, D.; Demattê, J.; Shepherd, K.D.; Shi, Z.; Stenberg, B.; Stevens, A.; Adamchuk, V. A global spectral library to characterize the world's soil. *Earth-Sci. Rev.* **2016**, *155*, 198–230. [CrossRef]
79. Ben-Dor, E.; Inbar, Y.; Chen, Y. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. *Remote Sens. Environ.* **1997**, *61*, 1–15. [CrossRef]
80. Chakraborty, P.; Singh, S.; Wood, D.A.; Hazra, B.; Kumar, P.; Kumar, P.V. Spatio-temporal variation, integrated quantification of source attribution, and health risk assessment of metal trace element contamination in coal mining soils of the Eastern Raniganj basin, India. *J. Environ. Chem. Eng.* **2025**, *13*, 116567. [CrossRef]
81. Zhao, J.; Cao, C.; Chen, X.; Zhang, W.; Ma, T.; Irfan, M.; Zheng, L. Source-specific ecological risk analysis and critical source identification of heavy metal (loid) s in the soil of typical abandoned coal mining area. *Sci. Total Environ.* **2024**, *947*, 174506. [CrossRef]
82. Wang, S.; Chen, Y.; Wang, M.; Zhao, Y.; Li, J. SPA-based methods for the quantitative estimation of the soil salt content in saline-alkali land from field spectroscopy data: A case study from the Yellow River Irrigation Regions. *Remote Sens.* **2019**, *11*, 967. [CrossRef]
83. Lobell, D.B.; Asner, G.P. Moisture effects on soil reflectance. *Soil Sci. Soc. Am. J.* **2002**, *66*, 722–727.
84. Zhao, D.; Xie, D.; Yin, F.; Liu, L.; Feng, J.; Ashraf, T. Estimation of Pb content using reflectance spectroscopy in farmland soil near metal mines, central China. *Remote Sens.* **2022**, *14*, 2420. [CrossRef]
85. Sun, W.; Zhang, X.; Sun, X.; Sun, Y.; Cen, Y. Predicting nickel concentration in soil using reflectance spectroscopy associated with organic matter and clay minerals. *Geoderma* **2018**, *327*, 25–35. [CrossRef]
86. Wu, Y.; Chen, J.; Ji, J.; Gong, P.; Liao, Q.; Tian, Q.; Ma, H. A mechanism study of reflectance spectroscopy for investigating heavy metals in soils. *Soil Sci. Soc. Am. J.* **2007**, *71*, 918–926. [CrossRef]
87. Li, T.; Mu, T.; Liu, G.; Yang, X.; Zhu, G.; Shang, C. A method of soil moisture content estimation at various soil organic matter conditions based on soil reflectance. *Remote Sens.* **2022**, *14*, 2411.
88. Wang, W.; Zhang, X.; Zheng, H.; Ding, S.; Shang, K.; Xiao, Q. Predicting soil organic matter with ZY1E hyperspectral images by correcting soil spectrum and expanding sample size. *Soil Tillage Res.* **2026**, *255*, 106815.
89. Bai, H.; Guo, B.; Pei, L.; Zhang, B.; Luo, P.; Guo, T.; Wu, M.; Ma, X. Retrieving soil heavy metal concentrations using hyperspectral remote sensing images based on three-band indices at a coal mine in China. *Adv. Space Res.* **2025**, *76*, 5973–5987. [CrossRef]
90. Li, W.; Xiang, Y.; Liu, X.; Tang, Z.; Wang, X.; Huang, X.; Shi, H.; Chen, M.; Duan, Y.; Ma, L. Estimation of soil moisture content based on fractional differential and optimal spectral index. *Agronomy* **2024**, *14*, 184. [CrossRef]
91. Wang, D.; Tan, Y.; Li, C.; Xin, J.; Wang, Y.; Hou, H.; Gao, L.; Zhong, C.; Pan, J.; Li, Z. Recognizing and reducing effects of moisture-salt coexistence on soil organic matter spectral prediction: From laboratory to satellite. *Soil Tillage Res.* **2025**, *248*, 106397. [CrossRef]

92. Li, J.; Zheng, K.; Liu, W.; Li, Z.; Yu, H.; Ni, L. Model-guided coarse-to-fine fusion network for unsupervised hyperspectral image super-resolution. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5.
93. Li, J.; Zheng, K.; Gao, L.; Han, Z.; Li, Z.; Chanussot, J. Enhanced deep image prior for unsupervised hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 1–18.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

remote sensing

MDPI

Article

Applications of Airborne Hyperspectral Imagery in Rare Earth Element Exploration: A Case Study of the World-Class Bayan Obo Deposit, China

Cai Liu [1,2], Junting Qiu [3], Junchuan Yu [2,*], Yanbo Zhao [4], Yuanquan Xu [2], Xin Zhang [4], Bin Chen [2], Rong Xu [2], Qianli Ma [2], Gang Liu [2] and Jinzhong Yang [2]

[1] School of Earth Sciences and Resources, China University of Geosciences, Beijing 100083, China; liucai@mail.cgs.gov.cn
[2] China Aero Geophysical Survey and Remote Sensing Center for Natural Resources, Beijing 100083, China; xuyuanquan@mail.cgs.gov.cn (Y.X.); chenbin@mail.cgs.gov.cn (B.C.); xurong@mail.cgs.gov.cn (R.X.); maqianli@mail.cgs.gov.cn (Q.M.); liugang@mail.cgs.gov.cn (G.L.); yangjinzhong@mail.cgs.gov.cn (J.Y.)
[3] Beijing Research Institute of Uranium Geology, Beijing 100029, China; qiujt@email.curb.edu.cn
[4] Inner Mongolia Land and Resources Exploration and Development Co., Ltd., Hohhot 010020, China; fang9hs@126.com (Y.Z.); 18747972480@163.com (X.Z.)
* Correspondence: yujunchuan@mail.cgs.gov.cn; Tel.: +86-151-0115-7141

Highlights

What are the main findings?

- This study proposed a hyperspectral workflow for REE identification, that enabled the detection of REE-bearing minerals regardless of the host rock types.
- The absorption features of the REE-bearing mineral monazite within the 400–1000 nm range can be ascribed to Nd^{3+}.

What are the implications of the main findings?

- Airborne hyperspectral imagery enables effective identification of REEs information on the Earth's surface.
- The CASI-1500h imagery performed most effectively in identifying the locations of REEs among the three sensors.

Abstract

Rare earth elements (REEs) play an important role in emerging renewable energy technology, the production of advanced materials, energy conservation, and high-end manufacturing industries, making them an irreplaceable strategic resource. The diagnostic spectral absorption features of REEs in the visible and near-infrared spectrum can be effectively used for identifying the occurrences of REEs on the Earth's surface. This study systematically compared three airborne hyperspectral sensors—HyMap, CASI-1500h, and AisaFENIX 1K—for detecting REEs in the Bayan Obo area of Inner Mongolia, China. The CASI-1500h imagery performed most effectively in identifying the locations of REEs among the three sensors evaluated here. Additionally, this study proposed a hyperspectral workflow for REE identification, which enabled the detection of REE-bearing minerals regardless of the host rock types—including carbonatites and associated dikes, fenite-syenites, and metamorphic feldspar-quartz sandstone. Laboratory-based spectroscopy and mineral chemistry analyses indicated that the absorption features of the REE-bearing mineral monazite within the 400–1000 nm range can be ascribed to Nd^{3+}. This study demonstrates the potential of airborne hyperspectral technology for efficient and large-scale exploration of REE deposits.

Keywords: rare earth elements; hyperspectral remote sensing; Bayan Obo deposit

1. Introduction

The rare earth elements (REEs) are a group of 17 elements comprising scandium (Sc), yttrium (Y), and the 15 lanthanide elements (La–Lu), as defined by the International Union of Pure and Applied Chemistry [1–3]. Based on their electron configurations, the REEs are commonly subdivided into two groups: light REEs (La-Eu) and heavy REEs (Gd-Lu), plus Sc and Y [4]. Sc and Pm are not considered in this study. These REEs are indispensable for the production of a wide array of modern technological fields spanning renewable energy, electric vehicles, electronics manufacturing, defense systems, and catalysts [2,3].

A fundamental understanding of the exploration principles, formation mechanisms and prospecting indicators for REE deposits is necessary for meeting the increasing demand for REEs. REE deposits are predominantly formed by high-temperature magmatic and hydrothermal processes, which are commonly associated with carbonatites, alkaline igneous rocks, and hydrothermal systems. In addition, such deposits may also be formed through low-temperature erosion and weathering processes, occurring in placers, bauxites, laterites, and ion-adsorption clays [1].

Ore-bearing carbonatites have distinct gravity and magnetic signatures because they are relatively dense and commonly contain magnetite or associated with magnetic rocks [5,6]. In addition, REEs are frequently associated with radioactive elements such as uranium and thorium, which often generate significant radioactive anomalies [7]. There, this makes geophysical methods effective for the exploration of deposits of REEs. Meanwhile, geochemical methods have also been widely employed, particularly for ion-adsorbing deposits of REEs [2,7].

Hyperspectral technology enables the rapid and efficient analysis of large land surface areas, allowing researchers to narrow down target areas for valuable mineral deposits [8,9]. This, make hyperspectral technology highly suitable for natural resource exploration. REE-bearing minerals, such as monazite, bastnäsite, synchysite, and parisite, produce diagnostic absorption features in the visible and near-infrared (VNIR) and shortwave infrared (SWIR) spectra [10,11], with prominent absorptions observed particularly at 580, 745, 800, and 880 nm [10–12]. Theoretically, this allows hyperspectral remote sensing to identify REE minerals by analyzing their spectral signatures. Several studies have explored the feasibility of hyperspectral techniques for REE identification, from drill cores [13] to outcrops [14]. Others have utilized satellite-borne [15–17] and UAV (unmanned aerial vehicle)-borne [18,19] hyperspectral imagery for REE mapping. Despite these achievements, hyperspectral REE identification and mapping face several challenges, such as the low spatial resolution of satellite-borne hyperspectral images, limited area coverage of UAV-borne hyperspectral images, detection limitations for low REE concentrations, and inadequate data processing methods.

The increasing application of airborne sensors in recent years has greatly facilitated mineral exploration [20–23], as they ensures high spatial resolution and broad area coverage. In this study, we acquired airborne hyperspectral imagery of the Bayan Obo area in China using three airborne hyperspectral sensors (i.e., HyMap, HyVista Corporation, Sydney, Australia; CASI-1500h, ITRES Research Limited, Calgary, AB, Canada; AisaFENIX 1K, Specrtal Imaging Ltd., Oulu, Finland), and we propose a new method that allows researchers to quickly and efficiently identify REE minerals. A field validation has also been conducted to verify the effectiveness of using airborne hyperspectral technology in REE identification.

https://doi.org/10.3390/rs18081110

2. Geological Setting

The Bayan Obo REE-Nb-Fe deposit is the largest REE resource in the world with REE reserves documented at 57.4 Mt (average grade 6 wt.% of REE_2O_3) [24,25]. It is located at the northern margin of the North China Craton, adjacent to the Central Asian Orogenic Belt (Figure 1). The deposit is hosted in the massive dolomite, occurs within a suite of low-grade metamorphic rocks of the Mesoproterozoic Bayan Obo Group [24,26]. This group predominantly consist of sandstone, slate, limestone, dolomite, and siltstone, which unconformably overlies the basement rocks [24,25]. The crystalline basement rocks are composed of Neoarchean mylonitic granite-gneiss (~2588 Ma), Paleoproterozoic syenite and granodiorite (~2018 Ma), as well as biotite granite-gneiss and garnet-bearing granite-gneiss (~1890 Ma) [26].

It is widely believed that the initial mineralization of REEs and the metal resources of REEs originated from the Mesoproterozoic carbonatite magmas or REE-rich hydrothermal fluids exsolved from the carbonatites [25,26]. The primary ores were subsequently affected by the Early Paleozoic subduction-related magmas or thermal perturbation, leading to the remobilization and reprecipitation of REEs [25,26].

The dominant REE-bearing minerals are bastnäsite and monazite, which occur in association with a diverse suite of REE and Nb minerals, including huanghoite, aeschynite, fergusonite, and columbite [26]. These ores are dominated by light REEs, mostly including Ce, La, and Nd [26]. Over 100 carbonatite dikes are distributed around the deposit, which intrude into the Bayan Obo Group [26,27]. Furthermore, these carbonatite dikes exhibit REE contents ranging from 0.02 to 20.00 wt.% [26,28].

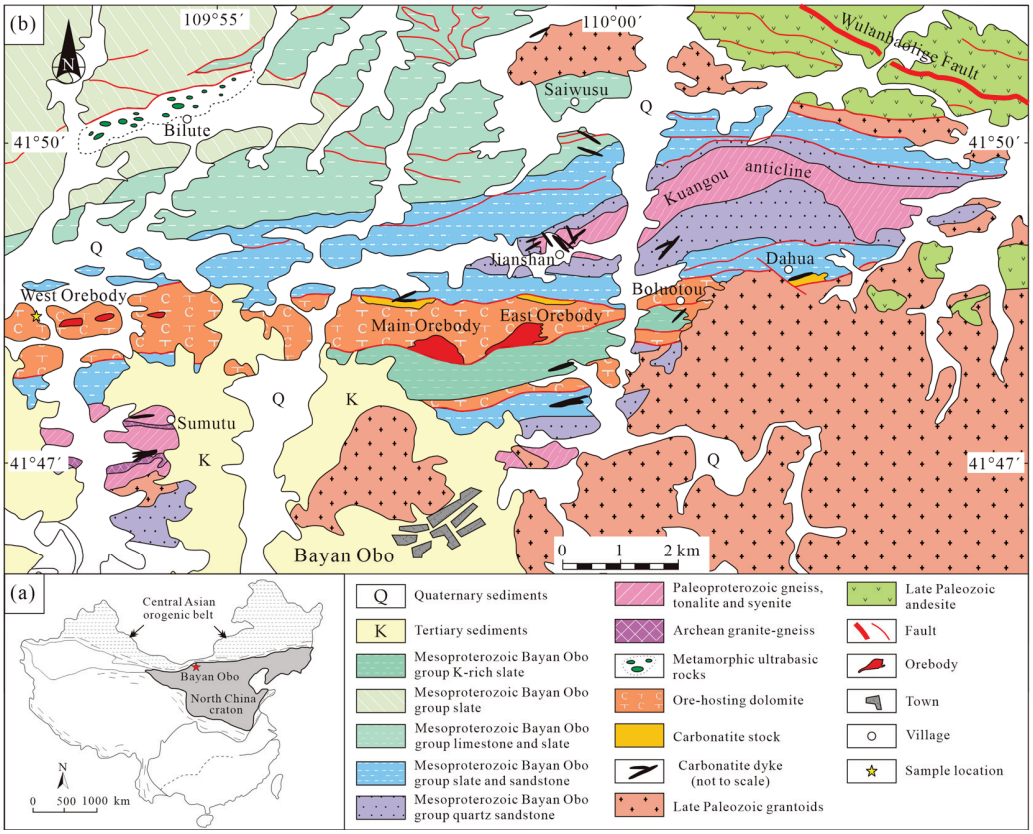


Figure 1. The location of the Bayan Obo area in the simple Chinese tectonic structure map (**a**), and its geological map (**b**), modified after Yang et al., 2011 [29].

3. Materials and Methods

3.1. Hyperspectral Data

The technical framework of this study is shown in Figure 2. This study collected three types of airborne hyperspectral imagery (HyMap, CASI-1500h, AisaFENIX 1K) with overlapping footprints covering the Bayan Obo area in China during 2023–2024. All imagery were collected between 10:00 and 14:30 local time under clear-sky conditions. The standard spectra of REE minerals were measured from the Bayan Obo REE deposit by using a field spectroradiometer (SR-6500; Table 1; Spectral Evolution, Inc., Haverhill, MA, USA). Table 1 lists the technical parameters of each type of hyperspectral data collected for this study.

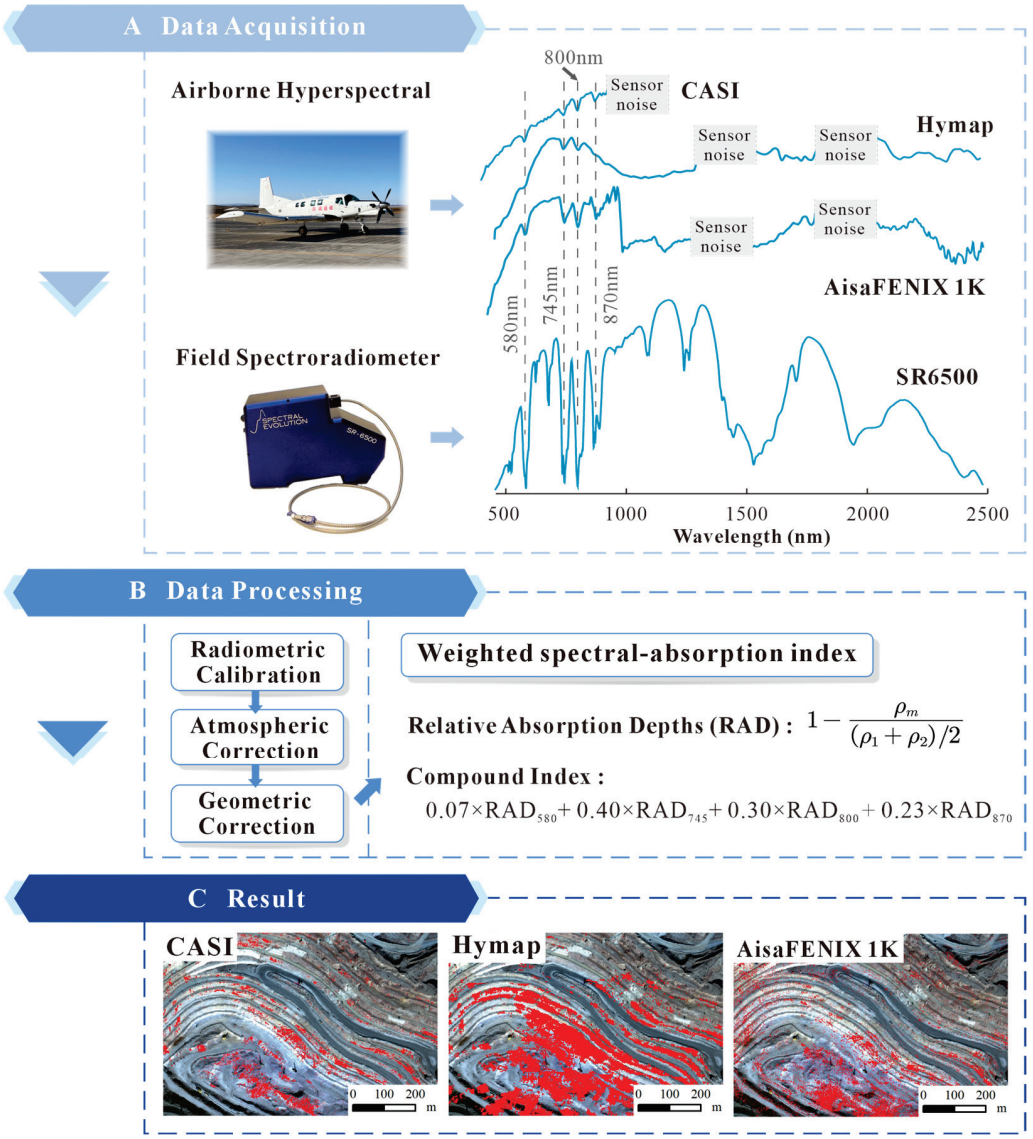


Figure 2. A generalized workflow illustrating the methods and tools employed in this study. (**A**) Airborne hyperspectral data were acquired by using Hymap, CASI-1500h, and AisaFENIX 1K sensors, while standard spectra of REE minerals were collected with a field spectroradiometer (SR6500); (**B**) Data processing includes data pre-processing (radiometric calibration, atmospheric correction, geometric correction), and REE-related information detecting; (**C**) The REE identification results (Red pixels).

Table 1. Technical parameters of each piece of hyperspectral data we collected.

	Airborne Hyperspectral			**Field Spectroradiometer**
	HyMap	**CASI-1500h**	**AisaFENIX 1K**	**SR-6500**
Spatial resolution	2.5 m	1.25 m	1 m	--
Spectral range	450–2470 nm	380–1050 nm	380–2500 nm	350–2500 nm
Spectral resolution	VNIR: 15 nm SWIR: 15–20 nm	VNIR: 7nm	VNIR: 3.5 nm SWIR: 6 nm	1.5 nm@700 nm 3.0 nm@1500 nm 3.8 nm@2100 nm
Signal-to-noise ratio	VNIR: ≥1000:1 SWIR: ≥600:1	--	VNIR:600–1000:1 SWIR:1250:1	--
Country	Australia	Canada	Finland	America

The Hyperspectral Mapper (HyMap) captures 144 wavebands covering the 450–2470 nm spectral region, with a spectral resolution of 15 nm in the VNIR bands and 15–20 nm in the SWIR bands [30]. At a relative flight height of 1000 m, the spatial resolution was 2.5 m, with a swath width of 1025 m.

The Compact Airborne Spectrographic Imager-1500h (CASI-1500h) can acquire up to 288 spectral bands across the 380–1050 nm spectral range [31]. The data obtained for this study included 96 bands with an average spectral resolution of 7 nm, and the relative flight height is ~2500 m with 1.25 m spatial resolution.

The AisaFENIX 1K airborne hyperspectral sensor can respectively acquire 344, 172, or 86 bands in VNIR range, whose spectral resolution varies from 1.75 to 7 nm; in addition, the 275 bands are available in the SWIR spectral range with a 6 nm spectral resolution [32]. The data acquired in this study included 174 bands in VNIR and 275 bands in SWIR, with a spatial resolution of 1 m.

Radiometric calibration for the HyMap data was conducted by multiplying the acquired DN values by the laboratory-derived radiometric calibration coefficients. Geometric correction was performed through image-to-base-map matching using a high-resolution base map that had been previously validated with ground control point data, while atmospherically corrected to reflectance using HyCorr and ENVI 5.6 software.

3.2. REE Identification Method

The standard spectra of REE minerals show five narrow absorption peaks at approximately 580, 680, 745, 800, and 870 nm. Since spectral curve shape and absorption are two criteria for hyperspectral mineral identification and mapping [33–36], we proposed a weighted spectral absorption index (WSAI) method to identify the REE minerals.

The WSAI quantifies relative absorption depths (RADs) at 580, 745, 800, and 870 nm and combines them into a compound index (CI). The RAD for a given peak is computed as

$$\mathrm{RAD} = 1 - \frac{\rho_m}{(\rho_1 + \rho_2)/2} \tag{1}$$

In the equation, ρ_m is the reflectance at specified peak m (e.g., ρ_{580}), and ρ_1 and ρ_2 are the reflectance on the left and right shoulders of peak m.

CI is subsequently defined by

$$\mathrm{CI} = 0.07 \times \mathrm{RAD}_{580} + 0.40 \times \mathrm{RAD}_{745} + 0.30 \times \mathrm{RAD}_{800} + 0.23 \times \mathrm{RAD}_{870} \tag{2}$$

where RAD_{580}, RAD_{745}, RAD_{800}, and RAD_{870} are the RADs at wavelengths of 580 nm, 745 nm, 800 nm, and 870 nm, respectively. The specific position can be fine-tuned according

to the sensor's band settings, such that the reflectance value at this position is minimized. The weight coefficients were determined based on the ratios among the absorption depths at the four spectral characteristic peaks, combined with the experimental results.

All hyperspectral data processing and analysis techniques were implemented using the Environment for Visualizing Images software (ENVI 5.6). The mineral distribution maps were plotted using ArcGIS 10.8.1 (ESRI, Redlands, CA, USA).

4. Results and Validations

By applying the proposed REE identification method, we successfully identified the signatures of REEs in carbonatites and carbonatite dikes, fenite-syenites, metamorphic feldspar-quartz sandstones, and even diluvial sediments. Notably, none of the airborne hyperspectral sensors can discriminate REE mineral species (e.g., monazite, bastnäsite, and parisite), as their diagnostic absorption features at 580, 745, 800, and 870 nm exhibit subtle spectral differences of only 1–7 nm, approaching the spectral resolution of hyperspectral sensors. In the future, with the improvement in the spectral resolution of airborne hyperspectral sensors, it may be possible to discriminate REE mineral species according to their precise characteristic peak positions. For example, bastnäsite exhibits absorption peaks at 864 nm, while monazite exhibits them at 871 nm.

4.1. Carbonatite Dikes

Carbonatite dikes occur in the vicinity of the Bayan Obo deposit, intruding into the Bayan Obo Group and the crystalline basement. These carbonatite dikes generally strike NW or NE and are relatively small in scale, with typical lengths ranging from 10 m to 200 m and widths ranging from 0.5 to 2.0 m [26]. The REE-bearing carbonatite dikes discovered in this study (Figure 3a–c) are 0.2–1.0 m wide and 70–100m long (Figure 4a–c), with a total rare earth oxide (TREO) content ranging from 0.79 to 15.90 wt.%. The major minerals are calcite and dolomite, accompanied by subordinate amounts of monazite, bastnäsite (Figure 5a,b), barite, allanite, magnetite, and apatite.

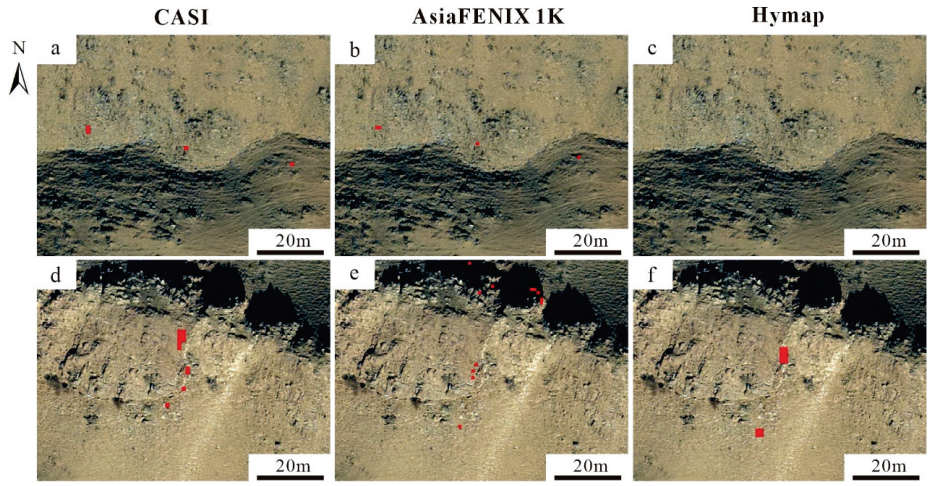


Figure 3. The REE identification results respectively obtained from three types of airborne hyperspectral sensors. Red pixels indicate the REE identification results. (**a**–**c**) Carbonatite dike: (**a**) CASI-1500h, (**b**) AisaFENIX 1K, (**c**) HyMap; (**d**–**f**) metamorphic feldspar-quartz sandstone: (**d**) CASI-1500h, (**e**) AisaFENIX 1K, (**f**) HyMap.

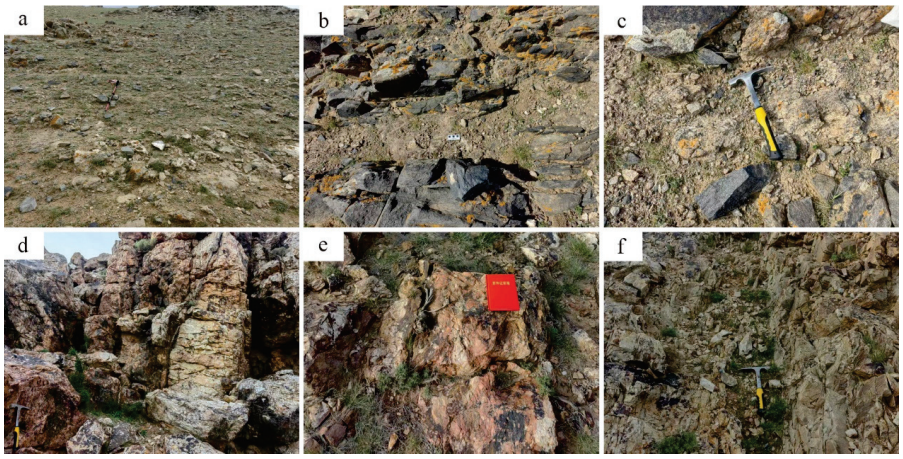
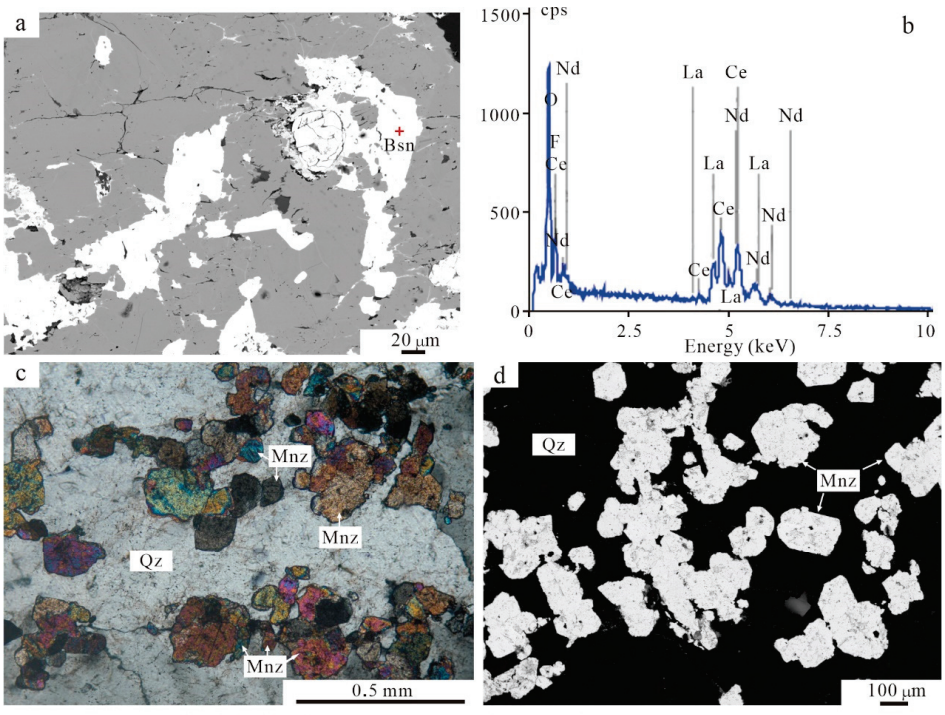


Figure 4. The outcrops of REE-bearing carbonatite dike and metamorphic feldspar-quartz sandstone. (**a**–**c**) Field photographs of REE-bearing carbonatite dyke; (**d**,**e**) the REE-bearing hydrothermal mineralization zone of metamorphosed feldspar-quartz sandstone, displaying intensive fracturing along with limonitization and silicification; (**f**) metamorphic feldspar-quartz sandstones.



e (1) Monazite EMPA Mean over 16 spots (in wt.%)
(2) Mixed sample

	(1)	(2)		(1)	(2)
La_2O_3	13.83	14.09	P_2O_5	28.09	29.98
Ce_2O_3	31.31	33.72	SiO_2	0.42	—
Pr_2O_3	7.16	7.43	Al_2O_3	0.06	—
Nd_2O_3	16.27	14.78	CaO	0.06	—
Gd_2O_3	1.26	—	TiO_2	< LOD	—
Er_2O_3	0.02	—	FeO	0.01	—
Yb_2O_3	0.03	—	F	0.80	—
Y_2O_3	0.14	—	Cl	0.06	—
Nb_2O_3	0.03	—	$Total$	99.55	100

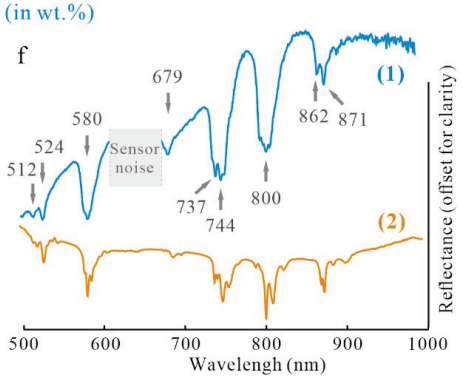


Figure 5. The microscopic characteristics of monazite and bastnäsite. (**a**) BSE image of bastnäsite; (**b**) EDS of bastnäsite; (**c**) microphotograph of monazite; (**d**) BSE image of monazite; (**e**) compositions of monazite and mixed sample; (**f**) spectra of monazite and mixed sample. Abbreviations of minerals: Mnz, monazite; Bsn, Bastnäsite; Qz, quartz.

https://doi.org/10.3390/rs18081110

These dikes are genetically related to REE mineralization [37] and are typically associated with intense fenitization. Fenitization, a characteristic alteration related to REE mineralization, is marked by an assemblage of aegirine, amphibole, phlogopite, and albite in the contact zones with the host rocks [38]. Furthermore, these carbonatite dikes exhibit REE contents ranging from 0.02 to 20.00 wt.% [26,28], supporting a genetic association with mineralization.

4.2. Metamorphic Feldspar-Quartz Sandstones

The monazite information identified by hyperspectral imagery (Figure 3d–f) is located within the hydrothermal mineralization zone of metamorphosed feldspar-quartz sandstone in the Dulahala Formation of the Bayan Obo Group. The REE-bearing mineralization zone, approximately 2–3 m in width (Figure 4d,e), exhibits intense fracturing along with limonitization and silicification. Geochemical analyses revealed TREO content ranging from 0.13 to 9.97 wt.%, dominated by light REEs such as La, Ce, Pr, and Nd.

Petrographic analysis confirmed the field observations, indicating that the rocks are predominantly composed of quartz and feldspar, with a subordinate amounts of monazite and iron oxides. Quartz and feldspar occur as anhedral grains, with feldspar surfaces exhibiting slight weathering. Monazite forms subhedral to euhedral crystals (Figure 5c,d), occurring as band-like aggregates on feldspar and quartz. The crystalline surfaces of monazite are generally clean, with some grains showing twinning and zoning textures, suggesting a hydrothermal origin. Iron oxides, predominantly magnetite, show signs of hydrothermal alteration such as hematitization and limonitization. Their coexistence with monazite further supports a hydrothermal genesis. It is inferred that REE-bearing hydrothermal fluids migrated upward along fractures and precipitated monazite within the metamorphosed feldspar-quartz sandstone.

REEs in the collected samples are primarily hosted in phosphates (monazite), fluorocarbonates (bastnäsite), and silicates (chevkinite-(Ce), fluorbritholite-(Ce)). All samples exhibited higher abundances of LREEs than HREEs. Monazite is the most representative REE mineral among these. Based on electron microprobe analysis (EPMA) results, the compositions of monazites from carbonatite dikes and metamorphic feldspar-quartz sandstones are dominated by Ce_2O_3 and P_2O_5, followed by Nd_2O_3, La_2O_3 and Pr_2O_3. They also contain minor amounts of heavy REEs (Gd_2O_3, Er_2O_3, Yb_2O_3, Y_2O_3) and impurity elements (Figure 5e).

5. Discussion

5.1. Method Evaluation

Based on the above findings, all airborne hyperspectral sensors (HyMap, CASI-1500h, AisaFENIX 1K) are capable of identifying REE-bearing minerals regardless of the host rock types by using WSAI. However, HyMap has an optimal spatial resolution of up to 2 m, which imposes constraints on its application in the exploration of REE deposits. In this study, the outcrop width of the REE-bearing carbonatite dikes ranges from 0.2 to 1 m, and the HyMap sensor fails to detect REE signals (Figure 3c). We employed the number of true positive and false positive pixels to evaluate the performance of each airborne hyperspectral sensor. The corresponding results showed (Figures 3 and 4; Table 2) that the CASI-1500h airborne hyperspectral sensor demonstrates the best performance in REE identification.

Owing to the multiple narrow and deep absorption peaks of REEs, they are easily confused with noise. Detecting REE-related information using spectral matching methods within the spectral range of 400–950 nm (e.g., spectral angle method) tends to cause false identification. The proposed WSAI method can alleviate this problem. Though the WSAI

method can effectively detect REEs, several factors, such as shadows and anthropogenic features (e.g., buildings and roads), may compromise classification accuracy. For example, shadowed regions exhibit weak reflectance signals and consequently high noise levels, which may be misclassified as REE signals. This issue can be mitigated by applying a reflectance threshold—values falling below this threshold are excluded from analysis. The threshold value was determined by analyzing the reflectance values of shadow areas in the hyperspectral image, and pixels with reflectance values below 0.05 were classified as shadow areas and discarded in this study.

Table 2. The number of true/false positive pixels.

Airborne Hyperspectral	Validation	Total Pixels	True Positive Pixels	False Positive Pixels
HyMap	Carbonatite dikes	1008	0	0
	Metamorphic feldspar-quartz sandstones	1036	3	0
CASI-1500h	Carbonatite dikes	4015	4	0
	Metamorphic feldspar-quartz sandstones	4015	12	0
AisaFENIX 1K	Carbonatite dikes	6188	4	0
	Metamorphic feldspar-quartz sandstones	6188	3	9

Additionally, the weighting approach may cause absorption features with small weight coefficients to be overlooked. Therefore, a double verification of characteristic peaks at 800 nm was conducted to prevent anthropogenic features from being erroneously identified as REEs. It should be noted that the shadow threshold and characteristic peak verification may vary depending on the specific study area. In subsequent research, we will investigate superior methods for precise REE identification.

5.2. Reflectance Spectra of REE Minerals

The absorption features of reflectance spectra are the result of electronic processes (such as crystalline fields, charge transfer absorptions, conduction bands, and color centers) and vibrational processes (resulting from the vibration of molecular bonds and ligands), which are diagnostic for the identification of minerals [39,40]. Monazite ((Ce, La, Y, Th)PO_4), a phosphate mineral, is one of the typical REE minerals enriched in light REEs [3]. The crystal structure of monazite is characterized by the REE^{3+} ions at a ninefold-coordinated site [11,41].

The spectra of monazite within the metamorphic feldspar-quartz sandstone show stronger sharp absorptions at 524, 580, 673–679, 743–745, 800, and 870–873 nm and three weaker absorptions at 511–513, 737, and 862–864 nm (Figure 5f). These spectral features are similar to those of Nd_2O_3 (Figure 6).

The elements La^{3+} to Lu^{3+} have 0–14 electrons in the 4f orbital [42]. The 4f–4f orbital intra-configurational electronic transitions cause REE^{3+} ions to have numerous sharp absorption features in the VNIR–SWIR spectra (400–2500 nm) [43], which are consistent with our experimental measurements (Figure 6). The partial filling of the 4*f* electron shell combined with a shielding effect caused by the fully filled $5s^2 5p^6$-electron shells makes lanthanide ions and their absorptions unique [44]. Since La^{3+} and Lu^{3+} are empty or filled in the 4*f* orbital, these elements do not have free electrons to generate absorption [11]. In

addition, Y^{3+} has the same electronic configuration as the noble gas krypton, which gives it particular stability [45], with no absorption features.

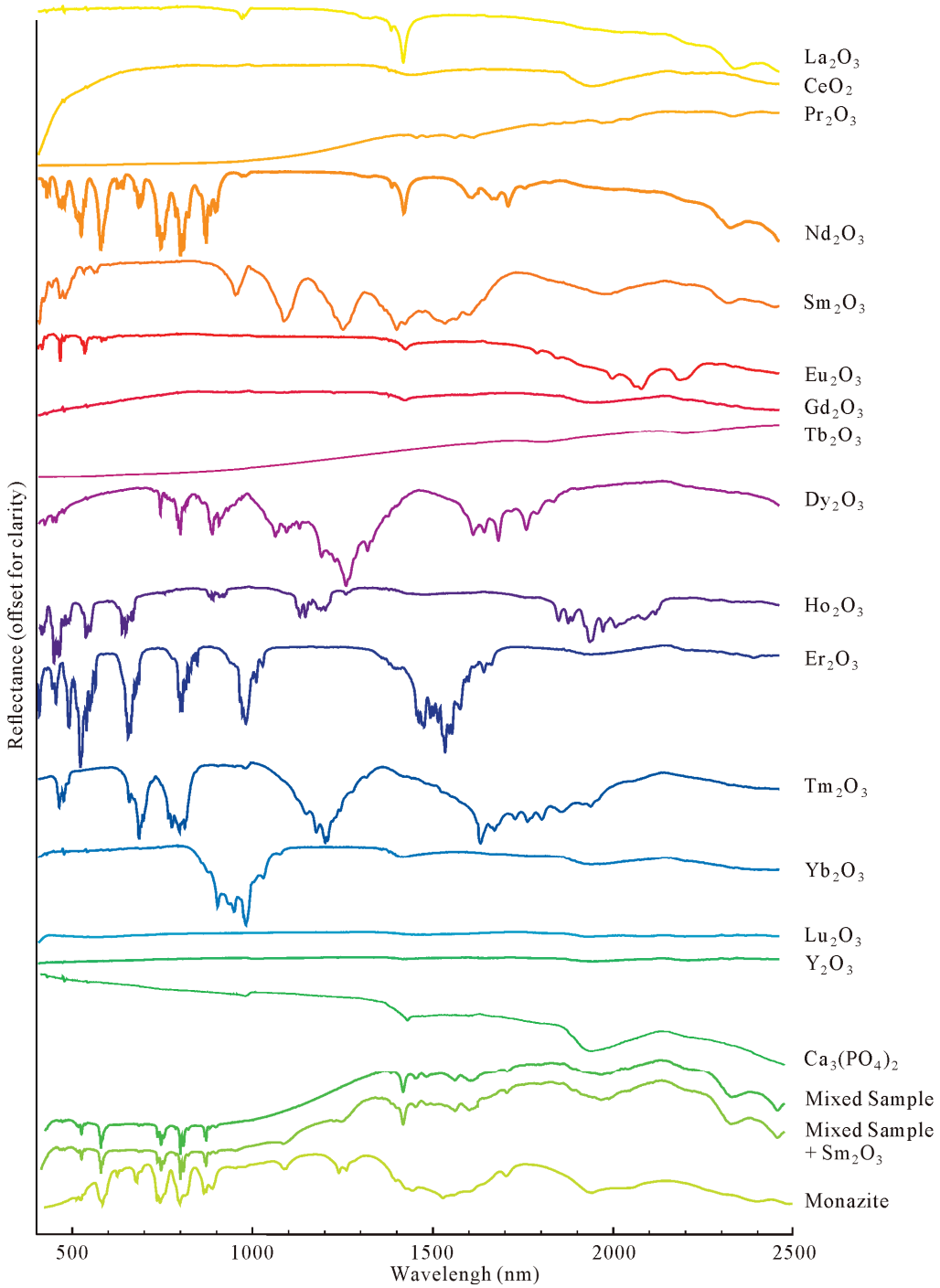


Figure 6. Spectra of single-element lanthanide oxides and other samples.

The EMPA results reveal that the principal constituents of monazite samples are Ce^{3+}, Nd^{3+}, La^{3+}, Pr^{3+}, and P^{5+} (Figure 5e). In most samples, Yb^{3+} and Er^{3+} contents are below the detection limit, while the total content of the remaining untested REE_2O_3 did not exceed 1 wt.%. These concentrations are too low to have a significant influence on the spectra of monazite. In addition, Ce^{3+} and PO_4^{3-} ($Ca_3(PO_4)_2$) show a wide absorption band centered at 1942 nm and 1967 nm, respectively, with no absorption in the range of 400–1000 nm (Figure 6). Pr^{3+} also has absorption features at 1969 nm and 2040 nm (Figure 6). Meanwhile,

Ce^{3+}, PO_4^{3-}, and Pr^{3+} are likely responsible for the wide absorption features of monazite at approximately 1970 nm.

Based on the EPMA results of monazite, La_2O_3, Ce_2O_3, Pr_2O_3, Nd_2O_3, and $Ca_3(PO_4)_2$ (all in 200-mesh powder) were mixed at a mass ratio of 14.09%, 33.72%, 7.43%, 14.78%, and 29.98%, respectively (Figure 5e). The spectra of the mixed sample exhibited strong sharp absorptions at 525, 579, 746, 800, and 872 nm and weaker absorptions at 478, 683, and 808–868 nm within the wavelength range of 400–1000 nm. These spectral characteristics are consistent with those of monazite. Nevertheless, the spectra of the mixed sample show prominent absorptions at 1418, 1564, 1609, and 2332 nm, which differ from those of monazite. Upon adding ~1% Sm_2O_3 (200-mesh powder) to the mixed sample, weak absorptions appeared at 1083 and 1252 nm, with no significant changes observed in other spectral features. Therefore, it is inferred that the spectrum of monazite is influenced not only by its REE composition but also potentially by its crystal structure; further in-depth investigations are warranted in subsequent studies.

6. Conclusions

Overall, this study confirmed that airborne hyperspectral imagery can effectively identify REE information. Among the three types of airborne hyperspectral sensors (HyMap, CASI-1500h, and AisaFENIX 1K), the CASI-1500h sensor is considered the most effective for this purpose. Additionally, the analytical approach introduced here provides a hyperspectral processing workflow that can be used to identify REE information, regardless of whether the REEs occur in carbonatite dikes, fenite-syenites, or metamorphic feldspar-quartz sandstones. Laboratory-based spectroscopy and mineral chemistry analyses indicated that the absorption features of the REE-bearing mineral monazite can be attributed to Nd^{3+} in the 400–1000 nm range.

Furthermore, the application of airborne hyperspectral remote sensing imagery has proven effective for significant effectiveness in the exploration of critical mineral resources, especially the REEs. This technology is expected to enhance exploration efficiency, thereby offering considerable savings in time, financial investment, and field effort.

Author Contributions: C.L.: writing—original draft preparation, methodology, investigation; J.Q. and J.Y. (Junchuan Yu): writing—review and editing, methodology; Y.X.: writing—original draft preparation, validation; Y.Z.: project administration.; X.Z., Q.M. and G.L.: validation; R.X.: data curation; B.C.: funding acquisition; J.Y. (Jinzhong Yang): supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Geological Exploration Foundation of Inner Mongolia of China (Grant No. 2023-JC05).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: We sincerely thank Zi-hong Gao, Ya-chao Han, Zhi-yang Nie and others for their invaluable assistance in acquiring the precious airborne hyperspectral data. Special thanks are due to the managements and staff of Inner Mongolia Geological Survey Institute, Baotou Iron and Steel Ltd., and Inner Mongolia Land and Resources Exploration and Development Co., Ltd. for their assistance during fieldwork. Furthermore, we would like to thank the anonymous reviewers for their constructive comments that helped to improve the manuscript.

Conflicts of Interest: Authors Yanbo Zhao and Xin Zhang were employed by the company Inner Mongolia Land and Resources Exploration and Development Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Goodenough, K.M.; Wall, F.; Merriman, D. The Rare Earth Elements: Demand, Global Resources, and Challenges for Resourcing Future Generations. *Nat. Resour. Res.* **2018**, *27*, 201–216. [CrossRef]
2. Balaram, V. Rare Earth Elements: A Review of Applications, Occurrence, Exploration, Analysis, Recycling, and Environmental Impact. *Geosci. Front.* **2019**, *10*, 1285–1303. [CrossRef]
3. Zepf, V. *Rare Earth Elements: A New Approach to the Nexus of Supply, Demand and Use: Exemplified Along the Use of Neodymium in Permanent Magnets*; Springer: Berlin/Heidelberg, Germany, 2013; ISBN 978-3-642-35457-1.
4. Voncken, J.H.L. *The Rare Earth Elements*; Springer Briefs in Earth Sciences; Springer International Publishing: Cham, Switzerland, 2016; ISBN 978-3-319-26807-1.
5. Chen, W.-Y.; Xue, G.-Q.; Xu, Y.; Wang, J.; Lv, P.-F.; Wu, X.; Song, W.-T.; Fan, H.-R.; Zhao, Y.-G.; Li, X.-H. Geophysical Exploration of the Giant Bayan Obo REE–Nb–Fe Deposit in Inner Mongolia, China: Progress and Challenges. *Ore Geol. Rev.* **2025**, *176*, 106415. [CrossRef]
6. Denton, K.M.; Ponce, D.A.; Peacock, J.R.; Miller, D.M. Geophysical Characterization of a Proterozoic REE Terrane at Mountain Pass, Eastern Mojave Desert, California, USA. *Geosphere* **2020**, *16*, 456–471. [CrossRef]
7. Taylor, R.D.; Shah, A.K.; Walsh, G.J.; Taylor, C.D. Geochemistry and Geophysics of Iron Oxide-Apatite Deposits and Associated Waste Piles with Implications for Potential Rare Earth Element Resources from Ore and Historical Mine Waste in the Eastern Adirondack Highlands, New York, USA. *Econ. Geol.* **2019**, *114*, 1569–1598. [CrossRef]
8. Kruse, F.A.; Boardman, J.W.; Huntington, J.F. Comparison of Airborne Hyperspectral Data and Eo-1 Hyperion for Mineral Mapping. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1388–1400. [CrossRef]
9. Van Der Meer, F.D.; Van Der Werff, H.M.A.; Van Ruitenbeek, F.J.A.; Hecker, C.A.; Bakker, W.H.; Noomen, M.F.; Van Der Meijde, M.; Carranza, E.J.M.; Smeth, J.B.D.; Woldai, T. Multi- and Hyperspectral Geologic Remote Sensing: A Review. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *14*, 112–128. [CrossRef]
10. Turner, D.J.; Rivard, B.; Groat, L.A. Visible and Short-Wave Infrared Reflectance Spectroscopy of REE Fluorocarbonates. *Am. Mineral.* **2014**, *99*, 1335–1346. [CrossRef]
11. Turner, D.J.; Rivard, B.; Groat, L.A. Visible and Short-Wave Infrared Reflectance Spectroscopy of REE Phosphate Minerals. *Am. Mineral.* **2016**, *101*, 2264–2278. [CrossRef]
12. Lorenz, S.; Beyer, J.; Fuchs, M.; Seidel, P.; Turner, D.; Heitmann, J.; Gloaguen, R. The Potential of Reflectance and Laser Induced Luminescence Spectroscopy for Near-Field Rare Earth Element Detection in Mineral Exploration. *Remote Sens.* **2018**, *11*, 21. [CrossRef]
13. Qasim, M.; Khan, S.D. Detection and Relative Quantification of Neodymium in Sillai Patti Carbonatite Using Decision Tree Classification of the Hyperspectral Data. *Sensors* **2022**, *22*, 7537. [CrossRef] [PubMed]
14. Boesche, N.; Rogass, C.; Lubitz, C.; Brell, M.; Herrmann, S.; Mielke, C.; Tonn, S.; Appelt, O.; Altenberger, U.; Kaufmann, H. Hyperspectral REE (Rare Earth Element) Mapping of Outcrops—Applications for Neodymium Detection. *Remote Sens.* **2015**, *7*, 5160–5186. [CrossRef]
15. Asadzadeh, S.; Koellner, N.; Chabrillat, S. Detecting Rare Earth Elements Using EnMAP Hyperspectral Satellite Data: A Case Study from Mountain Pass, California. *Sci. Rep.* **2024**, *14*, 20766. [CrossRef] [PubMed]
16. Imran, S.; Goswami, A.; Kumar, H.; Barman, B.J.; Saikia, A. Space and Ground Based Spectroscopic Studies and Mineral Chemistry of Rare Earth Element Bearing Peralkaline Rocks from Siwana Ring Complex, Rajasthan, India. *Remote Sens. Appl. Soc. Environ.* **2024**, *36*, 101383. [CrossRef]
17. Gadea, O.C.A.; Khan, S.D.; Sisson, V.B. Estimating Rare Earth Elements at Various Scales with Bastnäsite Indices for Mountain Pass. *Ore Geol. Rev.* **2024**, *173*, 106254. [CrossRef]
18. Booysen, R.; Jackisch, R.; Lorenz, S.; Zimmermann, R.; Kirsch, M.; Nex, P.A.M.; Gloaguen, R. Detection of REEs with Lightweight UAV-Based Hyperspectral Imaging. *Sci. Rep.* **2020**, *10*, 17450. [CrossRef]
19. Qasim, M.; Khan, S.D.; Sisson, V.; Greer, P.; Xia, L.; Okyay, U.; Franco, N. Identifying Rare Earth Elements Using a Tripod and Drone-Mounted Hyperspectral Camera: A Case Study of the Mountain Pass Birthday Stock and Sulphide Queen Mine Pit, California. *Remote Sens.* **2024**, *16*, 3353. [CrossRef]
20. Chakraborty, R.; Kereszturi, G.; Pullanagari, R.; Durance, P.; Ashraf, S.; Anderson, C. Mineral Prospecting from Biogeochemical and Geological Information Using Hyperspectral Remote Sensing—Feasibility and Challenges. *J. Geochem. Explor.* **2022**, *232*, 106900. [CrossRef]
21. Liang, S.; Sun, S.; Lu, H. Application of Airborne Electromagnetics and Magnetics for Mineral Exploration in the Baishiquan–Hongliujing Area, Northwest China. *Remote Sens.* **2021**, *13*, 903. [CrossRef]
22. Mishra, G.; Govil, H.; Srivastava, P.K. Identification of Malachite and Alteration Minerals Using Airborne AVIRIS-NG Hyperspectral Data. *Quat. Sci. Adv.* **2021**, *4*, 100036. [CrossRef]
23. Perikleous, D.; Margariti, K.; Velanas, P.; Blazquez, C.S.; Gonzalez-Aguilera, D. Aerial Drones for Geophysical Prospection in Mining: A Review. *Drones* **2025**, *9*, 383. [CrossRef]

24. Yang, K.; Fan, H.; Pirajno, F.; Li, X. The Bayan Obo (China) Giant REE Accumulation Conundrum Elucidated by Intense Magmatic Differentiation of Carbonatite. *Geology* **2019**, *47*, 1198–1202. [CrossRef]
25. Yang, J.; Song, W.; Liu, Y.; Zhu, X.; Kynicky, J.; Chen, Q. Mineralogy and Element Geochemistry of the Bayan Obo (China) Carbonatite Dykes: Implications for REE Mineralization. *Ore Geol. Rev.* **2024**, *165*, 105873. [CrossRef]
26. Fan, H.-R.; Yang, K.-F.; Hu, F.-F.; Liu, S.; Wang, K.-Y. The Giant Bayan Obo REE-Nb-Fe Deposit, China: Controversy and Ore Genesis. *Geosci. Front.* **2016**, *7*, 335–344. [CrossRef]
27. She, H.-D.; Fan, H.-R.; Yang, K.-F.; Li, X.-C.; Wang, Q.-W.; Zhang, L.-F.; Liu, S.; Li, X.-H.; Dai, Z.-H. In Situ Trace Elements of Magnetite in the Bayan Obo REE-Nb-Fe Deposit: Implications for the Genesis of Mesoproterozoic Iron Mineralization. *Ore Geol. Rev.* **2021**, *139*, 104574. [CrossRef]
28. Yang, X.-M.; Yang, X.-Y.; Zheng, Y.-F.; Le Bas, M.J. A Rare Earth Element-Rich Carbonatite Dyke at Bayan Obo, Inner Mongolia, North China. *Mineral. Petrol.* **2003**, *78*, 93–110. [CrossRef]
29. Yang, K.-F.; Fan, H.-R.; Santosh, M.; Hu, F.-F.; Wang, K.-Y. Mesoproterozoic Mafic and Carbonatitic Dykes from the Northern Margin of the North China Craton: Implications for the Final Breakup of Columbia Supercontinent. *Tectonophysics* **2011**, *498*, 1–10. [CrossRef]
30. Cocks, T.; Jenssen, R.; Stewart, A.; Wilson, I.; Shields, T. The Hymap^{TM} airborne Hyperspectral Sensor: The System, Calibration and performance. In Proceedings of the 1st EARSEL Workshop on Imaging Spectroscopy, Zurich, Switzerland, 6–8 October 1998.
31. ITRES Research Limited. Available online: https://www.itres.com/wp-content/uploads/2019/09/CASI1500.pdf (accessed on 25 January 2025).
32. Specim, Specrtal Imaging Ltd. 2017. Available online: https://www.specim.com/products/aisafenix-1k/# (accessed on 25 January 2025).
33. Kruse, F.A.; Heidebrecht, K.B.; Shapiro, A.T.; Barloon, P.J.; Goetz, A.F.H. The Spectral Image Processing System (SIPS) Interactive Visualization and Analysis of Imaging Spectrometer Data. *Remote Sens. Environ.* **1993**, *44*, 145–163. [CrossRef]
34. Qiu, J.-T.; Zhang, C.; Xu, Q.-J.; Yao, J.-L. Mapping of Carnallite along with Semi-Quantitative Estimation of Potassium Content of Drill Cores Using Hyperspectral Imagery. *Remote Sens. Lett.* **2017**, *8*, 859–868. [CrossRef]
35. Qiu, J.-T.; Qi, H.; Duan, J.-L. Reflectance Spectroscopy Characteristics of Turquoise. *Minerals* **2016**, *7*, 3. [CrossRef]
36. Han, Y.; Lee, T.C.M. Uncertainty Quantification for Sparse Estimation of Spectral Lines. *IEEE Trans. Signal Process.* **2022**, *70*, 6243–6256. [CrossRef]
37. Le Bas, M.J.; Xueming, Y.; Taylor, R.N.; Spiro, B.; Milton, J.A.; Peishan, Z. New Evidence from a Calcite-Dolomite Carbonatite Dyke for the Magmatic Origin of the Massive Bayan Obo Ore-Bearing Dolomite Marble, Inner Mongolia, China. *Mineral. Petrol.* **2007**, *90*, 223–248. [CrossRef]
38. Le Bas, M.J.; Kellere, J.; Kejie, T.; Wall, F.; William, C.T.; Peishan, Z. Carbonatite Dykes at Bayan Obo, Inner Mongolia, China. *Mineral. Petrol.* **1992**, *46*, 195–228. [CrossRef]
39. Clark, R.N. *Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy; Remote Sensing for the Earth Sciences*; John Wiley and Sons: New York, NY, USA, 1999; p. 3.
40. Guha, A. Mineral Exploration Using Hyperspectral Data. In *Hyperspectral Remote Sensing*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 293–318. ISBN 978-0-08-102894-0.
41. Clavier, N.; Podor, R.; Dacheux, N. Crystal Chemistry of the Monazite Structure. *J. Eur. Ceram. Soc.* **2011**, *31*, 941–976. [CrossRef]
42. Hasegawa, Y.; Kitagawa, Y.; Shoji, S. *Lanthanide-Based Wavelength Conversion Materials*; Springer Nature: Singapore, 2024; ISBN 978-981-97-5635-3.
43. Turner, D.J. Reflectance Spectroscopy and Imaging Spectroscopy of Rare Earth Element-Bearing Mineral and Rock Samples. Ph.D. Thesis, University of British Columbia, Vancouver, BC, Canada, 2015.
44. Hoefen, T.M.; Livo, K.E.; Giles, S.A.; Swayze, G.A. Ultraviolet to Near-Infrared Spectroscopy of REE-Bearing Materials. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 3422–3425.
45. Jones, A.P.; Wall, F.; Williams, C.T. *Rare Earth Minerals: Chemistry, Origin and Ore Deposits*; The Mineralogical Society Series; Chapman and Hall: London, UK, 1996; ISBN 978-0-412-61030-1.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

remote sensing

MDPI

Article

Double-Attention Context Interactive Network for Hyperspectral Image Classification

Nannan Hu [1,3,†], Zhongao Wang [1,3,†], Minghao Wang [1,2] and Yuefeng Zhao [1,2,3,*]

[1] School of Communication and Electronic Engineering, Shandong Normal University, Jinan 250358, China
[2] Shandong Provincial Engineering and Technical Center of Light Manipulations & Shandong Provincial Key Laboratory of Optics and Photonic Device, School of Physics and Electronics, Shandong Normal University, Jinan 250358, China
[3] Shandong Key Laboratory of Medical Physics and Image Processing, School of Communication and Electronic Engineering, Shandong Normal University, Jinan 250358, China
* Correspondence: yuefengzhao@sdnu.edu.cn
† These authors contributed equally to this work.

Highlights

What are the main findings?

- We propose a Double-Attention Context Interactive Network (DACINet) that strengthens long-range contextual interaction and 3D spectral–spatial feature learning for hyperspectral image classification, via a Context Interaction Fusion Module (CIFM) and Channel–Spatial Double-Attention (CSDA).
- Extensive experiments on Indian Pines, Pavia University, and Salinas datasets demonstrate that the DACINet achieves superior classification accuracy and robustness compared to state-of-the-art methods.

What are the implications of the main findings?

- By explicitly modeling long-range context interaction across distant spectral bands and introducing Channel–Spatial Double-Attention, the framework alleviates the limitations of local receptive fields in conventional CNNs, yielding more discriminative and robust hyperspectral representations under limited training samples and class imbalance.
- By further coupling context-interactive feature fusion with a 2D–3D hybrid convolutional design, the proposed CNN-based solution enhances spectral–spatial integration and boundary/detail sensitivity, improving overall classification reliability in practical hyperspectral image classification settings.

Abstract

Convolution is still the main method for hyperspectral image classification, since it takes into account both spatial and spectral characteristics. However, the convolution relies on local perceptual computation, ignoring the effective discriminant of context association for classification. In this paper, we propose a Double-Attention Context Interactive Network (DACINet) for hyperspectral image classification. Specifically, a Context Interaction Fusion Module (CIFM) is designed to enhance long-range contextual dependencies. By stacking multiple 3D convolutional layers, the module progressively enlarges its receptive field, while cross-layer residual connections facilitate the integration of features from different contextual scales, thereby strengthening the model's ability to capture complex relationships within the hyperspectral data. Then, a Channel–Spatial Double-Attention (CSDA) mechanism based on 3D is proposed for enhancing the two-dimensional spatial features and one-dimensional spectral features, respectively, and fusing the enhanced features. Furthermore, we also construct a hybrid convolutional layer, which combines 2D and 3D convolution to further enhance spectral bands on the basis of three-dimensional

understanding. Extensive experiments on the widely used IP, UP, SA and HU datasets show that the proposed DACINet achieves superior classification accuracy, reaching Overall Accuracies of 96.78%, 97.77%, 99.53% and 86.67% respectively, outperforming other state-of-the-art models.

Keywords: hyperspectral image classification; context interactive; double-attention; hybrid convolutional layer

1. Introduction

Hyperspectral images (HSIs) are common 3D remote sensing data, which contain both spectral information and spatial information of surface objects. Hyperspectral image classification (HSIC) aims to assign a unique semantic label to each HSI pixel based on the given land cover category [1,2]. It is widely used in the fields of resource investigations, environmental monitoring and agricultural production [3]. In the early days, machine learning methods such as SVMs were commonly adopted to extract effective discriminant features from HSIs, and different classifiers were then used to classify each pixel [4].

Benefiting from advances in deep learning [5–7], hyperspectral image processing has achieved remarkable progress. CNN-based methods were the first to be introduced into hyperspectral image classification. Hu et al. [8] proposed a five-layer 1D-CNN, which better extracts the spectral features of hyperspectral images. Fang et al. [9] employed discriminative band selection for pipeline hyperspectral images to better extract spatial features using 2D convolution. Subsequently, many new frameworks achieved remarkable results in hyperspectral classification. He et al. [10] employed a hybrid Mamba-Transformer framework to explore the multiscale properties of hyperspectral data effectively. Liang et al. [11] achieved efficient global dependency modeling using a double-branch Mamba-like linear attention mechanism. Zhang et al. [12] proposed the Center-Scan Mamba network to model spatial–spectral long-range dependency with linear complexity. Although new architectures such as Transformers and Mamba demonstrate great potential in hyperspectral image classification, the CNN and its variants remain the most widely used and mature solution due to their strong feature extraction capability and low computational complexity.

Recently, CNN variants have once again become the research focus in hyperspectral image classification. Ahmad et al. [13] proposed a 3D CNN to leverage spectral–spatial feature maps to enhance HSIC performance. Ghaderizadeh et al. [14] proposed a hybrid 3D–2D CNN, which employs a 3D fast learning block followed by a 2D CNN to extract spectral–spatial features. Alkhatib et al. [15] classified HSIs using a multiscale 3D CNN and three-branch feature fusion. Gündüz et al. [16] introduced a dual attention mechanism into both spectral and spatial modules to enhance the model's discriminative ability. These variants achieve excellent classification results. However, 1D networks focus on spectral information, 2D networks focus on spatial information, and 3D networks are limited to local perception and ignore contextual correlations. This separation between spectral and spatial processing leads to insufficient global perception, making it difficult for the model to fully exploit the inherent advantages of the graphical unification. Recent efforts such as the Online Spectral Information Compensation Network [17] address this by extracting multiscale spatial features through a multi-branch network and progressively compensating spectral information. However, the OSICN primarily focuses on spatial scale diversity rather than explicitly modeling contextual interactions across spectral bands.

To alleviate the above problems, we propose a novel Double-Attention Context Interactive Network (DACINet) for modeling the contextual interaction of spectral and spatial

https://doi.org/10.3390/rs18071059

features simultaneously. The DACINet mainly consists of a Context Interaction Fusion Module (CIFM), Channel–Spatial Double-Attention (CSDA), and a 2D–3D hybrid convolutional layer. CIFM can capture long-range correlations between spectral bands by incorporating cross-layer residual connections into the 3D convolutional network. Then, the CSDA enhances two-dimensional spatial features and one-dimensional spectral features, respectively, and fuses the enhanced features. In addition, a hybrid convolution layer that combines 2D and 3D convolution is introduced to further enhance spectral bands based on 3D feature extraction. To validate the effectiveness and robustness of our proposed model, we conduct extensive experiments on four challenging benchmark datasets: Indian Pines (IP), Pavia University (UP), Salinas (SA), and the 2013 University of Houston (HU) dataset.

The main contributions of this paper are as follows:

- A novel DACINet is proposed to improve HSIC performance by enhancing contextual interaction and leveraging 3D spectral–spatial features.
- A CIFM is proposed to capture contextual interaction features, while CSDA is designed to suppress irrelevant information and to enhance spectral bands and spatial information.
- A hybrid convolution layer combining 2D and 3D convolutions is proposed to further strengthen the representation of spectral–spatial features.

2. Related Work

In this section, we review two topics most relevant to our work: deep learning-based hyperspectral image classification and attention-based hyperspectral image classification.

2.1. Deep Learning-Based Hyperspectral Image Classification

Convolutional Neural Networks (CNNs) [18], especially 3D CNNs [19,20], have become the main approach for hyperspectral image classification, owing to their robust feature extraction capabilities. Ghaderizadeh et al. [14] extracted spectral–spatial features using 3D convolutions and further optimized them through 2D convolutions. To address the limitation of fixed receptive fields in capturing long-range contextual dependencies, Roy et al. [21] introduced an adaptive receptive field 3D residual module, dynamically adjusting convolution kernels to enhance feature representation. To overcome the Euclidean space constraints of CNNs and enhance the modeling of non-Euclidean geometric relationships, Graph Neural Networks (GNNs) [22] have been introduced into hyperspectral classification. SSGRAM [23] proposed a 3D spectral–spatial feature network combined with a Graph Attention Feature Processor. NESSGGCN [24] constructed a Gated GCN-CNN Non-Euclidean Spectral–Spatial Feature Mining Network to simultaneously extract features from non-Euclidean and Euclidean spaces. Recently, various novel deep learning architectures have been proposed for HSI classification. Yang et al. [25] proposed an Enhanced Multiscale Feature Fusion Network (EMFFN) that extracts multiscale features through a three-stage parallel multi-path architecture. For cross-scene classification, Liu et al. [26] proposed a Dual Classification Head Self-Training Network (DHSNet) that alleviates domain discrepancies through dual-head consistency learning. Wang et al. presented HyperSIGMA [27], the first billion-parameter foundation model for HSIs, which unifies multiple interpretation tasks via a sparse sampling attention mechanism. Unsupervised clustering is a fundamental task for HSI analysis when labeled samples are unavailable. Zhang et al. [28] proposed Elastic Graph Fusion Subspace Clustering (EGFSC) for large-scale HSI clustering via superpixel-level learning and dual-graph fusion. Huang et al. [29] introduced a structural prior-guided subspace clustering method incorporating local/non-local spatial priors and cluster structure priors. Jiang et al. [30] proposed Structured Anchor Projected Clustering with linear time complexity through anchor generation and

dual-graph learning. These unsupervised methods provide powerful tools for mining the intrinsic structures of HSIs. The structural information captured by these methods, such as superpixel-level organization and graph-based relationships, offers valuable insights that can inform the design of feature extraction modules in supervised classification frameworks like our DACINet.

2.2. Attention-Based Hyperspectral Image Classification

Attention mechanisms [31] were initially proposed in the field of natural language processing (NLP). Following the success of the Vision Transformer (ViT) in image classification, Chen et al. [32] pioneered its application to HSI classification, demonstrating its potential for capturing global context. To better harness the complementary strengths of different architectures, hybrid attention designs have since emerged. Arshad et al. [33] introduced HAT to integrate the local representational capabilities of 3D and 2D CNNs within a Transformer framework. Zhao et al. [34] combined group-wise separable convolution with spectral calibration attention, effectively reorganizing and weighting features in the spectral dimension. Similarly, Jing et al. [35] introduced dynamic convolution and spatial–spectral attention mechanisms to dynamically extract and integrate multi-level semantic features. Although the aforementioned methods have made significant progress in the design and application of attention mechanisms, most approaches treat spectral and spatial attention modules as separate units, whether sequential or parallel, thus failing to establish a unified dual-path interactive framework. For instance, the DBDA network [36] applies channel-wise and spatial-wise attention in two independent branches. Although effective, this design inherently decouples spectral and spatial attention, limiting their interactive fusion. In contrast, the CSDA module proposed in this paper operates within a unified 3D feature space, enabling parallel enhancement and interactive fusion of spectral and spatial features. This interactive design distinguishes CSDA from existing decoupled attention mechanisms and forms the core innovation of the proposed DACINet.

3. Proposed Method

In this section, we present the proposed Double-Attention Context Interactive Network (DACINet) in detail.

The diagram of the DACINet is illustrated in Figure 1. A hyperspectral image captures tens to hundreds of continuous spectral bands of a target region simultaneously, with each band containing a large amount of pixel information. In the DACINet, we first apply PCA to project the high-dimensional data into a low-dimensional space, which reduces dimensionality and eliminates redundant information. The reduced-dimensionality features are then input into the CIFM to model contextual interactions between spectral and spatial features simultaneously. Next, CSDA enhances spatial and spectral context interaction features, respectively, to retain more discriminative information. A hybrid convolutional layer combining 2D and 3D convolutions further enhances the discriminative power of spectral features. Finally, the resulting feature maps are classified by a fully connected layer.

https://doi.org/10.3390/rs18071059

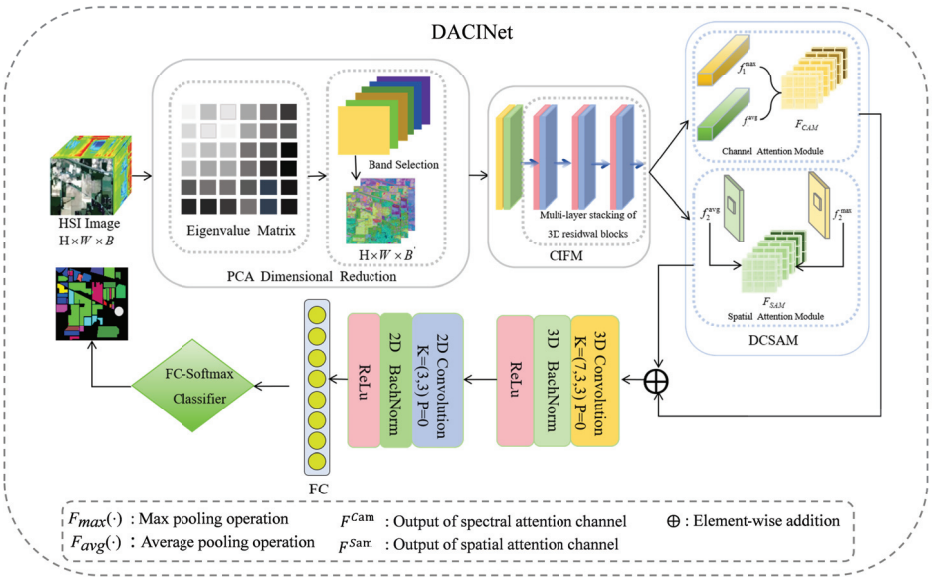


Figure 1. The structure of the proposed DACINet. The process begins with PCA for dimensionality reduction of the input HSI cube. The resulting patches are fed into the CIFM, which employs stacked 3D convolutions with cross-layer residual connections to capture long-range contextual information. The feature maps then enter the CSDA, which enhances channel (spectral) and spatial features in parallel before fusing them. Subsequently, a hybrid convolutional layer, comprising sequential 3D and 2D convolutions, further refines the spectral–spatial representation. Finally, a fully connected layer produces the pixel-wise classification results.

3.1. PCA

In order to solve the problem of hyperspectral image data redundancy, we still employ principal component analysis (PCA) to reduce the dimensionality of the hyperspectral data. The mapping process of PCA is as follows:

$$\begin{bmatrix} y_1^i \\ y_2^i \\ \vdots \\ y_n^i \end{bmatrix} = \begin{bmatrix} u_1^T \cdot (x_1^i, x_2^i, ..., x_n^i) \\ u_2^T \cdot (x_1^i, x_2^i, ..., x_n^i) \\ \vdots \\ u_n^T \cdot (x_1^i, x_2^i, ..., x_n^i) \end{bmatrix}, \tag{1}$$

where X^i is hyperspectral images, and $X^i = (x_1^i, x_2^i, ..., x_n^i)^T \in \mathcal{R}^{\mathcal{H}\times\mathcal{W}\times\mathcal{B}}$. u_i^T is the corresponding eigenvector. The feature after dimensionality reduction is represented as $Y^i = (y_1^i, y_2^i, ..., y_n^i)^T \in \mathcal{R}^{\mathcal{H}\times\mathcal{W}\times\mathcal{B}'}$. The PCA maps high-dimensional data to low-dimensional space through some linear projection, maximizing the retention of effective discriminative information in the projected dimension.

3.2. The Context Interaction Fusion Module (CIFM)

In hyperspectral images, homogeneous regions with different semantic and geometric properties but extremely similar appearance pose a significant challenge to the classification process. The long-distance contextual relationship between the image bands can provide effective discriminative constraints for such regions. To capture this contextual information, we design a Context Interaction Fusion Module (CIFM) consisting of stacked 3D convolutional residual blocks with cross-layer connections, as shown in Figure 2. While each individual block follows the standard residual 3D CNN design, the CIFM as a whole serves as a dedicated context encoder strategically positioned before the CSDA module to provide richly contextualized features for subsequent dual attention enhancement.

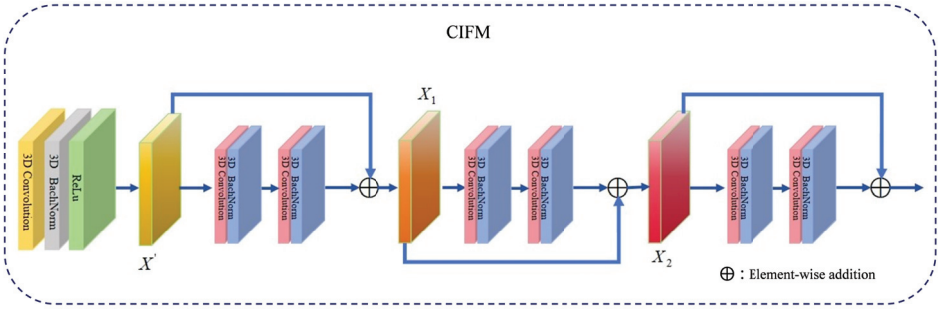


Figure 2. The structure of the CIFM.

In the CIFM, long-distance correlations between hyperspectral image bands are modeled through cross-layer residual connections. The input feature $Y \in \mathcal{R}^{\mathcal{H}\times W\times B'}$ is processed by 3D convolution, and the resulting feature is:

$$Y_j^l = 3DCov(Y) = W_{l+1}^{7\times3\times3} * Y + b^{l+1}, \tag{2}$$

$$Y' = R(\sum_{j=1}^{x^l} F_{bn}(Y_j^l) * W_i^{l+1} + b_i^{l+1}), \tag{3}$$

$$F_{bn}(Y_j^l) = \frac{Y_j^l - E(Y_j^l)}{\sqrt{Var^2(Y_j^l) + \epsilon}} \times \gamma + \beta, \tag{4}$$

where Y_j^l is the output features of the convolution layer, $W_{l+1}^{7\times3\times3}$ is the learnable weight with kernel size $(7 \times 3 \times 3)$, and b^{l+1} is the bias. $R(\cdot)$ represents the non-linear activation function ReLU. $F_{bn}(Y_j^l)$ is regularization, $E(Y_j^l)$ and $Var^2(Y_j^l)$ mean the batch mean and variance in input features, and ϵ, γ, β represent the stability constant, scaling factor, and offset, respectively. A double-layer 3D convolution is set in the CIFM, since the features are processed again according to Equations (2)–(4), and output features after two weighted layers are labeled as Y_j^{l+1}, and then two additional rounds of weight processing are performed:

$$Y_j^l = W_{l+1}^{1\times1\times1} * Y' + b^{l+1}, \tag{5}$$

$$Y_j^{l+1} = \sum_{j=1}^{x^l} (F_{bn}(Y_j^l) * W_i^{l+1} + b_i^{l+1}), \tag{6}$$

where Y_j^{l+1} are the output features after the two consecutive layers of weights. It further joins through cross-layer residuals:

$$Y_1 = Y + Y_j^{l+1}, \tag{7}$$

where Y_1 is the output of the first 3D residual block. The final output feature Y_f of the CIFM is obtained after f times of the same superimposed residual block processing. These cross-layer connections capture information associations over long distances, enhancing the context modeling capabilities of 3D convolution.

3.3. The Channel–Spatial Double-Attention (CSDA)

Hyperspectral images contain both spatial information and spectral band information. Effective integration of these two modalities can significantly improve classification performance. Considering the 3D spectral–spatial characteristics of hyperspectral data, we propose a Channel–Spatial Double-Attention (CSDA) mechanism based on pooling and feature fusion, as shown in Figure 3.

CSDA consists of two parallel branches: a channel attention mechanism and a spatial attention mechanism. Different from the original CBAM, which processes these two attentions sequentially, our CSDA processes them independently and in parallel.

Given the input feature map Y_f from the CIFM, the channel and spatial attention branches are computed as follows:

$$f_1^{max} = F_{max}(Y_f), f_1^{avg} = F_{avg}(Y_f), \tag{8}$$

$$f_2^{max} = Matr(Max(Y_f)), f_2^{avg} = Matr(Mean(Y_f)), \tag{9}$$

where f_1^{max}, f_1^{avg}, f_2^{max}, and f_2^{avg} denote the pooled feature maps from each operation. $F_{max}(\cdot)$ and $F_{avg}(\cdot)$ are the maximum pooling and global average pooling operations. *Matr* denotes a matrix transformation operation.

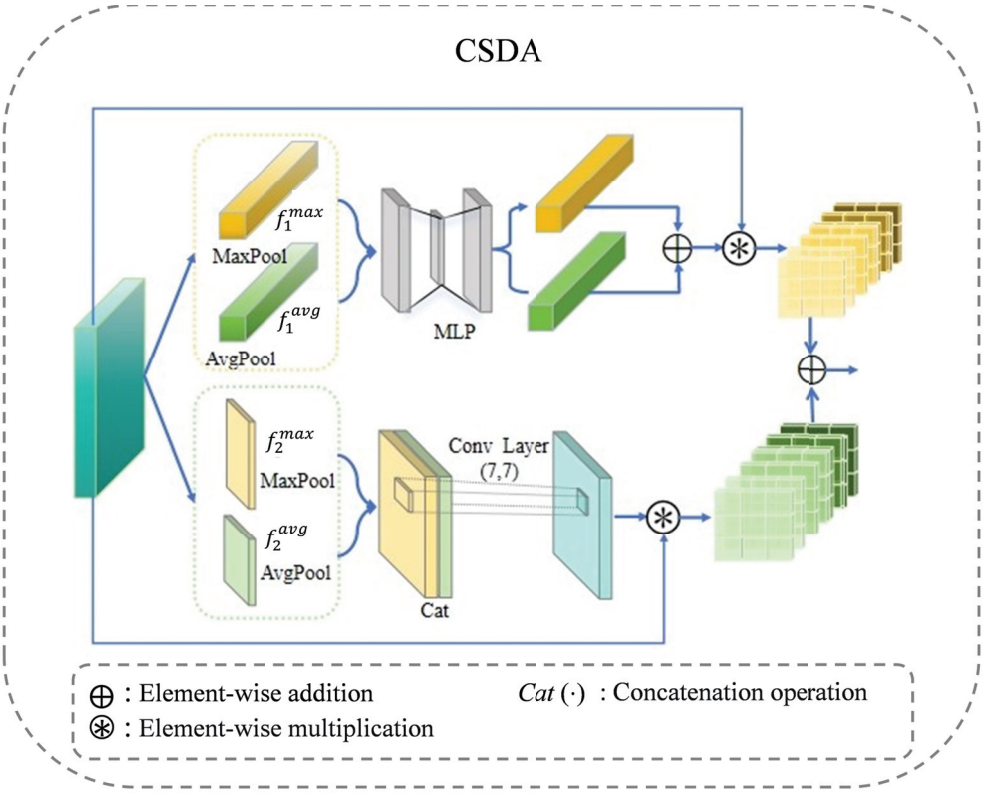


Figure 3. The structure of CSDA.

Then, f_1^{max} and f_1^{avg} are passed through an MLP for feature fusion:

$$\begin{aligned} Z_1^{max} &= W_{1\times1}^{max} * f_1^{max} + b^{l+1} \\ Z_1^{avg} &= W_{1\times1}^{avg} * f_1^{avg} + b^{l+1}. \\ Z &= Z_1^{max} + Z_1^{avg} \end{aligned} \tag{10}$$

The key to explicitly modeling spectral inter-band correlations lies within this shared MLP. Since its layers are fully connected, the MLP learns complex, non-linear interdependencies between all spectral bands. By considering the context provided by all bands simultaneously, it determines the relative importance of each band and produces an attention map Z that encodes these cross-band relationships.

https://doi.org/10.3390/rs18071059

For the spectral channel, the output features are obtained by element-wise multiplication of the attention map with the input:

$$F^{Cam} = Z \bigotimes Y_f. \tag{11}$$

For the spatial channel, f_2^{max} and f_2^{avg} are concatenated as:

$$U_1 = Cat(f_2^{max}, f_2^{avg}), \tag{12}$$

which are then passed through a convolution layer to obtain the weight matrix:

$$U = W_{(7\times7)}^{l+1} * U_1 + b^{l+1}. \tag{13}$$

The output features are also obtained by pixel dot multiplication of the weight matrix with the input

$$F^{Sam} = U \bigotimes Y_f. \tag{14}$$

The outputs of the two branches are added to obtain the final output of CSDA:

$$F = F^{Cam} + F^{Sam}. \tag{15}$$

By processing channel and spatial attention in parallel, CSDA ensures that both dimensions are enhanced independently before fusion, preserving the unique characteristics of spectral and spatial information. The dual-branch design effectively captures regional correlations in both domains, assigning higher weights to informative regions while suppressing irrelevant ones, thereby improving the discriminative power of the learned features for HSI classification.

3.4. The Hybrid Convolutional Layer

Multilayer 3D convolution is usually used to process hyperspectral images. In practice, the 1D spectrum and the 2D space are characterized independently. The use of 3D convolution increases the complexity of the model and prolongs the processing time. In the proposed DACINet, we also design a hybrid convolutional layer based on 2D and 3D convolutions. The layer consists of two 2D convolution layers and two 3D convolution layers. The size of the 3D convolutional layer kernel is $(7, 3, 3)$, and the size of the 2D convolutional kernel is $(3, 3)$. The hybrid layer begins with 3D convolutions because they are effective in capturing joint spectral–spatial features directly from the hyperspectral data, building a solid foundation for later processing. Given the feature map F output by the CSDA module, it is first processed by two consecutive 3D convolutional layers:

$$F' = 3D\text{Cov}(3D\text{Cov}(F)). \tag{16}$$

The output F′ is a 4D tensor, with dimensions corresponding to feature channels, spectral bands, height, and width. This 4D tensor is then reshaped into a 3D tensor by merging the channel and spectral dimensions, resulting in a feature map of the shape (channels × bands) × height × width. This reshaped tensor is then fed into two consecutive 2D convolutional layers:

$$F'' = 2D\text{Cov}(2D\text{Cov}(F')). \tag{17}$$

Through this reshaping, each channel in the resulting feature map encodes a specific combination of feature type and spectral band, allowing the 2D convolutions to learn cross-band interactions by combining these channels. This lets the model refine spatial patterns while

https://doi.org/10.3390/rs18071059

implicitly learning how different spectral bands interact, without the extra cost of additional 3D layers. After feature extraction, the output is passed through batch normalization and ReLU activation, then flattened and fed into a fully connected layer for classification. In this design, 3D convolutions extract features simultaneously in spectral and spatial directions, while 2D convolutions focus more on spatial feature refinement. When combined, they enable multi-level feature learning and improve the model's generalization ability and classification accuracy.

3.5. Classification

Finally, a fully softmax classifier is used for HSI classification. The class probability of each pixel is obtained through the softmax layer, and then the cross-entropy loss L is calculated:

$$\mathcal{L} = \frac{1}{M} \sum_{m=1}^{M} \sum_{k=1}^{K} ln(y_K^m) log(\widehat{y}_K^m), \tag{18}$$

where M is the number of samples in a small sample set. K is the total number of categories. y_K^m and $\widehat{y}_K^m$ represent actual and predicted sample labels respectively. The loss function tends to converge quickly on many kinds of problems, and the optimal solution can be found quickly.

4. Experiments

4.1. Datasets

In experiments, we adopted a stratified random sampling strategy, randomly selecting 5% of labeled samples per class for the IP dataset and 1% for the UP, SA, and Houston2013 datasets as training samples. The remaining samples were used for testing. Indian Pines (IP): The Indian Pines dataset is the first test data for HSIC, imaged by an Airborne Visual Infrared Imaging Spectrometer (AVIRIS) in 1992 on an Indian pine tree in Indiana, USA. Its spatial dimension is 145 × 145 pixels, the spectral dimension is 200 spectral bands, and it consists of 16 target categories. Table 1 presents the category names, number of categories of the Indian Pines dataset, as well as the corresponding color annotations for each category in the visualization of classification results.

Table 1. Category information of the Indian Pines dataset.

Number	Category	Color	Total Samples
1	Alfalfa		46
2	Corn-notill		1428
3	Corn-mintill		830
4	Corn		237
5	Grass/pasture		483
6	Grass/trees		730
7	Grass/pasture-mowed		28
8	Hay-windrowed		478
9	Oats		20
10	Soybean-notill		972
11	Soybean-mintill		2455
12	Soybean-clean		593
13	Wheat		205
14	Woods		1265
15	Bldg-grass-tree-drivers		386
16	Stone-steel-towers		93
Total	/	/	10,249

https://doi.org/10.3390/rs18071059

Pavia University (UP): The Pavia University dataset was acquired in 2001 over the University of Pavia campus, Italy, using the Reflective Optical System Imaging Spectrometer (ROSIS). The image comprises 610 × 340 pixels, with 115 spectral bands covering the wavelength range of 0.43–0.86 µm. After the removal of 12 noisy bands, the remaining 103 bands are commonly used for analysis. provides the category names, the number of categories, and the corresponding color annotations for this dataset. Table 2 presents the category names, number of categories of the Pavia University dataset, as well as the corresponding color annotations for each category in the visualization of classification results.

Table 2. Category information of the Pavia University dataset.

Number	Category	Color	Total Samples
1	Asphalt		6631
2	Meadows		18,649
3	Gravel		2099
4	Trees		3064
5	Painted Metal Sheets		1345
6	Bare Soil		5029
7	Bitumen		1330
8	Self-Blocking Bricks		3682
9	Shadows		947
Total	/	/	42,776

Salinas (SA): The Salinas hyperspectral dataset was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over the Salinas Valley, CA, USA. The image has dimensions of 512 × 217 pixels (totaling 111,104 pixels) and originally comprises 220 spectral bands covering a wavelength range of approximately 0.4–2.5 µm. The scene is annotated into 16 distinct land cover classes. Table 3 presents the category names, number of categories of the Salinas dataset, as well as the corresponding color annotations for each category in the visualization of classification results.

Table 3. Category information of the Salinas dataset.

Number	Category	Color	Total Samples
1	Broccoli-green-weeds_1		2009
2	Broccoli-green-weeds_2		3726
3	Fallow		1976
4	Fallow-rough-plow		1394
5	Fallow-smooth		2678
6	Stubble		3959
7	Celery		3579
8	Grapes-untrained		11,271
9	Soil-vinyard-develop		6203
10	Corn-senesced-green-weeds		3278
11	Lettuce-romaine-4wk		1068
12	Lettuce-romaine-5wk		1927
13	Lettuce-romaine-6wk		916
14	Lettuce-romaine-7wk		1070
15	Vinyard-untrained		7268
16	Vinyard-vertical-trellis		1807
Total	/	/	54,129

University of Houston 13 (HU): The HU dataset consists of 349 × 1905 pixels with 144 spectral channels ranging from 364 to 1046 nm and a spatial resolution of 2.5 m/pixel. In addition, the ground truth reference was subdivided into spatially disjoint subsets for training and testing, including 15 mutually exclusive urban land cover classes with 15,029 labeled pixels. Table 4 presents the category names, number of categories of the Houston 13 dataset, as well as the corresponding color annotations for each category in the visualization of classification results.

Table 4. Category information of the Houston 13 dataset.

Number	Category	Color	Total Samples
1	Grass-healthy		1238
2	Grass-stressed		1241
3	Grass-synth		690
4	Tree		1231
5	Soil		1229
6	Water		321
7	Residential		1255
8	Commercial		1231
9	Road		1239
10	Highway		1214
11	Railway		1222
12	Parking-lot1		1220
13	Parking-lot2		464
14	Tennis-court		423
15	Running-track		653
Total	/	/	14,871

4.2. Evaluation Metrics

To quantitatively compare the classification performance of different methods and modules from various aspects, in the following experiments, we adopt four commonly used evaluation metrics, namely, Overall Accuracy (OA), Average Accuracy (AA), Kappa Coefficient (Kap) and Accuracy per Class (AEC).

4.3. Implementations Details

All experiments are conducted on an NVIDIA T400 GPU (NVIDIA Corporation, Santa Clara, CA, USA), Windows 11 64-bit, Python 3.8.16 and PyTorch 2.2.1. For model training, the adaptive moment estimation (Adam) optimizer is employed, and the network starts with a learning rate of 0.001 and 100 Epochs. To better compare the advantages of different networks, the best results are highlighted in bold.

4.4. Comparison with State-of-the-Art Methods

The proposed model is compared with representative baselines spanning different methodological paradigms in HSI classification. These include: traditional machine learning methods RF [37] and SVM [38] to establish fundamental benchmarks; early deep learning models like Context [39] to trace the evolution from CNNs to advanced architectures; hybrid 2D/3D CNNs, such as HybridN [40] and CVSSN [41], as direct competitors sharing a similar hybrid convolutional design; attention-based networks RSSAN [42], SSTN [43], and SSAtt [44] to evaluate our CSDA module against prominent attention mechanisms; and graph convolutional networks like F-GCN [45] to assess performance across different paradigms for modeling non-Euclidean data relationships.

https://doi.org/10.3390/rs18071059

Table 5 shows the experimental results of different methods on the IP dataset. The results show that among the 16 categories of objects, nine categories of the proposed method achieved the best classification effect. The comprehensive classification effect is the best, with its OA reaching 96.78%, AA reaching 90.60%, and Kappa reaching 96.32%. Compared with the HybridN model with a suboptimal effect, our DACINet improves 1.54%, 0.04% and 4.02% on OA, AA and Kappa respectively. It is noteworthy that beyond the mean accuracy, DACINet often exhibits a lower standard deviation across multiple runs compared to other methods (as shown in Tables 5–8). This indicates a more stable and reliable classification performance, which is a significant advantage in practical applications. To quantify the impact of limited training samples on classification performance, we analyze the relationship between per class sample size and accuracy. As shown in Table 1, classes 1 (Alfalfa), 7 (Grass/pasture-mowed), and 9 (Oats) have the fewest training samples in the IP dataset, with only 46, 28, and 20 samples, respectively. Correspondingly, Table 5 reveals that these three classes consistently achieve the lowest accuracies across all compared methods. Even with our proposed DACINet, the accuracies on these minority classes are only 70.75%, 74.42%, and 73.33%, respectively. In contrast, classes with abundant training samples, such as class 2 (Corn-notill, 1428 samples) and class 11 (Soybean-mintill, 2455 samples), achieve accuracies exceeding 95% with the DACINet. This clear positive correlation between training sample size and classification accuracy demonstrates that limited sample availability is indeed the primary factor contributing to the suboptimal performance on these minority classes.

Table 5. Classification results (%) of various HSIC methods on IP dataset with 5% labeled samples per class. The best result for each class/metric is highlighted in bold.

Class	RF	SVM	Context	RSSAN	SSTN	SSAtt	HybridN	CVSSN	F-GCN	Ours
1	87.06 ± 5.63	42.86 ± 4.78	57.89 ± 5.27	88.89 ± 2.87	77.77 ± 3.98	90.93 ± 4.13	79.86 ± 6.24	64.71 ± 1.96	**93.83 ± 3.24**	70.75 ± 3.65
2	66.45 ± 6.12	66.93 ± 5.36	63.08 ± 5.32	62.27 ± 3.45	85.35 ± 4.53	83.72 ± 7.34	74.16 ± 2.19	93.33 ± 3.42	78.95 ± 6.25	**85.66 ± 4.32**
3	70.92 ± 3.42	75.45 ± 3.54	47.28 ± 2.65	60.00 ± 5.43	93.40 ± 4.65	81.05 ± 3.25	93.60 ± 1.78	96.24 ± 2.53	78.63 ± 3.32	**96.41 ± 3.21**
4	43.38 ± 2.65	48.56 ± 4.23	46.06 ± 1.68	82.72 ± 3.24	**98.95 ± 2.43**	88.94 ± 2.67	89.37 ± 1.32	99.98 ± 0.45	96.62 ± 2.13	96.07 ± 3.21
5	72.57 ± 6.78	85.49 ± 5.34	81.93 ± 5.67	79.41 ± 7.48	84.07 ± 4.38	87.57 ± 4.98	89.31 ± 3.78	92.22 ± 3.95	76.84 ± 8.12	**97.55 ± 2.45**
6	81.03 ± 3.25	84.48 ± 2.23	76.09 ± 1.98	91.33 ± 3.14	96.13 ± 3.21	**99.22 ± 1.13**	99.67 ± 1.02	98.56 ± 2.32	95.79 ± 2.11	99.14 ± 1.10
7	58.33 ± 10.11	80.00 ± 11.23	99.28 ± 3.54	**99.52 ± 2.21**	35.71 ± 12.23	72.00 ± 11.43	71.11 ± 9.87	94.73 ± 7.32	96.43 ± 3.21	74.42 ± 5.43
8	86.24 ± 3.12	88.82 ± 2.28	88.13 ± 5.45	93.64 ± 3.45	94.93 ± 4.52	94.57 ± 4.87	93.32 ± 5.43	92.26 ± 3.78	**100.00 ± 2.31**	97.09 ± 3.54
9	99.19 ± 2.33	60.00 ± 6.49	84.62 ± 5.46	55.56 ± 8.98	42.86 ± 7.56	93.33 ± 4.65	72.22 ± 3.21	80.00 ± 6.65	**99.03 ± 2.31**	73.33 ± 5.63
10	64.48 ± 4.32	76.44 ± 6.56	66.22 ± 7.66	59.84 ± 5.46	90.92 ± 5.34	85.23 ± 3.12	85.85 ± 7.65	93.46 ± 3.65	78.91 ± 5.43	**93.85 ± 4.23**
11	63.18 ± 3.44	69.17 ± 5.43	73.27 ± 4.53	72.88 ± 5.74	96.04 ± 2.35	90.25 ± 4.34	96.46 ± 1.29	97.84 ± 5.43	78.84 ± 6.12	**96.12 ± 3.67**
12	57.14 ± 9.49	67.51 ± 5.46	58.06 ± 8.87	55.46 ± 8.56	78.65 ± 5.69	67.91 ± 7.88	89.64 ± 4.57	85.85 ± 6.12	85.50 ± 5.32	**95.92 ± 4.17**
13	88.83 ± 5.34	92.67 ± 4.35	71.62 ± 6.54	86.11 ± 5.87	97.30 ± 2.89	97.00 ± 5.12	98.72 ± 4.30	95.42 ± 3.46	**100.00 ± 2.17**	98.73 ± 4.12
14	83.85 ± 1.65	84.88 ± 4.32	90.99 ± 3.19	88.77 ± 5.43	96.10 ± 3.10	96.54 ± 4.32	96.87 ± 2.90	97.62 ± 4.87	94.86 ± 5.29	**97.92 ± 3.82**
15	66.67 ± 6.54	80.15 ± 4.21	63.66 ± 4.20	72.49 ± 1.89	90.02 ± 4.12	88.83 ± 4.76	78.63 ± 3.21	**95.86 ± 2.69**	95.34 ± 4.01	93.31 ± 3.54
16	98.41 ± 2.10	98.25 ± 2.98	97.78 ± 4.12	60.53 ± 3.02	**98.82 ± 3.61**	93.33 ± 4.08	42.74 ± 5.12	81.90 ± 2.89	93.24 ± 3.06	94.12 ± 4.32
OA	70.42 ± 3.04	75.28 ± 4.23	71.18 ± 4.57	73.18 ± 4.34	91.32 ± 5.41	88.35 ± 3.21	93.95 ± 2.87	95.22 ± 5.34	95.24 ± 3.87	**96.78 ± 3.24**
AA	75.19 ± 2.65	75.21 ± 4.32	72.92 ± 3.12	75.62 ± 5.01	84.81 ± 5.65	88.72 ± 4.69	84.35 ± 3.20	90.50 ± 2.19	90.56 ± 2.97	**90.60 ± 3.05**
Kappa	65.71 ± 3.02	71.45 ± 3.21	66.98 ± 4.32	69.23 ± 3.76	90.13 ± 2.65	86.71 ± 1.78	93.09 ± 4.03	94.55 ± 3.78	92.30 ± 3.21	**96.32 ± 3.58**

Tables 6 and 7 present the experimental results of different methods on the UP and SA datasets. For the UP dataset, the results of Table 6 show that in the nine types of land features, our proposed method achieved the best classification effect in six of them. Due to the relatively small number of land feature types in the UP dataset, as can be observed from the data in the table, the accuracy of our proposed DACINet in eight types of land features reached over 90%, and in the second and fifth types, it was as high as over 99%. Kappa is mainly used to measure whether the final classification result is consistent with the actual observation value, and it measures the stability of the entire model's random classification. Our OA index has climbed to 97.77%, and the AA index has reached an excellent result of 96.72%. At the same time, the consistency test index Kappa is as high as 97.04%. Compared with the HybridN model that ranked second in comprehensive performance, our proposed method has improved by three percentage points in OA, jumped by 6.08% in AA, and achieved a 4.01% increase in Kappa. It can be seen that our proposed model has achieved the best overall classification effect, average classification effect, and model stability on the entire dataset.

https://doi.org/10.3390/rs18071059

Table 6. Classification results (%) of various HSIC methods on UP dataset with 1% labeled samples per class. The best result for each class/metric is highlighted in bold.

Class	RF	SVM	Context	RSSAN	SSTN	SSAtt	HybridN	CVSSN	Ours
1	79.58 ± 3.87	78.28 ± 2.56	89.65 ± 2.84	86.19 ± 4.23	89.66 ± 3.08	87.96 ± 2.19	93.16 ± 3.18	95.45 ± 2.76	**96.84 ± 2.85**
2	84.04 ± 1.98	84.47 ± 2.10	94.30 ± 1.28	96.92 ± 2.09	97.46 ± 3.21	98.55 ± 3.98	98.72 ± 2.76	98.71 ± 4.01	**99.45 ± 1.65**
3	55.06 ± 4.70	76.82 ± 3.09	67.47 ± 5.31	59.67 ± 5.42	74.78 ± 6.23	74.32 ± 2.65	84.95 ± 3.67	86.35 ± 2.37	**88.69 ± 3.20**
4	90.87 ± 4.32	91.73 ± 3.41	91.75 ± 4.79	**99.09 ± 2.12**	91.09 ± 3.56	98.39 ± 4.28	93.11 ± 1.67	96.77 ± 3.23	97.33 ± 4.11
5	95.00 ± 3.11	93.97 ± 3.33	99.98 ± 2.04	90.14 ± 3.97	98.59 ± 4.22	98.08 ± 3.65	95.63 ± 2.17	96.41 ± 5.04	**99.77 ± 0.78**
6	75.77 ± 1.67	94.05 ± 4.31	80.84 ± 3.44	86.04 ± 2.17	96.82 ± 2.04	96.41 ± 4.13	97.06 ± 3.22	96.36 ± 4.28	**98.26 ± 2.65**
7	73.85 ± 3.33	69.68 ± 5.67	81.13 ± 6.53	71.05 ± 3.76	**99.75 ± 1.54**	94.87 ± 3.47	92.42 ± 2.38	90.39 ± 3.42	98.13 ± 3.63
8	71.54 ± 3.24	72.87 ± 3.56	80.41 ± 4.44	76.00 ± 5.05	89.06 ± 3.27	86.13 ± 4.12	85.86 ± 6.55	89.04 ± 4.29	**90.80 ± 5.08**
9	99.46 ± 2.22	**99.89 ± 1.87**	86.31 ± 2.35	80.86 ± 5.68	91.65 ± 3.18	95.81 ± 5.23	77.95 ± 4.56	97.37 ± 2.33	97.25 ± 4.55
OA	81.71 ± 4.33	83.43 ± 1.21	88.86 ± 3.24	89.09 ± 4.01	93.64 ± 2.87	94.08 ± 1.66	94.77 ± 2.21	95.93 ± 3.06	**97.77 ± 2.70**
AA	80.58 ± 2.21	84.64 ± 3.24	85.72 ± 4.10	82.88 ± 3.17	92.10 ± 3.29	92.28 ± 4.19	90.64 ± 3.95	94.09 ± 2.79	**96.72 ± 2.64**
Kappa	75.02 ± 2.98	77.24 ± 4.19	85.18 ± 2.39	85.49 ± 1.89	91.56 ± 3.25	92.13 ± 2.13	93.03 ± 1.65	94.60 ± 3.10	**97.04 ± 1.23**

Table 7. Classification results (%) of various HSIC methods on SA dataset with 1% labeled samples per class. The best result for each class/metric is highlighted in bold.

Class	RF	SVM	Context	RSSAN	SSTN	SSAtt	HybridN	CVSSN	Ours
1	99.68 ± 4.09	99.78 ± 2.19	97.02 ± 3.21	99.95 ± 3.67	95.48 ± 2.79	99.98 ± 1.79	96.39 ± 5.04	96.27 ± 3.78	**100.00 ± 1.64**
2	98.75 ± 4.32	98.86 ± 3.08	98.15 ± 5.34	99.86 ± 2.79	97.55 ± 4.08	99.08 ± 2.18	99.76 ± 4.37	**99.98 ± 3.56**	99.94 ± 2.19
3	84.90 ± 7.64	87.95 ± 7.65	92.67 ± 3.89	90.73 ± 5.43	93.43 ± 5.22	93.22 ± 6.34	**99.31 ± 3.87**	95.46 ± 4.28	99.23 ± 2.86
4	97.16 ± 5.20	97.36 ± 4.39	96.86 ± 8.02	91.26 ± 7.04	96.27 ± 4.67	**98.20 ± 2.59**	97.34 ± 5.33	95.83 ± 3.27	96.18 ± 5.33
5	93.95 ± 6.33	95.49 ± 5.44	97.87 ± 4.75	99.05 ± 0.87	**99.59 ± 0.42**	96.92 ± 3.44	98.11 ± 2.12	99.17 ± 0.65	97.25 ± 2.19
6	99.59 ± 0.46	99.91 ± 0.10	98.44 ± 2.10	99.64 ± 1.01	99.95 ± 0.87	**100.00 ± 0.42**	99.01 ± 1.45	99.99 ± 1.01	99.92 ± 0.45
7	97.77 ± 2.11	97.75 ± 2.01	97.68 ± 3.21	97.19 ± 3.24	99.91 ± 1.02	98.22 ± 2.54	99.96 ± 0.78	98.44 ± 2.13	100.00 ± 0.29
8	72.04 ± 3.45	72.34 ± 5.65	85.63 ± 5.43	84.12 ± 2.98	87.10 ± 5.44	85.34 ± 4.66	97.19 ± 3.77	92.81 ± 5.67	**99.75 ± 1.32**
9	95.85 ± 4.35	98.47 ± 3.22	99.00 ± 2.10	98.59 ± 3.21	99.79 ± 1.02	**99.95 ± 0.12**	99.46 ± 0.79	99.46 ± 1.22	99.80 ± 1.22
10	82.79 ± 6.54	89.33 ± 4.58	91.71 ± 5.47	91.68 ± 7.12	95.41 ± 3.40	94.51 ± 3.29	98.44 ± 4.39	97.85 ± 3.21	**99.52 ± 1.23**
11	94.63 ± 2.30	90.14 ± 5.43	93.35 ± 3.89	96.93 ± 3.98	80.22 ± 5.43	94.66 ± 4.10	93.48 ± 4.29	99.41 ± 1.67	**99.72 ± 2.10**
12	95.08 ± 3.09	95.68 ± 5.67	98.10 ± 3.08	94.10 ± 5.30	97.21 ± 3.12	97.39 ± 4.33	98.03 ± 4.02	**99.86 ± 1.02**	98.92 ± 2.06
13	92.13 ± 6.22	92.84 ± 8.11	98.96 ± 4.52	99.65 ± 2.10	**99.89 ± 0.23**	97.88 ± 3.44	95.87 ± 4.32	99.88 ± 1.23	98.67 ± 3.22
14	91.86 ± 8.33	95.68 ± 5.54	98.10 ± 4.55	94.10 ± 6.43	97.21 ± 2.35	97.39 ± 5.41	98.03 ± 4.29	**99.86 ± 2.45**	98.92 ± 4.22
15	68.15 ± 5.66	74.12 ± 7.44	81.53 ± 3.49	79.95 ± 8.98	83.53 ± 4.30	80.47 ± 4.87	95.98 ± 5.66	89.69 ± 2.33	**99.39 ± 1.22**
16	94.38 ± 3.22	98.71 ± 4.23	97.59 ± 4.33	99.56 ± 1.23	98.18 ± 3.22	99.76 ± 1.29	98.89 ± 4.32	99.12 ± 1.27	**100.00 ± 1.02**
OA	86.51 ± 3.21	88.30 ± 3.20	92.58 ± 4.32	92.28 ± 5.34	93.46 ± 2.37	93.15 ± 3.24	97.95 ± 4.22	96.24 ± 4.55	**99.53 ± 2.87**
AA	91.17 ± 3.22	92.77 ± 5.44	95.18 ± 2.66	95.12 ± 3.42	95.17 ± 3.41	95.79 ± 3.12	97.64 ± 3.11	97.48 ± 3.09	**99.58 ± 1.03**
Kappa	84.95 ± 6.54	86.93 ± 3.90	91.74 ± 5.34	91.40 ± 4.26	92.71 ± 5.43	92.37 ± 3.89	97.72 ± 4.23	95.81 ± 2.98	**99.48 ± 2.10**

In Table 7, it can be seen that the OA index of our proposed DACINet model reached 99.53%, while the OA of the suboptimal HybridN network was 97.95%, an increase of 1.58% for the SA dataset. However, the OA of other methods generally fluctuated around 90%. The AA index is used to evaluate the overall classification performance of a model. According to the experimental results in the table, the overall classification effect of each model on the SA dataset was above 90%. The AA index of the DACINet model reached 99.58%, which was a significant improvement compared to SSTN and SSAtt, and it was 2.12% higher than the HybridN network. Kappa is also an important indicator for evaluating the effectiveness of a classification model. For the proposed method, Kappa reached 97.04%, and in 16 land cover categories, our proposed method achieved the best classification effect for 10 categories. This demonstrates that the DACINet can more accurately and stably complete the hyperspectral image classification.

Table 8 presents the experimental results of different methods on the HU dataset. From the experimental results, it can be observed that among the 15 land cover categories, our proposed DACINet achieves the best classification performance in nine categories, demonstrating its powerful feature discrimination capability. Particularly noteworthy is that on categories with complex textural characteristics, such as class 5 (Grass) and class 14 (Tennis Court), our method achieves accuracies of 97.55% and 96.32%, respectively, significantly outperforming other comparative methods. In terms of comprehensive evaluation metrics, our method achieves the best results across all three key indicators: OA, AA, and Kappa, reaching 86.67%, 89.20%, and 87.12%, respectively. Compared with the suboptimal A2S2K model, our DACINet improves by 1.72%, 3.85%, and 4.03% in OA, AA, and Kappa,

https://doi.org/10.3390/rs18071059

respectively. It is worth noting that even on categories with lower sample discriminability, such as class 10 (Coastal) and class 13 (Parking Lot), our method maintains relatively stable classification performance, benefiting from the effective spectral–spatial feature capability of the CSDA module. Overall, on this challenging HU dataset, the DACINet demonstrates excellent classification performance and good generalization ability, further verifying the effectiveness and robustness of the proposed framework.

Table 8. Classification results (%) of various HSIC methods on HU dataset with 1% labeled samples per class. The best result for each class/metric is highlighted in bold.

Class	RF	SVM	Context	RSSAN	SSTN	SSAtt	A2S2K	CVSSN	Ours
1	89.06 ± 4.63	82.86 ± 3.78	77.89 ± 3.27	90.89 ± 5.87	77.98 ± 4.98	91.33 ± 3.13	86.56 ± 2.24	78.59 ± 4.16	**92.95 ± 2.24**
2	85.76 ± 5.23	84.89 ± 6.16	65.28 ± 6.23	74.17 ± 2.87	85.55 ± 3.23	88.25 ± 8.49	89.16 ± 5.10	90.13 ± 3.22	**91.95 ± 4.15**
3	90.02 ± 8.03	92.45 ± 3.04	78.18 ± 9.61	82.02 ± 10.13	84.40 ± 5.50	82.15 ± 3.15	87.27 ± 7.18	**96.14 ± 3.50**	88.83 ± 5.23
4	90.43 ± 4.65	92.36 ± 6.53	84.26 ± 5.83	88.78 ± 7.14	86.85 ± 5.23	90.14 ± 4.70	93.01 ± 6.87	92.18 ± 4.65	**95.22 ± 4.23**
5	72.57 ± 6.78	85.49 ± 5.34	81.93 ± 5.67	79.41 ± 7.48	84.07 ± 4.38	87.57 ± 4.98	89.31 ± 3.78	92.22 ± 3.95	**97.55 ± 2.45**
6	**95.03 ± 6.85**	94.48 ± 7.13	76.87 ± 8.91	83.32 ± 7.24	85.43 ± 4.76	89.22 ± 5.13	91.02 ± 4.22	94.16 ± 3.23	93.23 ± 6.80
7	67.43 ± 8.91	67.40 ± 11.43	78.87 ± 6.54	79.34 ± 7.16	74.87 ± 9.23	78.13 ± 7.93	76.11 ± 6.37	**80.13 ± 8.32**	79.42 ± 7.43
8	73.74 ± 11.12	74.08 ± 8.28	65.13 ± 13.45	64.32 ± 11.45	76.83 ± 12.52	83.47 ± 8.87	83.32 ± 6.39	84.26 ± 8.78	**89.68 ± 9.41**
9	67.19 ± 7.93	65.38 ± 9.49	72.62 ± 10.46	73.56 ± 9.98	81.86 ± 7.56	81.03 ± 7.65	85.22 ± 8.21	87.10 ± 6.65	**89.13 ± 4.31**
10	59.48 ± 13.32	57.44 ± 12.56	54.01 ± 14.66	57.84 ± 14.46	59.42 ± 9.34	69.93 ± 8.92	**85.15 ± 9.65**	78.46 ± 10.65	78.65 ± 11.23
11	53.18 ± 8.44	63.17 ± 7.43	68.27 ± 8.53	65.88 ± 6.64	56.44 ± 12.35	71.25 ± 6.34	76.46 ± 6.29	80.84 ± 5.43	**85.32 ± 7.67**
12	52.64 ± 9.59	57.51 ± 8.46	55.56 ± 8.67	58.36 ± 9.56	64.15 ± 7.69	73.91 ± 9.88	74.14 ± 6.57	**80.15 ± 9.12**	79.42 ± 8.17
13	48.83 ± 13.34	50.67 ± 9.35	80.12 ± 7.54	79.41 ± 6.87	83.30 ± 6.89	90.10 ± 7.12	92.72 ± 5.30	91.42 ± 6.46	**94.12 ± 7.17**
14	80.85 ± 5.65	84.74 ± 7.32	79.59 ± 6.19	87.13 ± 8.43	82.45 ± 5.10	78.54 ± 9.32	88.17 ± 5.78	85.62 ± 4.37	**96.32 ± 6.82**
15	96.27 ± 4.54	**99.85 ± 0.71**	84.26 ± 6.20	78.59 ± 6.89	93.22 ± 4.12	89.83 ± 4.76	97.63 ± 3.21	92.16 ± 3.69	94.34 ± 2.01
OA	71.82 ± 3.04	75.78 ± 2.23	69.48 ± 4.17	70.18 ± 4.24	74.32 ± 4.41	80.35 ± 3.21	84.95 ± 4.17	83.24 ± 3.32	**86.67 ± 4.24**
AA	75.79 ± 2.55	78.11 ± 3.42	70.12 ± 4.24	75.62 ± 4.21	79.81 ± 5.65	81.72 ± 4.29	85.35 ± 3.23	84.98 ± 4.37	**89.20 ± 4.15**
Kappa	69.11 ± 5.12	73.41 ± 5.21	67.28 ± 7.32	68.54 ± 5.76	74.13 ± 5.65	78.91 ± 4.78	83.09 ± 3.67	85.30 ± 3.34	**87.12 ± 5.18**

4.5. Ablation Studies

Ablation of different modules. In this section, ablation experiments are conducted to verify the functions of different modules in the DACINet. The CIFM, CSDA and hybrid convolution layer are introduced to the backbone network step by step. The results are shown in Table 9, where $\surd$ indicates the module is used, and $\times$ indicates it is not used. It can be seen that the classification results only through the CIFM are the worst, with OA of 85.20% and Kappa of 92.90 on the IP dataset, and with OA of 90.40% and Kappa of 87.14 on the UP dataset. This is because it focuses on fusion and lacks judgment on the validity. When combined with the CIFM and hybrid convolution layer, the overall classification accuracy and stability of the three are improved, which indicates that the feature extraction after the context interaction fusion is effective. Obviously, the DACINet with all modules obtains the best classification accuracy in OA and Kappa on the three datasets, specifically OA of 99.53% and Kappa of 99.48 on the SA dataset. This indicates that all modules play their own effective roles and jointly improve the hyperspectral classification.

Table 9. Ablation results of different modules. Bold values indicate the best results among all compared methods.

Different Modules			Metrics	IP	UP	SA
CIFM	CSDA	HCL				
×	×	$\surd$	OA(%)	93.17	97.05	97.47
			Kappa	93.63	96.07	98.41
$\surd$	×	×	OA(%)	85.20	90.40	97.71
			Kappa	82.90	87.14	97.58
$\surd$	×	$\surd$	OA(%)	94.76	97.68	98.56
			Kappa	94.02	96.93	98.51
$\surd$	$\surd$	$\surd$	OA(%)	**96.78**	**97.77**	**99.53**
			Kappa	**96.32**	**97.04**	**99.48**

https://doi.org/10.3390/rs18071059

Effectiveness of CSDA. In Table 10, we show the difference between the proposed CSDA and CBAM, where $\surd$ indicates the module is used, and $\times$ indicates it is not used. It is observed that when compared with the baseline without any attention mechanism, the CBAM mechanism dose not improve the performance on the UP and SA datasets effectively. On the IP dataset, the OA increases by 1.85%, and Kappa increases by 2.11; the performance is improved significantly. This is because the sample distribution of the IP dataset is unbalanced, and the discriminant information can be effectively screened by introducing the attention mechanism. The proposed CSDA achieves performance improvement on all three datasets, and it is higher than CBAM. This indicates that screening of spectral and spatial features through dual-channels can effectively alleviate sample imbalance, while it also improves classification performance in common scenarios.

Table 10. Ablation results of different attention. Bold values indicate the best results among all compared methods.

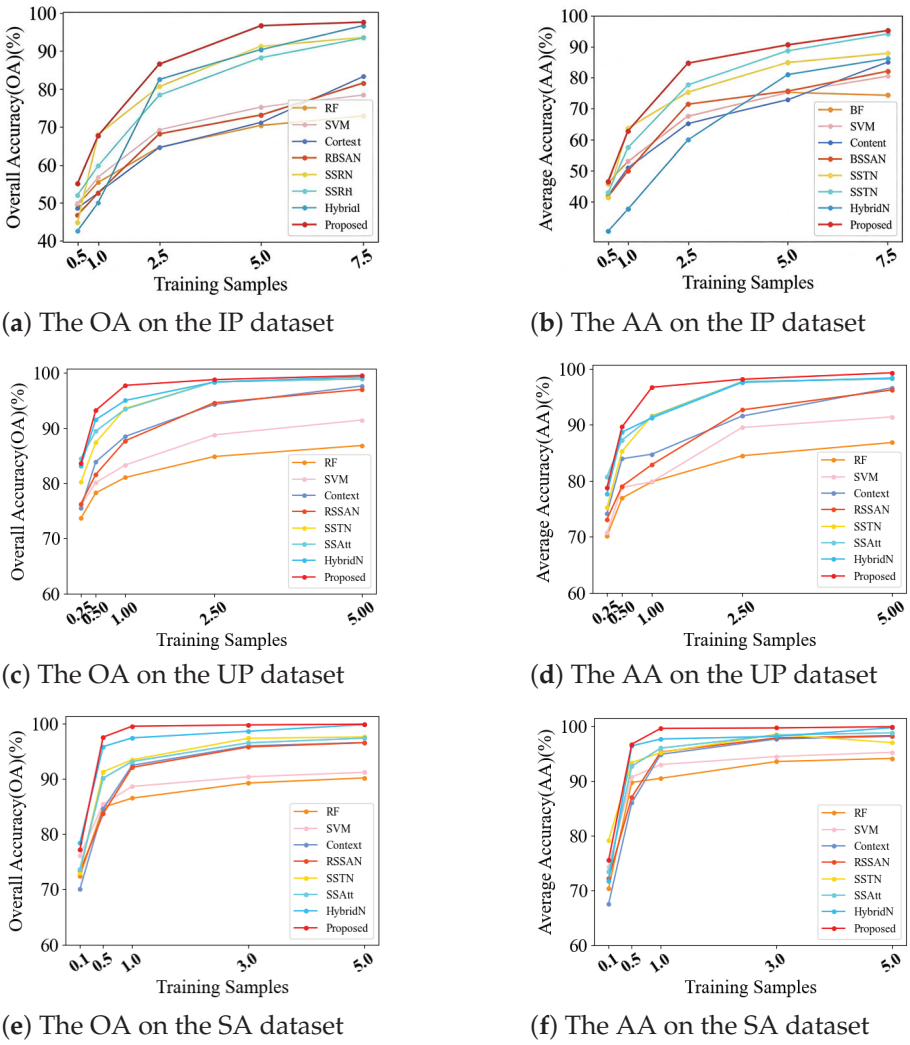
Attention Mechanism		Metrics	IP	UP	SA
CBAM	**CSDA**				
$\times$	$\times$	OA(%)	94.76	97.68	98.56
		Kappa	94.02	96.93	98.51
$\surd$	$\times$	OA(%)	96.61	97.28	98.90
		Kappa	96.13	96.38	98.77
$\times$	$\surd$	OA(%)	**96.78**	**97.77**	**99.53**
		Kappa	**96.32**	**97.04**	**99.48**

Validity of Robustness. To verify the robustness of the proposed model, we conduct experiments on three datasets with different samples. The results on the IP dataset are shown in Figure 4. It can be seen that both OA and AA increase as the percentage of training samples increases. In detail, deep learning models significantly improve the classification performance of machine learning models (RF, SVM) across all sample distributions. HybridN has a lower OA and AA score on the IP dataset in few-shot samples, especially when the sample is less than 5%. Overall comparison shows that the proposed DACINet results in the best OA and AA scores across all sample distributions, which indicates the DACINet has stronger robustness than others. To further verify the impact of sample size on classification, we calculated the classification confusion matrix for the nine categories of the UP dataset, as shown in Figure 5. The classification accuracy of the category with a smaller sample size (such as Shadows) is significantly lower than that of the category with a larger sample size (such as Meadows). This also verifies the contribution of sample size to classification accuracy. Small sample studies still need to be strengthened.

Analysis of Band Dimension. The dimensionality reduction has a certain impact on the final classification performance of the model. For this reason, we conducted relevant experiments on three datasets while keeping other conditions unchanged and varying the band dimensions. The experimental results are shown in Figure 6. The results show that the performance of the three different datasets varies in different band dimension parameters. When the dimension reduction number is set to 36 for the IP dataset, the values of the three evaluation parameters (OA, AA, and Kappa) are the highest. When it is set to 38, the performance of the model is the worst. Among them, due to the unbalanced data sample distribution in the IP dataset, the values of the OA and Kappa parameters are not much different, while the value of the AA parameter fluctuates significantly. When the dimension reduction band number is set to 13 for the UP dataset, the performance is the best. As the number of bands set increases, the effect gradually decreases, and when it is set to 19, the decline is the greatest, and the classification effect is the worst. For the SA dataset, the

https://doi.org/10.3390/rs18071059

values of the three evaluation parameters increase first and then decrease as the number of bands set increases. When the parameters are set to 17, the effect reaches the optimal state. Based on the experimental results, the band parameters of the IP, UP, and SA datasets are set to 36, 13, and 17 respectively.



(**a**) The OA on the IP dataset

(**b**) The AA on the IP dataset

(**c**) The OA on the UP dataset

(**d**) The AA on the UP dataset

(**e**) The OA on the SA dataset

(**f**) The AA on the SA dataset

Figure 4. Robustness verification of the proposed model on the IP, UP, and SA datasets. The x-axis represents the percentage of labeled samples (from each class) used for training. The y-axis represents the classification accuracy (Overall Accuracy, OA, or Average Accuracy, AA).

Selection of the Convolution Kernel. The size of the convolution kernel directly affects the amount of information about the adjacent pixels in the selected space. To better explore the different datasets' requirements for the spatial dimension in the model, this paper also conducts comparative experiments with different convolution kernels. A 3 × 3 small kernel in the spatial dimension is sufficient to extract effective spatial features, and it has a relatively small number of parameters. In the spectral dimension, experiments are conducted on the selected UP, IP, SA, and HU datasets with intervals of two pixels ranging from 1× to 13×. OA and AA are used as the classification results for the four datasets. In the experiment, when the length of the original spectral bands is less than the length of the vector that needs to be mapped and filled, the triangular principle is adopted, that is, it is filled in two steps. The experimental results are shown in Figure 7. The results show that the overall trend on the four datasets is that it first increases with the increase in the spatial window and then tends to be balanced. When the spectral size is less than seven,

https://doi.org/10.3390/rs18071059

the performance gradually improves, and after seven, the classification accuracy tends to stabilize. Considering the overall computational cost, in all subsequent experiments, the convolution kernel is set to 7 × 3 × 3.

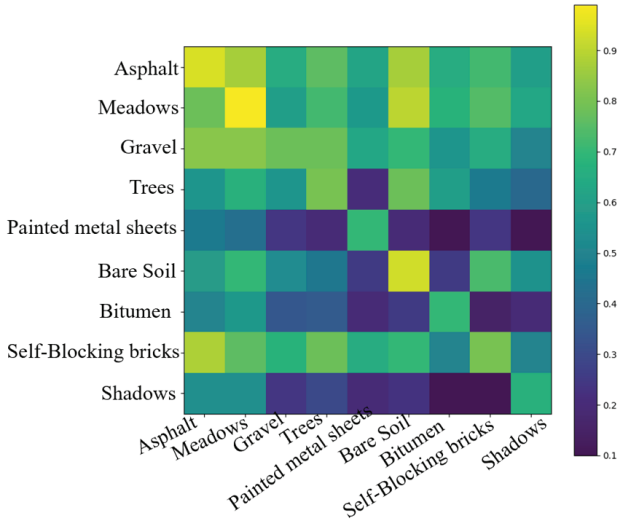
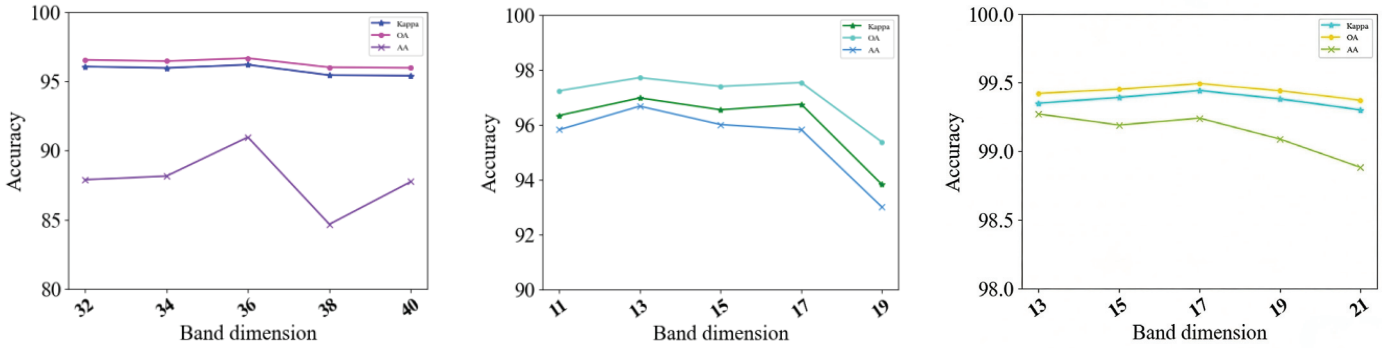


Figure 5. The confusion matrix for 9 categories in the UP dataset.



(**a**) IP (**b**) UP (**c**) SA

Figure 6. The OA, AA, and Kappa values of different bands on the IP, UP, and SA datasets.

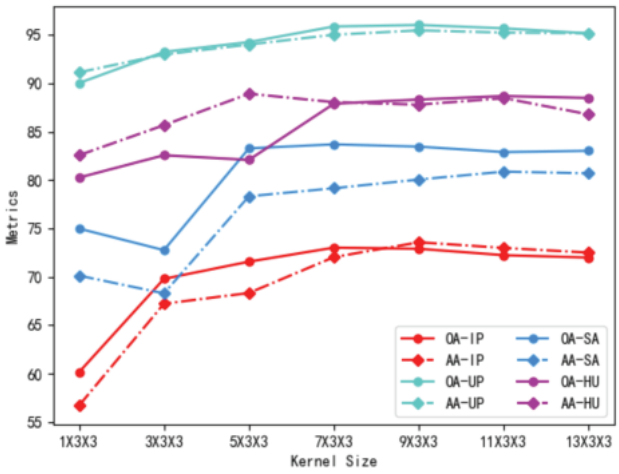


Figure 7. Different kernel size on IP, UP, SA and HU datasets.

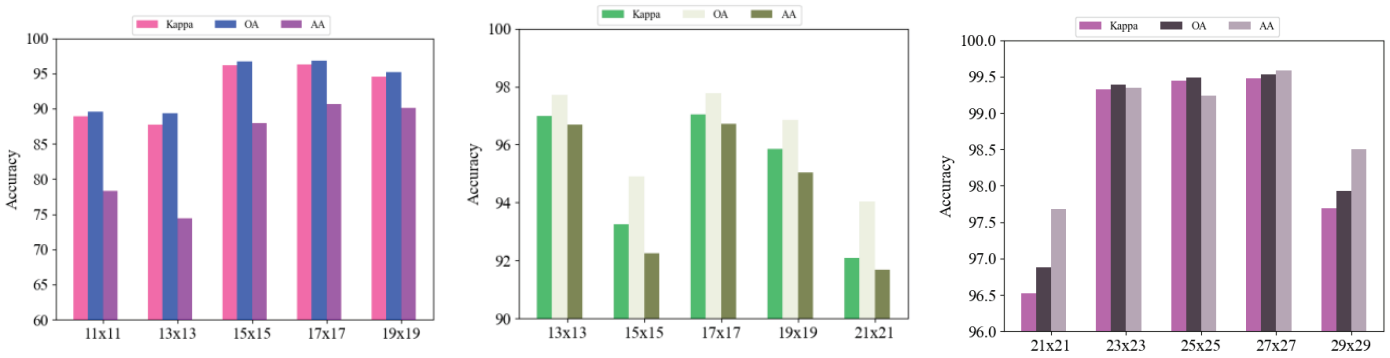
Analysis of the Complexity. Table 11 presents the complexity comparison of classification networks using 3D convolution kernels on the IP and UP datasets, mainly focusing

on the parameters and floating-point numbers of neural networks. The results show that when using only a 3D CNN, the network parameters are the least, but the floating-point numbers are more. Since the main body of the A2S2K network adopts the residual network, its parameters are the least, but the computational load is relatively large. The proposed method ranks second in terms of both network parameters and floating-point numbers, achieving the best comprehensive evaluation effect. The proposed hybrid convolution layer integrates spatial and spectral information. Among them, the 3D convolution simultaneously captures feature information in both the spectral and spatial dimensions, and introduces 2D convolution to enhance spatial feature extraction. Compared with using only 3D convolution, this design significantly reduces the total number of parameters and the number of floating-point operations, while ensuring the model's performance.

Table 11. Comparison of Params and FLOPs for different methods. Bold values indicate the best results among all compared methods.

Dataset	Metrics	Context	SSAN	3D CNN	A2S2K	Ours
IP	Params (M)	1.211	87.020	0.975	**0.371**	0.578
	FLOPs (M)	84.79	7062.96	**30.284**	170.45	50.98
UP	Params (M)	0.703	25.024	4.657	**0.221**	0.631
	FLOPs (M)	49.52	1982.81	**2.298**	87.30	43.561

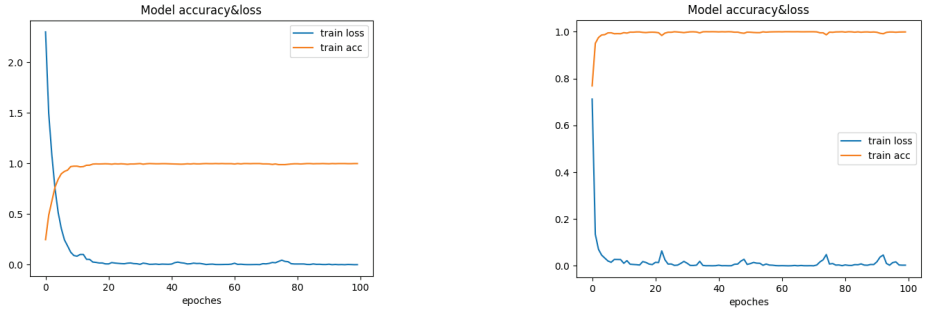
Analysis of Input Spatial. The size of the input space dimensions also has an impact on the final classification effect of the model. In the three datasets, while setting the optimal band parameters and keeping other conditions consistent, we changed the input space size parameter of the model for experiments to find the most suitable input space size for the three datasets. The experimental results are shown in Figure 8. The results show that the classification performance of the three datasets varies under different input space sizes. For the IP and UP datasets, the fluctuation amplitudes of the three evaluation parameters (OA, AA, and Kappa) are consistent. As the input space size increases, they first decrease, then increase, and then decrease again. Both the IP and UP datasets achieve the best classification effect at an input size of 17 × 17. However, for the SA dataset, the three evaluation parameters show an overall trend of increasing first and then decreasing as the input size increases. The best classification effect is achieved when the input size is set to 27 × 27. Therefore, based on the experimental results, we set the input size of the IP and UP datasets to 17 × 17, and the SA dataset to 27 × 27.



(**a**) IP (**b**) UP (**c**) SA

Figure 8. The OA, AA, and Kappa values under different input spatial sizes on the IP, UP, and SA datasets. The x-axis values represent the spatial dimension of input patches.

Convergence Analysis of the DACINet. To verify the rationality of the Epoch parameter setting of the model, this section visualizes the training convergence process of the Indian Pines and Salinas datasets. The experimental results are shown in Figure 9. The results indicate that the convergence speed of the IP dataset is relatively slow, and it converges completely at around 30. However, the loss function and accuracy curves of this dataset are relatively stable. The SA dataset converges faster and has reached a stable state at 10, but the loss function curve fluctuates significantly, while the accuracy curve is stable. This proves that the Epoch setting we have made is reasonable, and the proposed model has good stability and generalization ability.



(**a**) The OA on the IP dataset (**b**) The OA on the SA dataset

Figure 9. Convergence analysis of the proposed DACINet. The training loss and accuracy curves on (**a**) the IP dataset and (**b**) the SA dataset.

4.6. Visualization Analysis

To show the effectiveness of the DACINet more intuitively, the classification results of the proposed DACINet and other representative methods on the IP dataset are visualized in Figure 10, colors follow the same scheme as described in Table 1. Among them, a small area in the figure indicates a small number of ground objects, and fewer samples are selected in the classification process. As observed, SSTN and SSAtt classification methods have more errors in the classification of ground objects with fewer samples. The F-GCN method is improved compared with other methods, but it tends to produce wrong results at the intersection of two categories. Clearly, the DACINet has the best overall classification performance, and the boundary accuracy has been significantly improved. The classification visualization results of datasets UP, SA and HU are also presented in Figures 11–13, colors follow the same scheme as described in Tables 2–4.

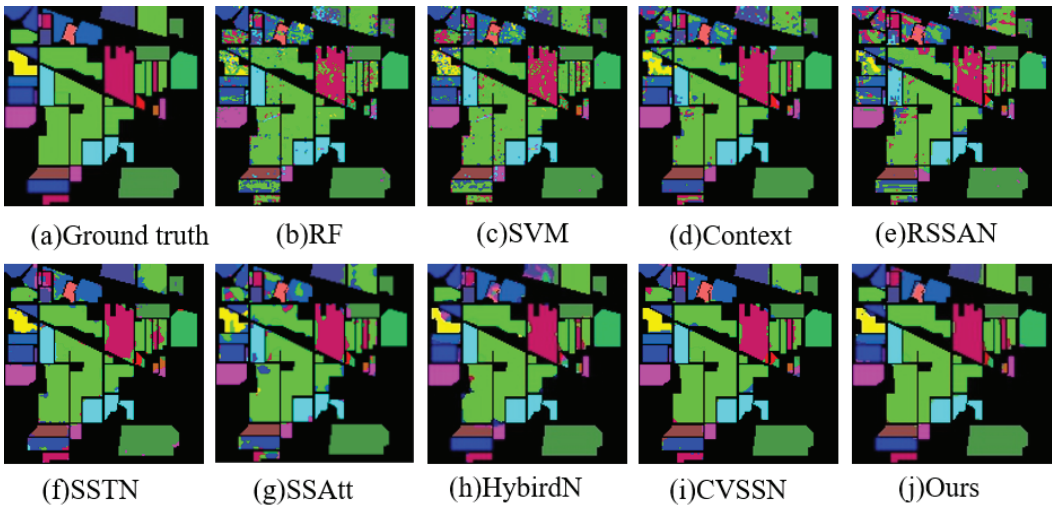


Figure 10. Visualization of different models on IP dataset.

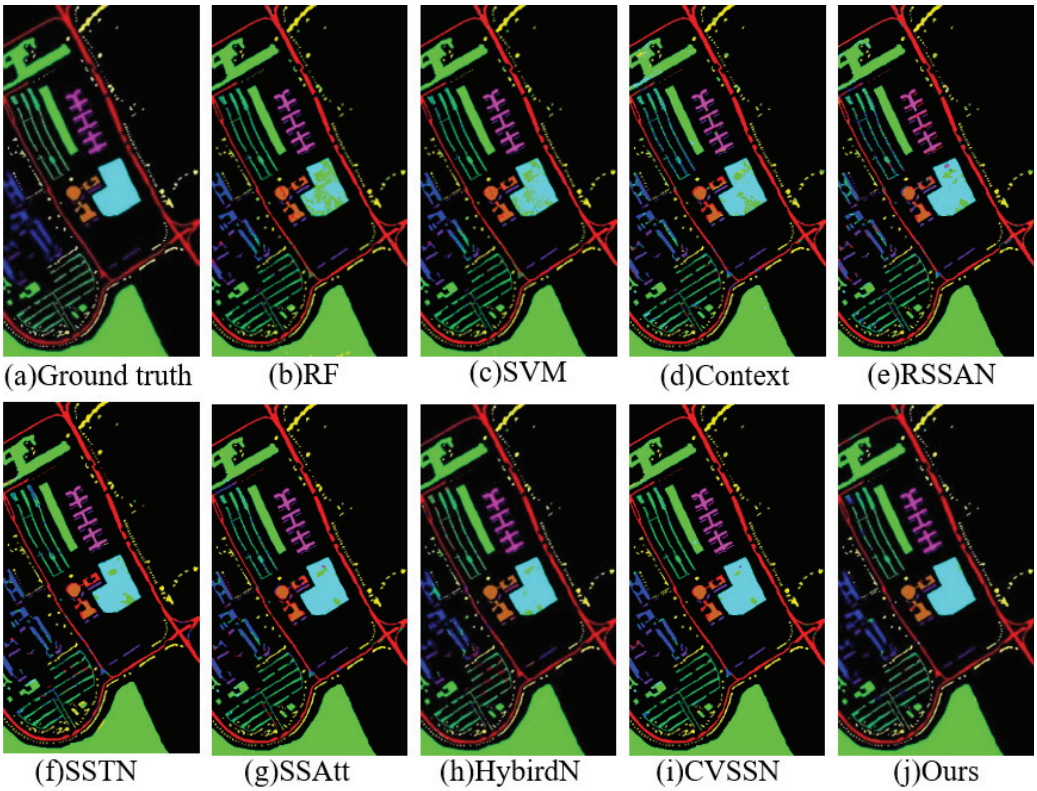


Figure 11. Visualization of different models on UP dataset.

The classification performance of the RF and SVM machine learning models on the four datasets is highly susceptible to other factors. For the IP dataset, SSTN and SSAtt are two classification methods that have more errors in the classification of land features with few samples. The HybridN method has been improved compared to other methods, but it is prone to generating incorrect results at the boundary of the two categories. For the UP dataset, SSTN, SSAtt and HybridN are three classification methods that have large errors in the classification of wheat stubble represented by sky blue. Since the UP dataset is an image captured in a university area, the land features are often distributed in narrow and elongated strips, which are prone to generating errors at the edges. The DACINet proposed in this section effectively solves the problem of classification errors in the UP dataset classification task. For the SA dataset, the visual graph results of classification methods such as RF, SSAtt and HybridN have obvious errors. The DACINet, while maintaining the advantages of other methods, significantly improves the classification accuracy of each category. For the HU dataset, the classification task is more challenging due to the complex urban scenes with diverse land cover categories. As shown in the classification visualization results, traditional machine learning methods RF and SVM produce substantial misclassifications, particularly in categories with similar spectral characteristics such as Grass-healthy, Grass-stressed, and Grass-synth. The Context and RSSAN methods show some improvement but still struggle with categories like Parking-lot1 and Parking-lot2, which have irregular shapes and scattered distributions. SSTN and SSAtt exhibit better performance in homogeneous regions like Water and Tree, yet they generate noticeable errors in complex categories such as Residential and Commercial, where mixed pixels are prevalent. A2S2K, as one of the advanced methods, achieves relatively good results but still fails to accurately classify challenging categories like Tennis-court and Running-track with limited samples. In contrast, the proposed DACINet significantly reduces misclassifications across all categories, achieving the most accurate and complete classification maps. The experimental results show that the DACINet network can better

complete the classification task on the four datasets. Therefore, by assigning greater weight to spectral–spatial features via the dual attention channel, the context interaction facilitates better feature fusion. Combined with the hybrid network, this approach can alleviate classification errors at object boundaries and edge blurring, thereby improving the overall classification performance of the model.

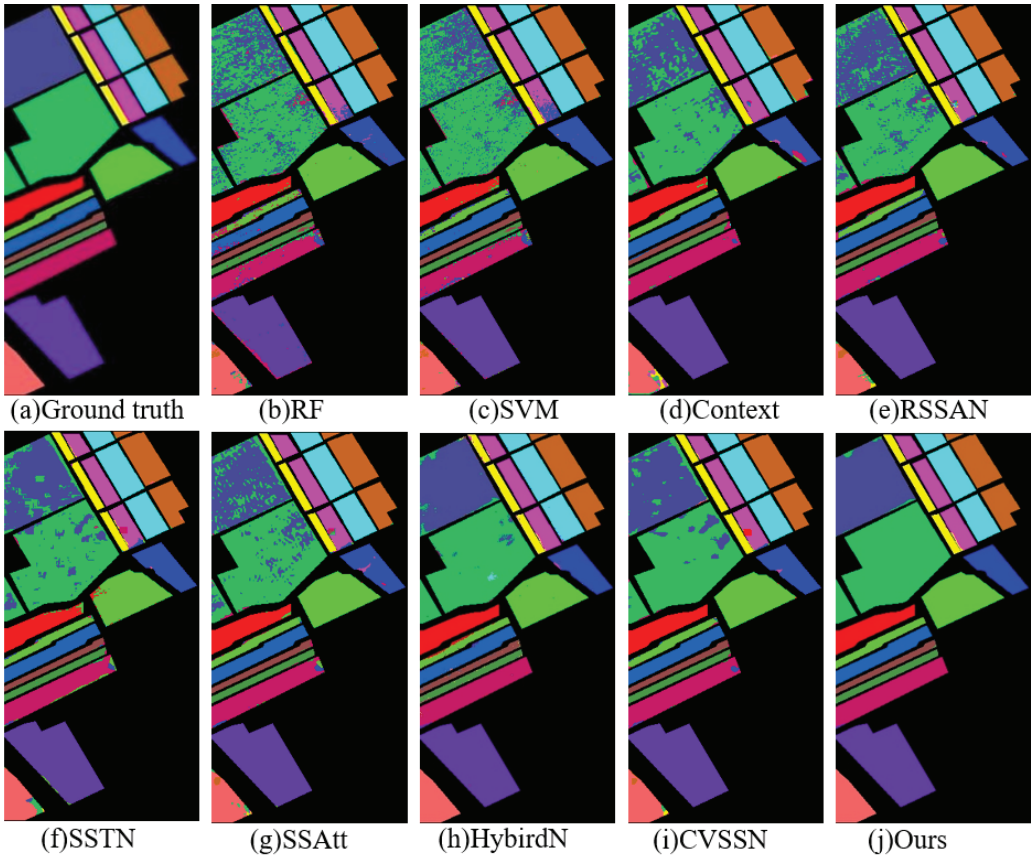


Figure 12. Visualization of different models on SA dataset.

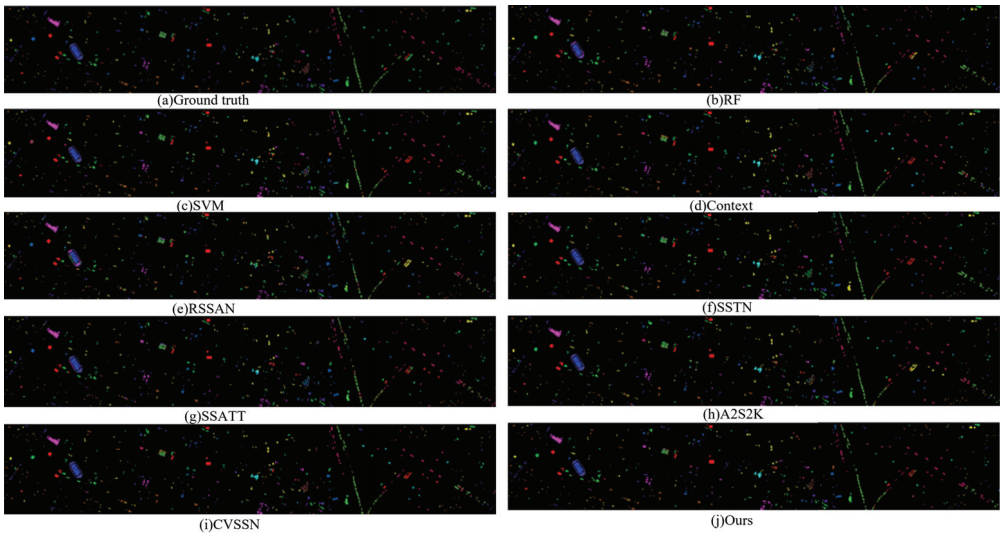


Figure 13. Visualization of different models on HU dataset.

5. Discussion

The DACINet was designed to address a central challenge in HSI classification: jointly modeling spectral and spatial information while capturing long-range contextual depen-

dencies. Standard CNNs, constrained by local receptive fields, struggle to aggregate information from distant pixels or spectral bands. Our framework tackles this by integrating contextual modeling, dual attention enhancement, and a hybrid convolutional strategy. The result is a model that learns features that are both spectrally discriminative and spatially coherent, leading to its consistent performance across diverse datasets.

The CIFM extends standard 3D CNNs by stacking layers with cross-layer residual connections. This design progressively expands the effective receptive field, allowing the model to draw information from larger spectral–spatial neighborhoods. The residual connections then integrate features across these different scales, a capability crucial for distinguishing classes that are spectrally similar but differ in their spatial context. The innovation of the CSDA module lies in its parallel and interactive computation. Unlike sequential mechanisms like CBAM, where channel attention biases spatial processing, or decoupled approaches like DBDA that lack cross-branch interaction, CSDA computes both attention maps from the same 3D feature map and fuses them via addition. This enables complementary learning between spectral and spatial features, aligning with the intrinsic structure of HSI data. The hybrid convolutional layer balances performance and efficiency through its sequential 3D-to-2D design. Initial 3D layers build a rich spectral–spatial representation, while the subsequent reshape encodes spectral information into the channel dimension, allowing lighter 2D convolutions to refine spatial patterns. This explains the high accuracy and low complexity of the DACINet, which achieves far fewer FLOPs than other attention-based methods while maintaining strong performance.

Despite its robust overall performance, the accuracy of the DACINet remains constrained in categories with extremely few training samples. This limitation arises because, like most deep learning frameworks, the model struggles to learn sufficiently discriminative representations from only a handful of examples—especially in datasets with severe class imbalance. A natural direction for future work is therefore to integrate few-shot learning strategies into the framework. One promising approach involves embedding meta-learning into the DACINet, where the model is trained across many episodes to learn a more generalizable metric space. This would enable the network to compare new samples against a small support set and make predictions based on feature similarity, rather than relying solely on class statistics learned from abundant data.

6. Conclusions

This paper proposes a Double-Attention Context Interactive Network (DACINet) for hyperspectral image classification. The DACINet is mainly composed of a CIFM, a CSDA mechanism and a hybrid convolutional layer. The CIFM captures correlations between long-distance image bands through cross-layer residual connections to enhance contextual interaction. The CSDA is a new dual-channel attention mechanism for 3D features, which enhances 2D spatial and 1D spectral features, respectively, and fuses them to strengthen 3D associations. The hybrid convolutional layer combines 2D and 3D convolution to further enhance the discriminability of spectral information. Experiments are carried out on the datasets IP, UP, SA and HU to verify the performance of the proposed DACINet. The results show that the proposed DACINet is superior to other state-of-the-art methods. Ablation studies validate the effectiveness of each core component, while visualization analysis reveals the model's capability for the modeling of spectral–spatial features. In the future, more attention will be paid to few-shot hyperspectral image classification.

Author Contributions: Investigation, M.W.; Resources, Y.Z.; Writing—original draft, N.H.; Writing—review & editing, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grants from the National Natural Science Foundation of China (42271093), the Natural Science Foundation of Shandong Province (ZR2024QF060, ZR2025QC1571), and the Science and Technology Support Plan for Youth Innovation of Colleges and Universities of Shandong Province of China (2025KJH134).

Data Availability Statement: The data presented in this study are available in public hyperspectral remote sensing scene repositories. These data were derived from the following resources available in the public domain: 1. Indian Pines (IP) dataset: https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes#Indian_Pines (accessed on 30 July 2025) (alternative domestic mirror: https://opendatalab.org.cn/OpenDataLab/Indian_Pines (accessed on 30 July 2025)); 2. Pavia University (UP) dataset: https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes#Pavia_University (accessed on 30 July 2025) (alternative domestic mirror: https://hf-mirror.com/datasets/danaroth/pavia (accessed on 30 July 2025)); 3. Salinas (SA) dataset: https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes#Salinas (accessed on 28 July 2025) (alternative domestic mirror: https://opendatalab.org.cn/OpenDataLab/Salinas (accessed on 28 July 2025)); and 4. Houston 2013 (HU) dataset: https://www.grss-ieee.org/community/technical-committees/2013-ieee-grss-data-fusion-contest/ (accessed on 10 February 2026) (alternative domestic mirror: https://drive.uc.cn/s/3fe4f55a213f4?public=1 (accessed on 10 February 2026)). All datasets used in the experiments are publicly accessible without restrictions. The original data files, including hyperspectral images and ground truth labels, can be downloaded directly from the provided official URLs for research reproduction.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Imani, M.; Ghassemian, H. An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges. *Inf. Fusion* **2020**, *59*, 59–83. [CrossRef]
2. Wang, X.; Liu, J.; Chi, W.; Wang, W.; Ni, Y. Advances in Hyperspectral Image Classification Methods with Small Samples: A Review. *Remote Sens.* **2023**, *15*, 3795. [CrossRef]
3. Zhang, Y.; Li, W.; Zhang, M.; Qu, Y.; Tao, R.; Qi, H. Topological structure and semantic information transfer network for cross-scene hyperspectral image classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *34*, 2817–2830.
4. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]
5. Xue, Z.; Zhou, Y.; Du, P. S3Net: Spectral–spatial Siamese network for few-shot hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5531219. [CrossRef]
6. Hu, L.; He, W.; Zhang, L.; Zhang, H. Cross-Domain Meta-Learning Under Dual Adjustment Mode for Few-Shot Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5526416. [CrossRef]
7. Wang, Z.; Zhao, S.; Zhao, G.; Song, X. Dual-Branch Domain Adaptation Few-Shot Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5506116. [CrossRef]
8. Thoreau, R.; Achard, V.; Risser, L.; Berthelot, B.; Briottet, X. Active Learning for Hyperspectral Image Classification: A comparative review. *IEEE Geosci. Remote Sens. Mag.* **2022**, *10*, 256–278. [CrossRef]
9. Fang, L.; Liu, G.; Li, S.; Ghamisi, P.; Benediktsson, J.A. Hyperspectral image classification with squeeze multibias network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 1291–1301. [CrossRef]
10. He, Y.; Tu, B.; Liu, B.; Li, J.; Plaza, A. HSI-MFormer: Integrating Mamba and Transformer Experts for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5621916. [CrossRef]
11. Liang, L.; Xie, P.; Zhang, Y.; Li, J.; Zhang, Z.; Li, J.; Plaza, A. DBMLLA: Double-branch Mamba-like linear attention network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5524315. [CrossRef]
12. Zhang, T.; Xuan, C.; Cheng, F.; Tang, Z.; Gao, X.; Song, Y. CenterMamba: Enhancing semantic representation with center-scan mamba network for hyperspectral image classification. *Expert Syst. Appl.* **2025**, *287*, 127985. [CrossRef]
13. Ahmad, M.; Khan, A.M.; Mazzara, M.; Distefano, S.; Ali, M.; Sarfraz, M.S. A Fast and Compact 3-D CNN for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5502205. [CrossRef]
14. Ghaderizadeh, S.; Abbasi-Moghadam, D.; Sharifi, A.; Zhao, N.; Tariq, A. Hyperspectral Image Classification Using a Hybrid 3D-2D Convolutional Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7570–7588. [CrossRef]
15. Alkhatib, M.Q.; Al-Saad, M.; Aburaed, N.; Almansoori, S.; Zabalza, J.; Marshall, S.; Al-Ahmad, H. Tri-CNN: A Three Branch Model for Hyperspectral Image Classification. *Remote Sens.* **2023**, *15*, 316. [CrossRef]

16. Gündüz, A.; Orman, Z. Hyperspectral image classification using a hybrid RNN-CNN with enhanced attention mechanisms. *J. Indian Soc. Remote Sens.* **2025**, *53*, 613–629. [CrossRef]
17. Yang, J.; Du, B.; Xu, Y.; Zhang, L. Can Spectral Information Work While Extracting Spatial Distribution?—An Online Spectral Information Compensation Network for HSI Classification. *IEEE Trans. Image Process.* **2023**, *32*, 2360–2373. [CrossRef]
18. Şakaci, S.A.; Urhan, O. Spectral-Spatial Classification of Hyperspectral Imagery with Convolutional Neural Network. In Proceedings of the 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), Istanbul, Turkey, 15–17 October 2020; pp. 1–4. [CrossRef]
19. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6232–6251. [CrossRef]
20. He, M.; Li, B.; Chen, H. Multi-scale 3D deep convolutional neural network for hyperspectral image classification. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 3904–3908. [CrossRef]
21. Roy, S.K.; Manna, S.; Song, T.; Bruzzone, L. Attention-Based Adaptive Spectral–Spatial Kernel ResNet for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 7831–7843. [CrossRef]
22. Krichen, M. Generative Adversarial Networks. In Proceedings of the 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 6–8 July 2023; pp. 1–7. [CrossRef]
23. Paul, B.; Fattah, S.A.; Rajib, A.; Saquib, M. SSGRAM: 3-D Spectral-Spatial Feature Network Enhanced by Graph Attention Map for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5516715. [CrossRef]
24. Zhang, Z.; Huang, L.; Tang, B.H.; Wang, Q.; Ge, Z.; Jiang, L. Non-Euclidean Spectral-Spatial feature mining network with Gated GCN-CNN for hyperspectral image classification. *Expert Syst. Appl.* **2025**, *272*, 126811. [CrossRef]
25. Yang, J.; Wu, C.; Du, B.; Zhang, L. Enhanced Multiscale Feature Fusion Network for HSI Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10328–10347. [CrossRef]
26. Liu, R.; Liang, J.; Yang, J.; Hu, M.; He, J.; Zhu, P.; Zhang, L. DHSNet: Dual Classification Head Self-Training Network for Cross-Scene Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5534515. [CrossRef]
27. Wang, D.; Hu, M.; Jin, Y.; Miao, Y.; Yang, J.; Xu, Y.; Qin, X.; Ma, J.; Sun, L.; Li, C.; et al. HyperSIGMA: Hyperspectral Intelligence Comprehension Foundation Model. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 6427–6444. [CrossRef]
28. Zhang, Y.; Wang, X.; Jiang, X.; Zhang, L.; Du, B. Elastic Graph Fusion Subspace Clustering for Large Hyperspectral Image. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 6300–6312. [CrossRef]
29. Huang, S.; Zeng, H.; Chen, H.; Zhang, H. Spatial and Cluster Structural Prior-Guided Subspace Clustering for Hyperspectral Image. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5511115. [CrossRef]
30. Jiang, G.; Zhang, Y.; Wang, X.; Jiang, X.; Zhang, L. Structured anchor learning for large-scale hyperspectral image projected clustering. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *35*, 2328–2340. [CrossRef]
31. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2023**, arXiv:1706.03762. [CrossRef]
32. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929. [CrossRef]
33. Arshad, T.; Zhang, J. Hierarchical Attention Transformer for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 5504605. [CrossRef]
34. Zhao, Z.; Xu, X.; Li, S.; Plaza, A. Hyperspectral Image Classification Using Groupwise Separable Convolutional Vision Transformer Network. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5511817. [CrossRef]
35. Jing, C.; Sun, G.; Zhang, A.; Fu, H.; Cheng, J.; Shi, Z. A Dynamic Attention Unet Network for Hyperspectral Image Classification. In Proceedings of the IGARSS 2025—2025 IEEE International Geoscience and Remote Sensing Symposium, Brisbane, Australia, 3–8 August 2025; pp. 8458–8461. [CrossRef]
36. Li, R.; Zheng, S.; Duan, C.; Yang, Y.; Wang, X. Classification of hyperspectral image based on double-branch dual-attention mechanism network. *Remote Sens.* **2020**, *12*, 582. [CrossRef]
37. Nhaila, H.; Elmaizi, A.; Sarhrouni, E.; Hammouch, A. Supervised classification methods applied to airborne hyperspectral images: Comparative study using mutual information. *Procedia Comput. Sci.* **2019**, *148*, 97–106. . [CrossRef]
38. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [CrossRef]
39. Lee, H.; Kwon, H. Going deeper with contextual CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **2017**, *26*, 4843–4855. [CrossRef] [PubMed]
40. Roy, S.K.; Krishna, G.; Dubey, S.R.; Chaudhuri, B.B. HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 277–281.

41. Li, M.; Liu, Y.; Xue, G.; Huang, Y.; Yang, G. Exploring the relationship between center and neighborhoods: Central vector oriented self-similarity network for hyperspectral image classification. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *33*, 1979–1993. [CrossRef]
42. Zhu, M.; Jiao, L.; Liu, F.; Yang, S.; Wang, J. Residual spectral–spatial attention network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 449–462.
43. Zhong, Z.; Li, Y.; Ma, L.; Li, J.; Zheng, W.S. Spectral–spatial transformer network for hyperspectral image classification: A factorized architecture search framework. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5514715.
44. Hang, R.; Li, Z.; Liu, Q.; Ghamisi, P.; Bhattacharyya, S.S. Hyperspectral image classification with attention-aided CNNs. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 2281–2293. [CrossRef]
45. Xu, J.; Li, K.; Li, Z.; Chong, Q.; Xing, H.; Xing, Q.; Ni, M. Fuzzy graph convolutional network for hyperspectral image classification. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107280. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

remote sensing

MDPI

Systematic Review

Retrieval of Multiple Variables from Hyperspectral Data: A PRISMA-Aligned Systematic Review of Classical Physics-Based Machine Learning and Hybrid Algorithms in Vegetation and Raw Materials Application Domains

Andrea Taramelli [1,2], Sara Liburdi [1,*], Alessandra Nguyen Xuan [2], Simone Mancon [3], Serena Sapio [1,4] and Emiliana Valentini [1,4]

[1] Department of Science, Technology and Society (STS), University School for Advanced Studies IUSS Pavia, Palazzo del Broletto, Piazza della Vittoria 15, 27100 Pavia, Italy; andrea.taramelli@iusspavia.it (A.T.); serena.sapio@iusspavia.it (S.S.); emiliana.valentini@cnr.it (E.V.)
[2] Institute for Environmental Protection and Research (ISPRA), Via Vitaliano Brancati 48, 00144 Roma, Italy; alessandra.nguyenxuan@isprambiente.it
[3] ARESYS s.r.l., Via Luigi Cadorna 66, 20055 Vimodrone, Milan, Italy; simone.mancon@aresys.it
[4] National Research Council of Italy, Institute of Polar Sciences, Via Salaria km 29,300, 00015 Montelibretti, Italy
* Correspondence: sara.liburdi@uniroma1.it

Highlights

What are the main findings?

- Machine learning and deep learning approaches dominate vegetation applications, while classical physics-based methods remain prevalent in raw materials, with hybrid models achieving the highest overall performance across domains.
- Across vegetation, mineral and methane mapping, insufficient standardization of accuracy metrics and incomplete methodological reporting represent the main sources of bias rather than algorithmic limitations.

What are the implications of the main findings?

- Hybrid physics-informed machine learning approaches emerge as the most promising pathway to ensure robustness, transferability and operational readiness for upcoming hyperspectral missions.
- The results provide evidence-based guidance for algorithm selection and development aligned with user requirements of current and future hyperspectral missions, such as PRISMA, CHIME, SBG and IRIDE.

Abstract

Hyperspectral (HSI) remote sensing has emerged as a transformative technology for Earth Observation, enabling detailed assessments across different domains. The current PRISMA-aligned systematic review aims to compare classical physics-based algorithms with emerging machine learning (ML), deep learning (DL) and hybrid approaches across two relevant application domains (vegetation and raw materials), analyzing over 350 peer-reviewed studies (194 after the screening) sourced from Scopus and Web of Science and accessed in July 2025. Specific domain-related studies have been considered, excluding duplicates and studies not strictly related to HSI. Risk of bias was assessed qualitatively based on different criteria. The efficiency of the techniques was analyzed by comparing the accuracy metrics reported in the studies. The heterogeneity of the evaluation metrics used across the different categories of the studies and the underrepresentation of some application domains is the final baseline of the work. The results were synthesized, grouping by application

domains and algorithm category: ML and DL models dominate vegetation applications, and physics-based methods remain prevalent in raw materials. Hybrid models achieve the highest performances across all domains. This review highlights the importance of the hyperspectral operational requirements identified for upcoming missions (CHIME, SBG and IRIDE) and points out the opportunity for algorithm development.

Keywords: hyperspectral data; canopy water content; chlorophyll content; mineral mapping; methane mapping; hyperspectral missions; raw materials; vegetation

1. Introduction

Hyperspectral remote sensing has transformed how we observe and understand Earth, offering a window into its radiometric response complexity through hundreds of contiguous bands. This extraordinary level of detail enables applications ranging from raw material exploration [1] and vegetation biophysical assessment [2] to mapping atmospheric gases like methane [3]. Products derived from hyperspectral imaging are essential for tackling pressing challenges like climate change, sustainable resource management, and ecosystem conservation. Despite its remarkable capabilities, unlocking the full potential of hyperspectral data is far from straightforward [4]. Classical/physics-based (C/PB) algorithms, such as Spectral Angle Mapper (SAM), have long served as the backbone of hyperspectral data analysis. While effective for basic tasks, these methods often falter when confronted with the increasing volume, variety, and complexity of new hyperspectral datasets. This is highlighted by the advent of new hyperspectral missions: (I) Copernicus Hyperspectral Imaging Mission for the Environment (CHIME), developed by the European Space Agency as part of the Copernicus programme; (II) PLATiNO hyperspectral constellation of the IRIDE Earth Observation programme (a constellation of constellations funded through the Italian National Recovery and Resilience Plan); and (III) NASA's Surface Biology and Geology (SBG) mission. The relevance of these imminent missions lies in their design to yield operational, user-driven, and higher-level products. Notably, CHIME will furnish an operational mapping of vegetation traits and raw materials domains, derived from a preliminary user needs analysis. The mission outcomes will support food security, particularly in the domains of soil and vegetation variable assessment and mineral resource exploration. The SBG mission will be focused on global surface composition, with the aim of enabling supporting applications through an increased revisit frequency, particularly for mineral mapping and land monitoring. Finally, the IRIDE-PLATiNO will contribute high-resolution imaging capabilities, within a multi-mission system architecture, to obtain dynamic products, mainly related to land cover and land use. These missions highlight the need for robust, transferable and physically consistent algorithmic solutions.

It is imperative to emphasize that hyperspectral data analysis is accompanied by numerous challenges, which necessitate the development of sophisticated retrieval algorithms and an in-depth data examination and evaluation. The high dimensionality of hyperspectral imagery (HSI) gives rise to spectral collinearity, noise sensitivity and mixed-pixel effects, thereby rendering feature extraction and model generalization more difficult. Additionally, scene variability, atmospheric correction uncertainties and sensor-specific artefacts further exacerbate cross-sensor transferability. Furthermore, the escalating volume and intricacy of HSI produced by forthcoming missions necessitate scalable, reproducible and physically consistent models [5–8]. Machine learning (ML) offers tools to deal with these large and diverse datasets but can also be used to uncover complex patterns and relationships within them. Furthermore, ML algorithms have demonstrated their potential

https://doi.org/10.3390/rs18050798

to address the shortcomings of classical methods, such as sub-pixel analysis algorithms, like the Linear Spectral Mixture Analysis (LSMA) or spectral matching (SM) [9]. Moreover, hybrid approaches that integrate ML with C/PB methodologies further enhance analytical precision and adaptability, paving the way for more robust solutions [10,11]. The current and upcoming challenges facing HIS analysis delineate the need for a systematic comparison of C/PB, ML and hybrid methods. This will help to explain how these algorithms, operating in isolation or in combination, can overcome these challenges.

This review sets out to understand which category of algorithms is the most efficient and suitable for hyperspectral data analysis considering C/PB algorithms, ML algorithms and hybrid (ML + C/PB) models. The aim is to identify existing gaps and opportunities, aligning them with the user needs of future hyperspectral missions, like CHIME and PLATiNO, providing a comprehensive overview of current methodologies [5,7]. This study focuses on two application domains: vegetation biophysical assessment and raw materials exploration. The selection of these two application domains is underpinned by two considerations. Firstly, these are considered the most relevant in the context of the current policy framework [7]. In particular, in the raw materials domain, included are strategic priorities related to industrial competitiveness, supply chain resilience, and strategic autonomy and security of resources. Regulatory frameworks, such as the Critical Raw Material Act (CRMA), explicitly identify a set of critical and strategic materials, which are essential in the context of the green and digital transition [12]. The Global Methane Pledge calls for a reduction in methane emissions as a short-term priority, requiring the capability to detect, map and analyze methane emissions, in particular from localized source points [13]. Although mineral and methane mapping are very different physical phenomena, they share common methodological challenges related to spectral discrimination, noise sensitivity, spatial heterogeneity and the need for robustness under operational conditions. For this reason, they are both considered in the raw materials domain in this work. In parallel, the vegetation domain is primarily related to climate adaptation, biodiversity protection and ecosystem resilience policies from a broader perspective to food security [14]. Schiavon et al. [7] collected the requirements from a user's communities to analyze the potential of hyperspectral satellites for responding to their needs. Most of these requirements were related to agriculture and ecosystem assessment. Despite this, these application domains show gaps due the current technical limits in hyperspectral data processing [7]. Secondly, both application domains are deemed the most suitable for further study and investigation given prevailing user requirements, mission objectives and the availability and focus of hyperspectral data. Furthermore, these domains are also included in CHIME's Mission Requirements Document [15]. In fact, this future Copernicus hyperspectral mission aims to satisfy the large community of users and address many needs in two main application domains, sustainable agriculture and food security, also including biodiversity protection, forest vegetation and related environmental degradation and hazards; the second one is raw materials, including, in particular, mine environment management [16]. The work of Berger et al. [17] systematically linked, for the first time in the European Union (EU), land-related agricultural and environmental policies to EO-derived variables that can be generated from upcoming satellite missions in the next ten years, such as CHIME. In this work, the authors analyzed the current Technologic Readiness Level (TRL) reached for each remotely sensed variables with proven added values, such as mineral abundances and compositions, leaf pigment content and vegetation water content, which support the potential for targeting critical raw materials and achieving the EU's biodiversity objectives and greenhouse gas emission reduction target, respectively. They revealed that it will be crucial to improve scientific algorithms and enhance the accuracy of the cited variables to develop mature products tailored to stakeholders and policymakers.

The analysis starts from the evaluation of user requirements considering the current and new hyperspectral missions' potential to implement algorithm innovations. Ultimately, this review aims to guide the hyperspectral research community toward scalable, innovative, and user-focused solutions that meet the challenges of a multi-satellite mission [12].

2. Background on Hyperspectral Data and Missions

2.1. Hyperspectral Data Background

Initiatives such as the European Critical Raw Material Act (CRMA) [12], Paris Agreement [18], and European Sustainable Agenda [19] provide the framework to underline and promote improvements in current technologies and methodologies related to the Earth Observation domain applied to environmental monitoring [7]. Hyperspectral imaging (HSI) remote sensing technology provides innovations, collecting surface reflectance information in multiple and continuous wavebands [20]. The potential of hyperspectral data is related to their application in different fields, such as raw material mapping, vegetation biophysical and biodiversity assessment and GHG mapping (e.g., methane). The strength of this type of data is related to the spectral resolution: hundreds of contiguous bands, comprising the spectral range between VNIR-SWIR from 400 to 2500 nm [21–23]. These types of characteristics can be used to better analyze the spectral composition of pixels, capturing detailed spectral information across the bands and enabling the identification and analysis of materials and processes with unmatched precision.

HSI can be acquired using different platforms: (1) satellites [24], (2) airborne [25], (3) Unmanned Aerial Vehicles (UAVs) [26] and (4) field instruments such as field spectroradiometers [27]. Unlike traditional multispectral systems, which divide the electromagnetic spectrum into a few broad bands, HSI resolves hundreds of narrow, contiguous bands, spanning a wide spectral range.

For this study, the focus is mainly on the Visible–Near Infrared (VNIR) range (400–1000 nm) and the Shortwave Infrared (SWIR) range (1000–2500 nm), which are both critical for numerous environmental and resource-based applications [28]. Table 1 provides a non-exhaustive summary of the most relevant spectral ranges for the two application domains selected in this review [29].

Table 1. Spectral features relevant for the application domain. Several products have been identified for each domain and are defined as "High Priority Products" or "Additional products", according to their classification in CHIME's Mission Requirements Document.

Application Domain	Spectral Range (nm)	Spectral Region	High Priority Products/Additional Products
Vegetation biophysical assessment	670, 710–730	Red Band, Red-Edge	Canopy Chlorophyll Content
	950–970, 1150–1260, 1450, 1950, 2250	NIR, SWIR	Canopy Water Content
Raw Materials	400–1000, 1000–2500, 8000–12,000	VNIR, SWIR, LWIR	Minerals (e.g., Fe-bearing minerals, Clay minerals, Carbonates, Silica)
	1650–2350	SWIR	Methane

2.2. Hyperspectral Missions Overview

This study leverages a diverse set of hyperspectral data sources, which comprise operational, decommissioned, and planned missions, across satellite, airborne, and UAV platforms. These platforms provide unique combinations of spatial, spectral, and temporal characteristics, enabling their application in different domains. Table 2 reports the most commonly used hyperspectral sensors, including information on the status of the mission and technical characteristics of sensors.

https://doi.org/10.3390/rs18050798

Table 2. Most commonly used and studied hyperspectral sensors: missions status with technical specifications (partially based on [7]).

Hyperspectral Mission/Platform and Owner	Type	Status and Year of Launch	Spatial Resolution	Revisit Time	Spectral Bands	Spectral Range (nm)	Spectral Resolution (nm)
PRISMA, HyperSpectral Precursor of the Application Mission (ASI-Italy)	Satellite	Operational (2019)	30 m	29 days On-Demand	66 VNIR-174 SWIR	400–2500	~12 nm
EnMAP (DLR-GFZ Germany)	Satellite	Operational (2022)	30 m	27 days	244	420–2450	6.5 nm (VNIR), 10 nm (SWIR)
EMIT (Nasa-USA)	Spaceborne (ISS)	Operational (2022)	60 m	Targeted areas for dust regions	285	380–2500	7.5 nm
DESIS (DLR-Germany)	Spaceborne (ISS)	Operational (2018)	30 m	On demand	235	402–1000	~2.55 nm
Hyperion (EO-1) (Nasa-USA)	Satellite	Decommissioned (2017)	30 m	16 days	70 VNIR-150 SWIR	426–2395	~10 nm
CHIME, Copernicus Hyperspectral Imaging Mission for the Environment (ESA-Europe)	Satellite	Planned (2028)	20–30 m	12–15 days	400+ (Planned)	400–2500	~10 nm (planned)
SBG, Surface Biology and Geology (Nasa-USA)	Satellite	Planned (2027–2028)	30 m	3 days	300+ (Planned)	380–2500	5–10 nm (targeted)
IRIDE, Hyperspectral constellation (PLATiNO) (ASI-ESA Italy-)	Satellite	Planned (2025–2026)	5 m PAN, 21 m SPOTLIGHT, 31 m	27 days	60 VNIR-256 SWIR	400–2500	10 nm (targeted)
AVIRIS (Nasa-JPL USA)	Airborne	Operational (1980s updated versions)	4–20 m	On demand	224 VNIR	380–2500	5–10 nm
HyMap (Commercial)	Airborne	Operational (1990s ongoing)	5–10 m	On demand	128	400–2500	15 nm
Specim FX17 (Specim-Finland)	UAV	Operational (2018)	0.5–2 m	On demand	224 SWIR	900–1700	8 nm

2.3. Hyperspectral Data: Gaps and User Requirements

Despite the potential of HSI, there are notable gaps between the capabilities of current missions and the operational needs expressed by diverse user communities [7]. Research conducted by the Italian Copernicus User Forum (CUF) [7] and the SBG science community (Surface Biology and Geology) [30] has identified high-priority needs for improved spatial resolution ($\leq$30 m), frequent revisit times ($\leq$2 weeks), and better exploitation of the VNIR-SWIR spectral range (400–2500 nm). Related to this, recent studies, such as those conducted for ESA's CHIME mission, underscore the importance of aligning mission planning and algorithm development with specific user requirements [15]. As shown in Figure 1, the current operational satellite sensors are unable to fully meet user requirements. Whilst UAV and airborne-based sensors can satisfy the requested spatial resolution, their number of bands is limited, and acquisition is only on demand. Conversely, the satellite sensors currently in orbit are close to meeting the requirements, both in terms of spatial resolution (almost always 30 m) and temporal resolution. Future satellite missions (i.e., CHIME and PLATiNO-IRIDE) will fully meet user demands, increasing the spatial resolution and the number of bands [17].

https://doi.org/10.3390/rs18050798

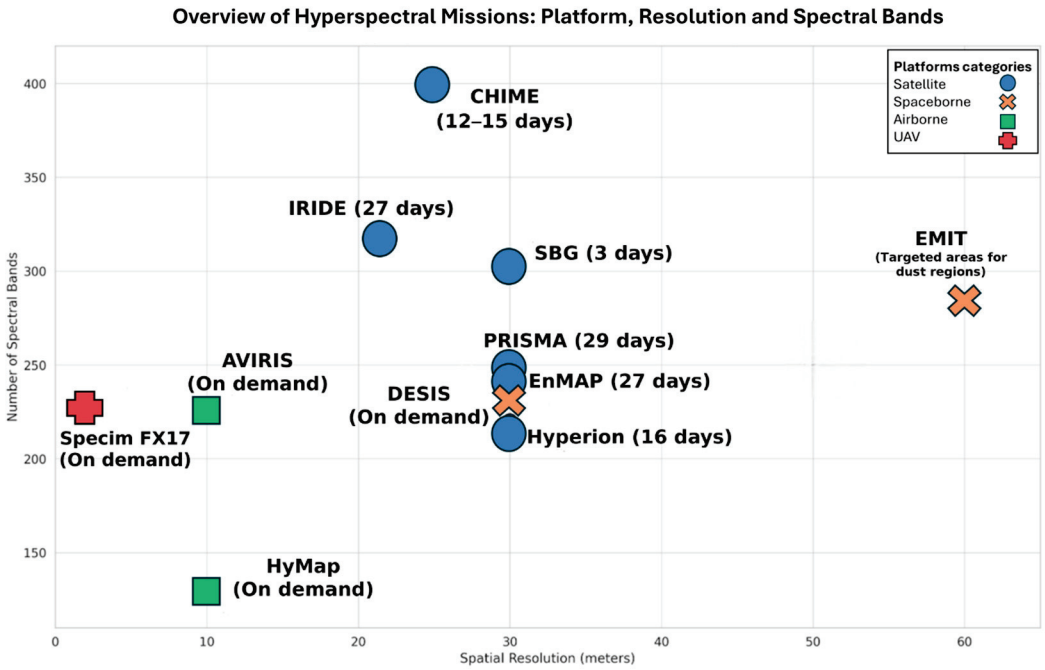


Figure 1. Most commonly used and studied hyperspectral sensors, distributed per number of spectral bands and spatial resolution. The symbols represent the type of platform, and, in brackets, the revisit time is reported.

These gaps, ranging from limited spatial resolution and temporal revisit time to underutilized spectral ranges (Figure 1), highlight the need for a concerted effort to enhance (1) data processing algorithms, (2) multi-mission synergy and (3) product retrieval [16]. Central to addressing these gaps is a user-centric approach that incorporates feedback from diverse stakeholders.

3. Algorithmic Framework for Hyperspectral Data

3.1. Algorithms for Hyperspectral Data Analysis in Different Application Domains

Regarding the development of data processing algorithms, user requests are different according to different applications [4]. Simple C/PB methodologies, such as Matched Filtering (MF) [31] and Continuum Removal [32], are computationally efficient and well-suited for specific, well-characterized applications but often struggle with high-dimensional data and overlapping spectral features. ML approaches, including Random Forests, Support Vector Machines (SVMs), and deep learning architectures like Convolutional Neural Networks (CNNs) or Mixture Density Networks (MDNs), offer the flexibility to handle complex datasets and extract nuanced patterns [11]. Recent deep architectures, specifically designed for hyperspectral image classification and segmentation, further confirm this trend, by combining multi-scale spectral–spatial encoding with attention mechanisms. Examples include multiarea target attention networks for classification, multi-scale convolutional and multi-attention models, and recent semantic segmentation frameworks such as SegHSI, which achieve accurate pixel-wise labeling with limited annotated samples [33]. Although these methods are not all directly applied to the specific products considered in this review, they illustrate the rapid evolution of deep learning-based HSI processing and point to algorithmic directions that could benefit future HPP retrievals. However, they rely heavily on the availability of large, high-quality training datasets, which are often scarce in the hyperspectral domain. A key contribution of this review lies in its focus on bridging the gaps identified in recent user requirement studies. The high number of available and future hyperspectral missions, and so the increasing number of data, will be a valuable

https://doi.org/10.3390/rs18050798

opportunity to better investigate the multi-mission synergy and the requirements from user communities so that all services and specifications can be developed. This strategy can satisfy the three main user requirements related to hyperspectral data: higher spectral resolution, higher spatial resolution, higher temporal resolution related to revisit time [7]. Some of these gaps will be filled partially by new hyperspectral missions but also through synergy with multispectral and radar missions [34,35]. Such synergies also enable the development of specific algorithms, which are fundamental for ensuring interoperability across missions [16].

To advance in Earth Observation technologies, it is important to match the technical requirements on the specific thematic area with user requirements. Considering the future CHIME mission, the main data products related to the exploitation of hyperspectral data are provided at Level-2A (geometrically corrected and in surface reflectance). In addition to these, there are a set of Higher-Level Data Products that will be a part of the mission services offered by the EO to insert it in the downstream industry for business development. Inserted in Higher-Level Data Products, there are High-Priority Products (HPPs) (e.g., leaf and canopy chlorophyll content and canopy water content) for the vegetation application domain or, in the raw materials context, the composition and abundance of non-renewable raw materials (mainly minerals). Another additional product is large gas leakage detection, not an HPP, but a potential one, all of them leveraged on a specific spectrum portion [15].

For example, in the CHIME mission, HPPs are a set of higher-level products proposed to users as a part of the mission catalogue or services offered by the program. There are many different proposed HPPs for the CHIME mission, divided by application domains. In this review, the HPPs and the additional products considered are related to two of these domains: vegetation and raw materials (mineral mapping) [7,17,26,29].

3.2. Conceptual Differences Between AI, ML, DL, C/PB, and Hybrid

All the algorithms mentioned in the previous paragraph have been defined as C/PB or as ML techniques [24]. To fully understand the comparative framework developed in this review and the conceptual differences between algorithm classes, Table 3 reports the definition of Artificial Intelligence (AI), machine learning (ML), deep learning (DL), classical/physics-based (C/PB) algorithms and hybrid models, highlighting the strength and the limitations of each category [4,9].

Table 3. Algorithm categories description: definition, strengths and limitations of artificial intelligence, machine learning, deep learning, classical/physics-based and hybrid models.

Category	Definition	Key Methods	Strengths	Limitations
Artificial Intelligence (AI)	Development of systems capable of performing tasks requiring human intelligence, combining symbolic and data-driven approaches.	Symbolic AI, Expert Systems, ML	Combines reasoning and learning; general-purpose intelligence framework	Conceptual umbrella; not directly operational
Machine Learning (ML)	Subset of AI; algorithms that learn patterns from data to enable predictive or descriptive modeling.	Supervised, Unsupervised, Semi-supervised, Active, Transfer Learning	Flexible, adaptive, effective on noisy and high-dimensional data	Requires labeled data and tuning; can struggle with physical interpretability
Deep Learning (DL)	Subset of ML; multi-layer neural networks for hierarchical feature learning from raw data.	CNNs, RNNs, autoencoders, Transformers	Learns complex hierarchical features; high automation of feature extraction	High computational cost; needs large datasets

Table 3. *Cont.*

Category	Definition	Key Methods	Strengths	Limitations
Classical/Physics-based Algorithms	Algorithms based on deterministic physical models describing radiation-substrate interactions.	RTMs, MF, SAM, Linear/Non-linear Unmixing	Physically interpretable; transferable across conditions	Limited in handling non-linear relations; computational cost for RTMs
Hybrid Models	Models integrating physical constraints into ML/DL frameworks.	Physics-guided NN, Hybrid inversion (RTM + ML), ML corrections of physical outputs	Combines physical consistency with learning flexibility; improves generalization and uncertainty quantification	Requires careful integration design; still emerging in operational EO

In considering these categories, particular emphasis should be placed on the hybrid model, which involves the incorporation of physical constraints within machine learning and deep learning frameworks. Figure 2 provides a conceptual scheme to better clarify the interconnection and relationship between AI, ML, DL, C/PB and hybrid models.

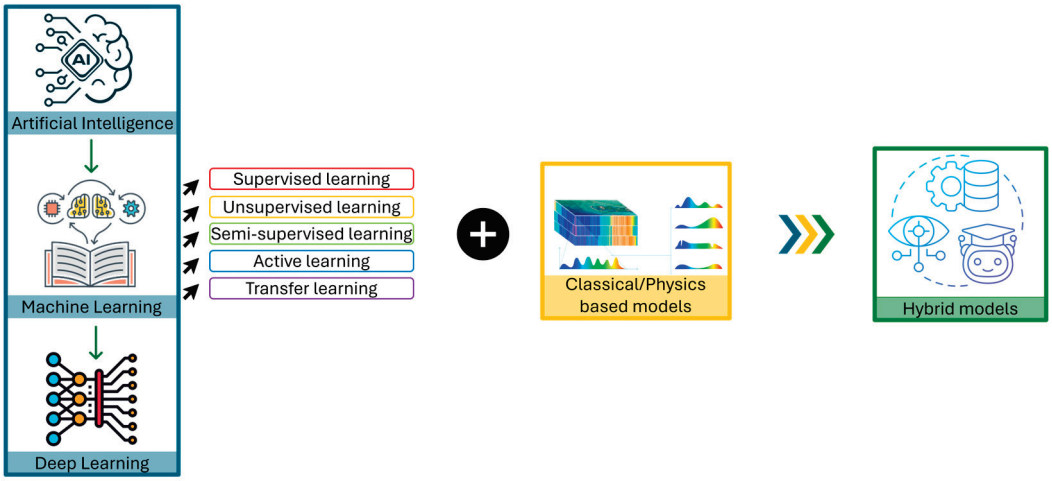


Figure 2. Conceptual scheme linking AI/ML learning framework (blue), classical/physics-based models (yellow) and their fusion into hybrid models (ML + C/PB) (green).

4. CHIME High-Priority Products (HPPs) and Selected Application Domains

HPPs, as mentioned above, are high-level mission deliverables derived from CHIME's core specifications and explicitly aligned with the Mission Requirements Document (MRD), to ensure strategic value for both scientific communities and downstream industry. Each HPP is tuned to CHIME's instrument specification (spectral bands, signal-to-noise ratio, spatial resolution). In this framework, HPPs are able to support the mission not only scientifically but also operationally, ready for integration in different Earth Observation downstream services [36,37].

4.1. Selected CHIME HPPs and Additional Products

4.1.1. Vegetation

Vegetation status assessment and monitoring require hyperspectral data to meet specific user requirements and to ensure accurate and meaningful analysis. These include the following:

- Spectral Requirements: High spectral resolution ($\leq$10 nm) is essential to capture fine variations in spectral signatures, allowing differentiation of species-

specific traits, detection of subtle stress signals, and analysis of biochemical and structural attributes [38];

- Spatial Requirements: High spatial resolution ($\leq$20 m) is critical for mapping ecosystem heterogeneity, particularly in diverse or fragmented landscapes, enabling detailed monitoring of species distributions and habitat conditions [39];
- Temporal Requirements: Frequent temporal resolution, with revisit times of less than 10 days, is necessary to monitor phenological changes, track seasonal dynamics, and assess short-term responses to environmental stressors [40];
- Calibration and Validation: Accurate calibration, georeferencing, and access to field validation data are fundamental for ensuring that hyperspectral observations are reliable and comparable across time and space [41,42];
- Data Accessibility and Processing: Availability of user-friendly platforms and processing tools is vital to allow researchers and users to analyze data effectively, ensuring timely application of insights in biodiversity management and policymaking.

Orbiting satellite hyperspectral sensors, such as PRISMA and EnMAP, partially meet these requirements, enabling transformative advances on different scales [24]. Future satellite missions, such as CHIME, SBG and IRIDE, are developed to fully meet these technical requirements [7,43].

4.1.2. Raw Materials—Mineral Mapping

Hyperspectral imaging has become a revolutionary tool in mineral mapping, in particular due to its ability to detect diagnostic spectral features associated with specific minerals and geological processes and components [44]. The electromagnetic spectrum provides a powerful tool for understanding the characteristics of rocks and minerals through the interpretation of spectral features across different wavelength regions. Sensors like EMIT [45], PRISMA [46], and EnMAP [47] provide high-quality hyperspectral data in the VNIR-SWIR ranges, with improved spatial, spectral, and radiometric performance. EMIT, for example, is particularly suited for detecting surface mineral dust, offering insights into the mineralogical composition of arid and semi-arid regions. PRISMA and EnMAP provide broad applications for mineral alteration mapping and lithological boundary delineation with resolutions suitable for regional and site-specific investigations [48]. With spectral resolutions of around 10–20 nm and spatial resolutions between 20 and 30 m, these sensors are ideal for resolving fine mineral absorption features while capturing regional-scale geological variability [49]. High radiometric accuracy ensures reliable quantitative classifications of minerals, in terms of percentage in the pixel, even in complex terrains [50]. The integration of these advanced sensors into mineral exploration workflows has enabled detailed mapping of mineralogical processes, supported sustainable resource exploration, and improved environmental monitoring in mining regions [47,49,51].

4.1.3. Raw Materials—Methane Mapping

Gas detection is a critical application of hyperspectral imaging, especially for monitoring greenhouse gases (GHGs), such as methane (CH_4) and carbon dioxide (CO_2). This has garnered increased attention in recent times, in particular for climate change-related studies [52]. The SWIR range (1600–2500 nm) is particularly well suited for this purpose, as it encompasses the strong absorption features of these gases. These spectral features enable the identification of gas source points and proxy for the quantification of their concentration, considering the specific absorption feature in the SWIR region [53]. For instance, methane exhibits distinctive absorption bands at approximately 1660–1680 nm and 2300–2400 nm, making the SWIR range indispensable for plume detection, flux quantification, and monitoring multi-source emissions [22,54]. High spectral resolution, typically below 5 nm, is

https://doi.org/10.3390/rs18050798

crucial for resolving these fine absorption features. Additionally, a high signal-to-noise ratio (SNR) is essential for capturing faint gas signatures, especially in areas with mixed or noisy backgrounds.

Spatial resolution of $\leq$50 m is required to pinpoint localized emission sources and effectively track plume dispersion patterns [55], which is critical for assessing industrial emissions, natural gas leaks, and other point sources [54]. Operational hyperspectral sensors, such as PRISMA and EMIT, have demonstrated exceptional capabilities in gas detection and monitoring. PRISMA, with its VNIR-SWIR coverage, has been successfully used to identify methane plumes, offering critical insights into emission dynamics [56]. EMIT, originally focused on mineral dust source mapping, has also shown potential for detecting greenhouse gases in specific applications [57]. These advancements are instrumental in air quality research and climate change mitigation, providing actionable data to reduce greenhouse gas emissions and address environmental challenges [58].

5. Materials and Methods

5.1. Review Design and Reporting

This work presents a systematic review of literature studies dealing with the retrieval of selected HPP products, such as vegetation-related variables and mineral and methane mapping. This review has been conducted using two scientific databases, Scopus and Web of Science (WOS). The algorithms applied in each study were evaluated, considering C/PB, ML, DL algorithms and hybrid models and comparing recent innovations and traditional approaches. This manuscript is aligned with the PRISMA checklist for systematic reviews (Figure 3). This systematic review was conducted by a single reviewer; thus, some PRISMA items were not fully met, specifically those related to duplicate screening and duplicate data extraction. All other steps of the review process adhered to PRISMA recommendations, using a standardized Excel form.

5.2. Data Sources and Search Strategy

The comparison between C/PB algorithms, ML algorithms and hybrid models was carried out following specific criteria, also with the aim to underline the specific role of hyperspectral data in the algorithm's development and application.

The systematic review was conducted following three main steps (Figure 4):

1. Application domains and product selection: Two application domains have been identified, vegetation and raw materials. Both application domains were selected in light of their central role in the CHIME mission framework and their relevance in the current European and international policy priorities, including ecosystem monitoring and Critical Raw Materials Act. For each application domain, the HPPs (marked by *) or additional products investigated are as follows:
    - Vegetation: Canopy Water Content (CWC)*, Vegetation Chlorophyll Content (CC)*;
    - Raw materials: mineral mapping (e.g., kaolinite)*, methane mapping.
2. Algorithms categorization: C/PB algorithms—ML/DL algorithms—hybrid models. Studies were classified as C/PB if variable retrieval was primarily driven by explicit physical models, even if ML is used as a secondary tool for the optimization of parameter or post-processing phase. Conversely, studies are categorized as ML-based when the target variable is directly evaluated through statistical learning retrieval. Finally, when physical models and ML are jointly explicitly integrated within the processing chain, the study is classified as hybrid.
3. Metrics of algorithm evaluation approach. This phase is deeply explained in the following sections. This organization categorizes the methods according to the specific needs of hyperspectral data processing, providing a clear view of existing strategies.

https://doi.org/10.3390/rs18050798

Then, a comparative analysis was carried out to highlight the main limitations of classical approaches and assess how ML/DL can address them in each application domain. To make this analysis clearer, comparative tables were constructed, integrating up-to-date scientific references.

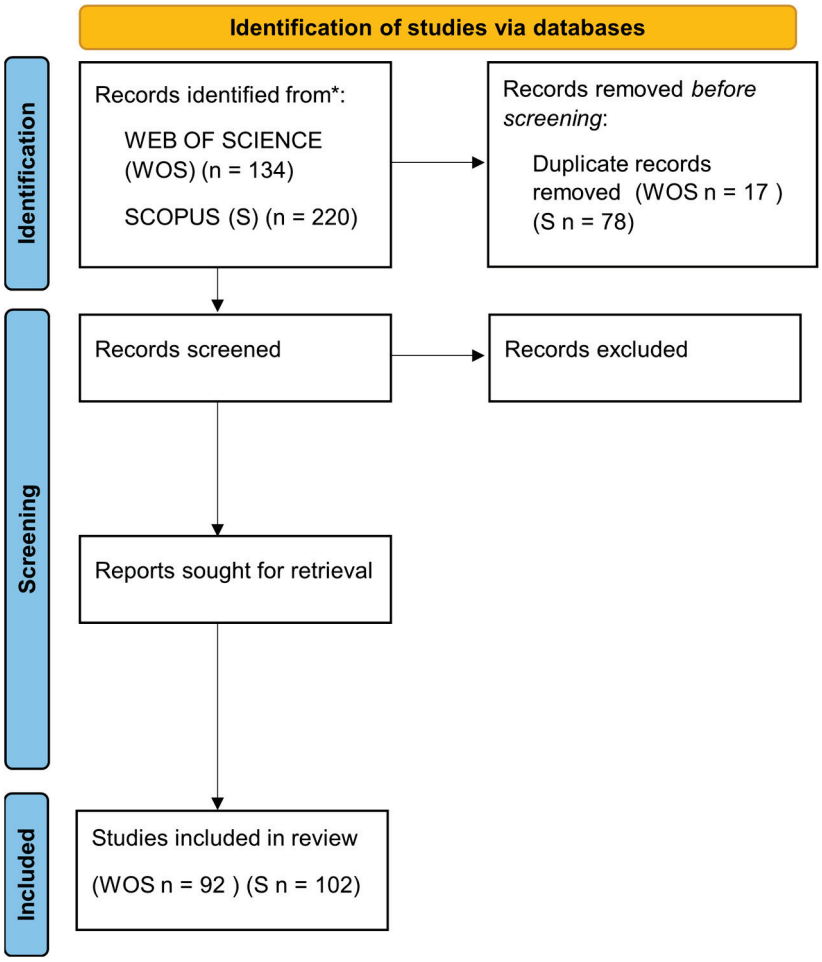


Figure 3. PRISMA 2020 flow diagram illustrating the systematic search and study selection process (Sections 5.1–5.5), including identification, screening, eligibility, and inclusion phases. The diagram reports the number of records retrieved from Scopus and Web of Science, duplicates removed, records screened, and studies included in the final synthesis. The asterisk (*) indicates that the reported number corresponds to the combined total of records retrieved from the two databases considered. Arrows represent the sequential progression of records through the review process. Colors are used solely to visually distinguish the different PRISMA phases.

5.3. Eligibility Criteria and Study Selection

The research strings customized for WoS and SCOPUS databases are reported as Supplementary Materials. A general visualization of used strings for each product is shown in Tables 4 and 5, related to C/PB algorithms and AI/ML/DL/hybrid models, respectively. No sensitivity analysis was conducted due to the heterogeneity of the methods. Eligibility studies included peer-reviewed articles published from 2015 to 2025 [24], written in English, which exploit hyperspectral remote sensing data for the analysis of the selected application domains [59]. The timeframe selected reflects key development period in hyperspectral sensors and algorithmic updates, including the transition toward new kinds of approaches where ML/DL and hybrid frameworks are involved. Duplicates and not strictly related studies were excluded. In addition, studies not reporting accuracy metrics were excluded in the eligibility screening stage of the papers considered.

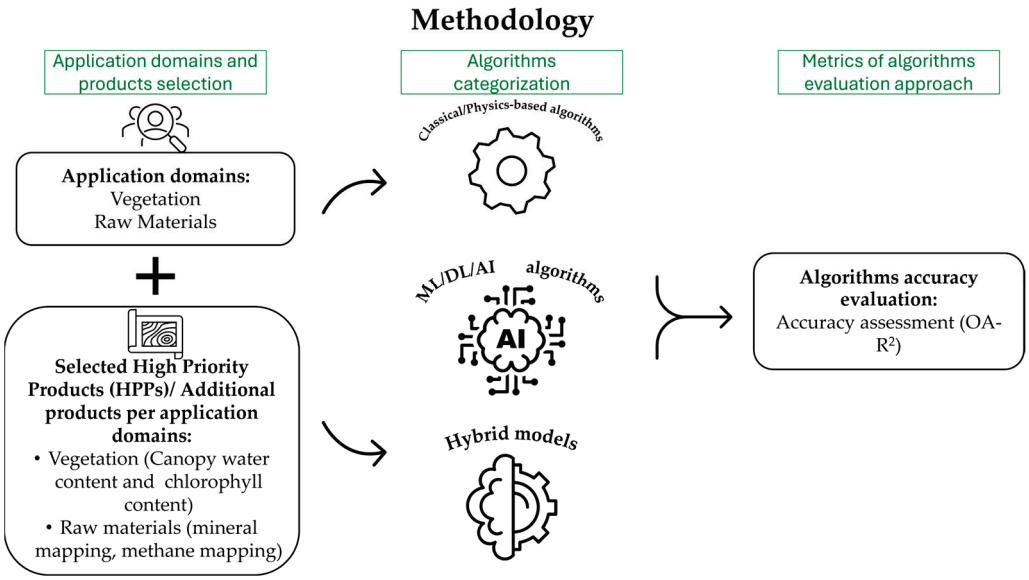


Figure 4. Methodology workflow of the review. The research followed a three-stage process: (1) two application domains were identified and the corresponding High-Priority Products were selected (**left**); (2) studies addressing the retrieval of these products were categorized according to the algorithm used (**center**); (3) the accuracy metrics reported in the studies were compared to evaluate which algorithm achieved the highest accuracy (**right**).

Table 4. Research strings for C/PB algorithms analysis.

Research String	Application Domain	HPP/Additional Product
Results for AB = (“hyperspectral data”) AND AB = (“canopy water content”) AND AB = (“algorithms”)	Vegetation	Canopy Water Content (CWC)
Results for AB = (“hyperspectral data”) AND AB = (“chlorophyll”) AND AB = (“algorithms”)	Vegetation	Chlorophyll Content (CC)
Results for AB = (“hyperspectral data”) AND AB = (“mineral mapping”) AND AB = (“algorithms”)	Raw materials (mineral)	Mineral abundance mapping
Results for AB = (“hyperspectral data”) AND ALL = (gas) AND AB = (“algorithms”)	Raw materials (gas)	Methane mapping

It is important to underline that, despite the filtering process considering only hyperspectral data, to overcome the specific limitation of these data (i.e., low revisit time), several works used HSI in synergy with other optical sensors. This information has been collected and analyzed together with hyperspectral applications.

Table 5. Research key for AI/ML/DL and hybrid algorithms.

Research String (ML/AI/DL)	Application Domain	HPP/Additional Product
AB = (“hyperspectral data”) AND AB = (“canopy water content”) AND AB = (“algorithms”) AND AB = (“machine learning” OR “artificial intelligence” OR “deep learning”)	Vegetation	Canopy Water Content (CWC)
AB = (“hyperspectral data”) AND AB = (“chlorophyll”) AND AB = (“algorithms”) AND AB = (“machine learning” OR “artificial intelligence” OR “deep learning”)	Vegetation	Chlorophyll Content (CC)

Table 5. *Cont.*

Research String (ML/AI/DL)	Application Domain	HPP/Additional Product
AB = (“hyperspectral data”) AND AB = (“mineral mapping”) AND AB = (“algorithms”) AND AB = (“machine learning” OR “artificial intelligence” OR “deep learning”)	Raw materials (mineral)	Mineral abundance mapping
AB = (“hyperspectral data”) AND ALL = (methane) AND AB = (“algorithms”) AND AB = (“machine learning” OR “artificial intelligence” OR “deep learning”) AND ALL = (gas) AND AB = (“algorithms”)	Raw materials (gas)	Methane mapping

5.4. Risk-of-Bias Assessment

To evaluate the methodological robustness across the studies included in the research, a specific risk-of-bias (RoB) tool was developed. Standard risk-of-bias instruments, such as QUADAS-2, ROBINS-I, PROBAST, Cochrane RoB, were not used because they were designed for clinical trials, diagnostic accuracy studies, or predictive models with human subjects. These are not designed for algorithm development studies in hyperspectral remote sensing, where biases arise mainly from the dataset construction, validation and metric reporting.

Starting with this assumption, a tailored PRISMA-aligned approach was conducted to assess the risk of bias. Four characteristics were considered: (C1) dataset quality and representativeness, (C2) validation design and overfitting control, (C3) accuracy and reporting, (C4) transparency and reproducibility. These characteristics were used to rate each work as low, high or unclear RoB, according to the explicit analysis criteria better explained in Table 6.

Table 6. Risk-of-bias analysis criteria categories.

Category	Low Risk	High Risk	Unclear
C1—Dataset quality	Adequate sample size; multi-scene or multi-site dataset; clear ground truth	Very small or single-scene dataset; poorly defined ground truth	Insufficient dataset description
C2—Validation design	Independent test set and/or well-defined CV	No clear train/test separation; ambiguous validation	Validation referenced but not described
C3—Accuracy reporting	Complete metrics (e.g., R^2 + RMSE; OA + F1) with uncertainty estimates	Incomplete or inappropriate metrics; no uncertainty; training-only results	Metrics reported without context or clarity
C4—Transparency and reproducibility	Clear description of methods, preprocessing, parameters; code available	Major methodological details missing	Partial information

For each study, a value was assigned to each characteristic (C1–C4): 0 (satisfied), 1 (not satisfied). If the overall value was higher than 1, the RoB for the selected paper was rated as high; if the risk was low within the four-category system, the final classification of bias was low; and if some of the criteria were not satisfied or not specified in the work analyzed, the paper was rated as unclear. The distribution of high-risk ratings (C1 and C2) showed very few high-risk cases but a high proportion of unclear ratings due to insufficient reporting. C3 (accuracy reporting) exhibited the largest concentration of high-risk ratings, with approximately 55% of vegetation and methane studies and approximately 40% of mineral studies failing to report complete or standardized metrics. C4 (transparency) showed almost no high-risk ratings but widespread unclear assessments,

especially in machine learning (ML)/deep learning (DL) papers, where preprocessing and hyperparameters were often underreported. Overall, insufficient reporting, rather than poor methodological design, was the primary contributor to RoB in hyperspectral studies.

5.5. Performance Metrics and Comparative Analysis

In order to compare the different algorithm categories, the following accuracy metrics were considered: (1) coefficient of determination (R^2), mainly for regression-based studies (e.g., methane, CWC), and (2) Overall Accuracy (OA). The R^2 quantifies how well the model's predictions reproduce the variability in the value of a reference variable. Higher R^2 values indicate that the algorithm captures a larger proportion of the true spectral–biophysical variability. The OA expresses the proportion of correctly classified samples over the total number of validation samples. It is the standard metric for mineral mapping, where outputs are categorical classes (e.g., kaolinite), and continuous regression metrics such as R^2 are not applicable. These two metrics were selected because they were used in all the papers analyzed and reported most consistently across the scientific literature, facilitating the comparison among heterogeneous algorithm categories These enable a more in-depth explanation of two dominant retrieval problem types in hyperspectral data analysis: regression and classification. Due to the heterogeneity in sensors, datasets, and validation strategies, a descriptive comparative analysis was performed instead of a meta-analysis. To capture performance variability, highlight methodological trends, and identify differences between C/PB, ML/DL, and hybrid approaches, the minimum and maximum R^2 or OA values were extracted for each product and algorithm category. When relevant, the use of hyperspectral multi-sensor synergies was also noted, particularly in methane and mineral mapping applications.

6. Results

The Results Section provides a synthesis of the reviewed studies, organized by application domain, target product, algorithm category, sensor platform and covered spectral ranges. Considering the large number and heterogeneity of the included studies, their specific key characteristics, such as data source (UAV, airborne, multi-sensor, etc.), methodology and the primary research objective, are reported in an aggregated and comparative form instead of an individual study summary. This approach enables the identification of robust methodological and sensor ware trends across the selected literature. In particular, in this systematic review, a total of 194 scientific papers were analyzed across the two considered application domains (i.e., vegetation and raw materials), after a first phase of screening to remove duplicates and studies which did not report accuracy metrics. The literature was collected from two leading scientific databases: Web of Science, (WoS) with 92 papers, and Scopus, contributing 102 papers (Figure 4). An overview of the data extraction template, listing the column names extracted during the screening phase, is provided in the Supplementary Material (Table S1).

Figure 5a shows a summary of the number of scientific papers analyzed, dividing them by application domain and product. More papers were related to chlorophyll content retrieval in the vegetation application domain (107), followed by methane mapping (55) and mineral mapping (23) for the raw materials domain. Finally, the canopy water content in the vegetation domain is the least studied, with only nine papers. Figure 5b shows the growing use of hyperspectral data over the time period considered in this review divided by application domain and product.

The tailored risk-of-bias (RoB) assessment revealed that many studies were limited by incomplete reporting. This analysis suggested that the main limitation lies in the under-reporting of methodological details.

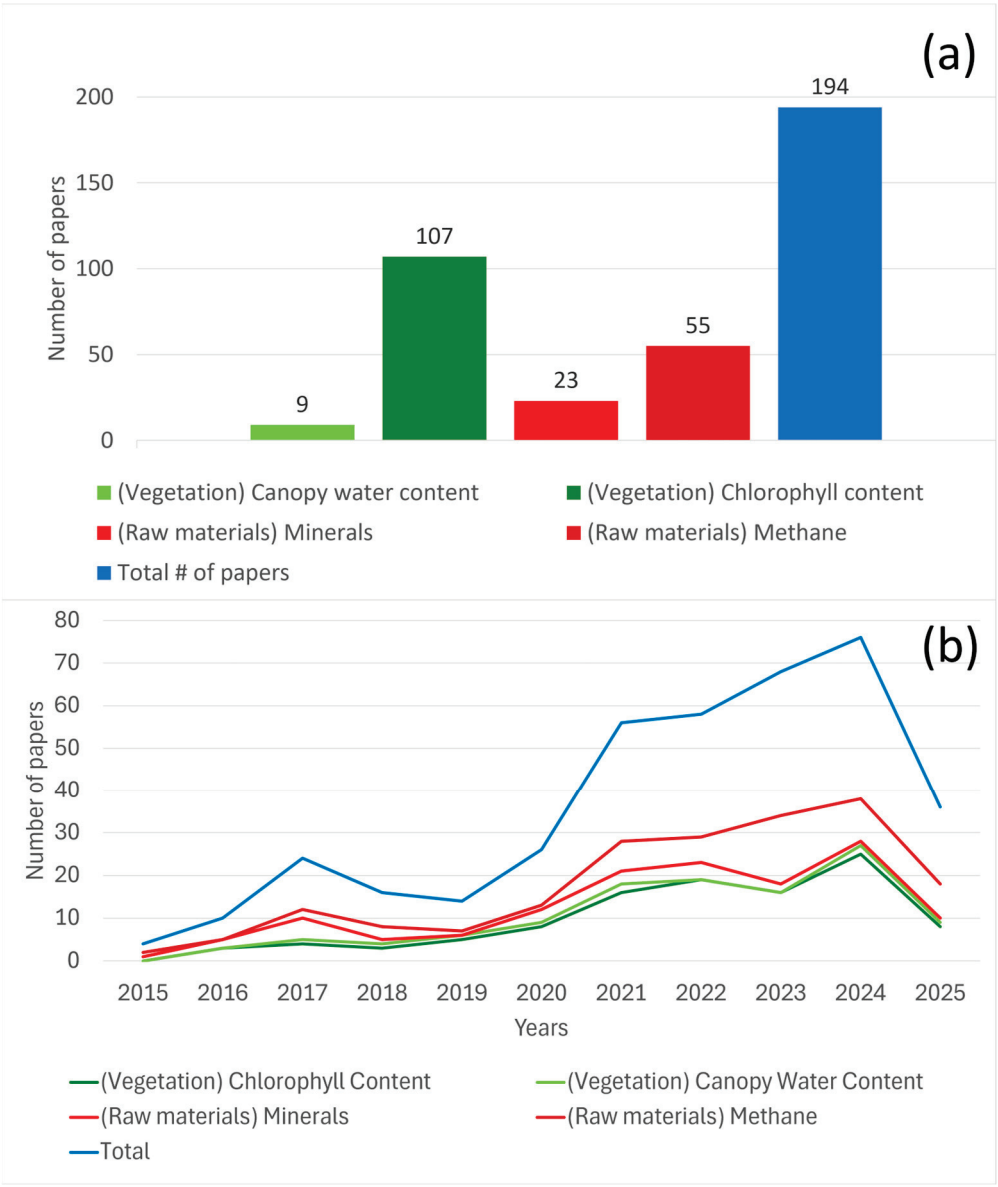


Figure 5. Number of papers analyzed in this systematic review: (**a**). Total number of papers analyzed for each application domain and product; (**b**) timeline of published papers for each application domain and product. The negative trend in 2025 is representative only of the first part of the year (until July 2025).

Considering the large number and the heterogeneity of the reviewed studies, the main characteristics, such as primary research objective, dominant algorithmic approach and data source, are summarized in Table 7.

The first result extracted from the review is related to the most relevant spectral ranges used for each application domain considered. This result allow us to more accurately detail the information compared to Table 1. This confirmed that the VNIR-SWIR range (400–2500 nm) is the core spectral domain for the retrieval of vegetation properties and mineral mapping, with almost 88%, 56% and 91% of the studies using this range for chlorophyll, CWC and mineral retrieval, respectively. The SWIR range (900–2500 nm) is particularly used for mapping methane.

For methane and mineral mapping, thermal IR (7500–13500 nm) is also used, while MWIR (3000–5000 nm) has limited importance for these application domains (Figure 6).

https://doi.org/10.3390/rs18050798

Table 7. Aggregated overview of the main characteristics of the reviewed studies across application domains and products.

Application Domain	Product	Primary Research Objective	Dominant Algorithm Category	Main Data Source
Vegetation	CWC	Quantitative biophysical retrieval	ML/Hybrid	Field, UAV
Vegetation	CC	Quantitative biophysical retrieval	ML/DL	Field, UAV, Satellite
Raw materials	Mineral mapping	Detection/abundance mapping	C/PB	Airborne, Satellite
Raw materials	Methane mapping	Detection and mapping	C/PB/Hybrid	Airborne, Satellite

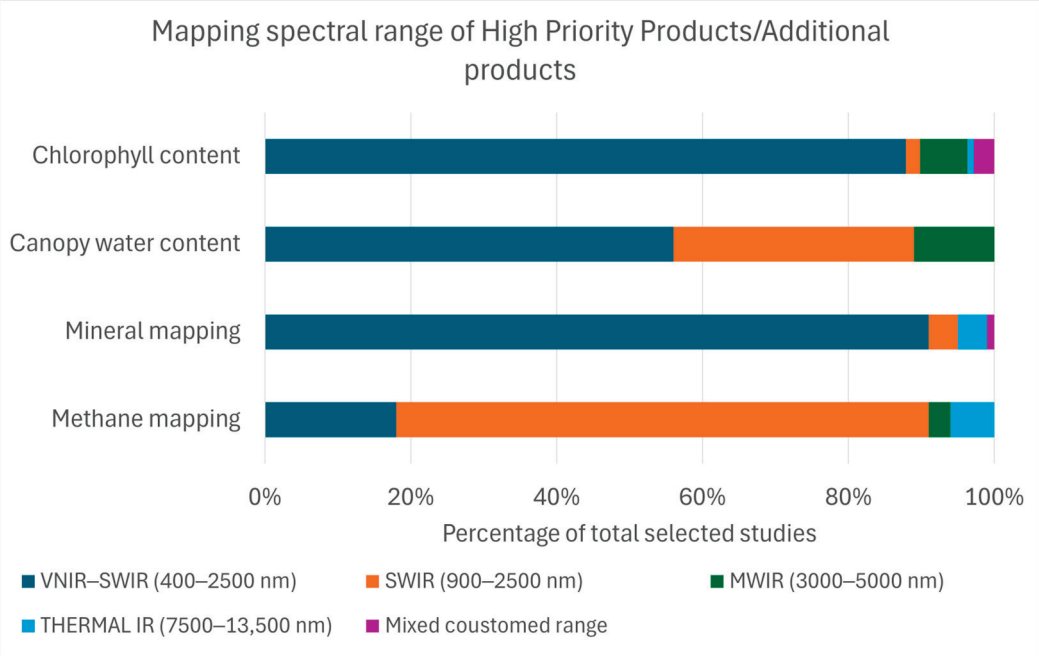


Figure 6. Percentage of papers that used specific electromagnetic spectrum ranges for the retrieval of the products.

6.1. Vegetation

The following paragraphs will repots the findings of this review for the vegetation domain, divided by canopy water content and chlorophill content. Four results will be reported for each product: (1) the most commonly used algorithm, (2) the most common sensor platform, (3) the most used spectral ranges, and (4) the accuracy metric values.

6.1.1. Canopy Water Content (CWC)

The evaluation of a parameter like the canopy water content (CWC) has undergone a notable evolution from C/PB models to modern algorithms (ML or hybrid approach). The most used algorithm for this parameter is ML, which accounts for 56% of the total methodologies applied for CWC evaluation (Figure 7). On the other hand, C/PB approaches, for example, Radiative Transfer Models (RTMs) and empirical relationships, represent 22% of the total. Hybrid approaches that combine physical modelling with ML account for 11%, as well as deep learning (DL). Specifically, the algorithms used are as follows (Figure 7):

- The C/PB partial least squares regression (PLSR) and its relative hybrid combination with the artificial neural network (PLSR-ANN hybrids) account for 23%, em-

phasizing the traditional stronghold of linear latent space techniques adapted to hyperspectral complexities.

- We found that 18% of the studies investigated show that ML algorithms, such as Boosting and tree-based ensemble methods (e.g., Random Forest, XGBoost, CatBoost, LightGBM, Extremely Randomized Trees), are among the most popular. Their ability to capture non-linearities and handle high-dimensional data makes them particularly suited to hyperspectral analysis.
- The ML Gaussian Process Regression (GPR) represents 18%, although powerful for uncertainty estimation, likely due to its computational cost on large hyperspectral datasets.
- The ML Support Vector Machines (SVM/SVR) account for 12%, typically preferred for their robust generalization in medium-sized datasets.
- CNNs, autoencoders and MDN variants, which are DL models, represent only 6%.
- C/PB approaches, bundled under "other" (RTMs, LUTs, empirical relations), retain a significant niche on this domain (23%).

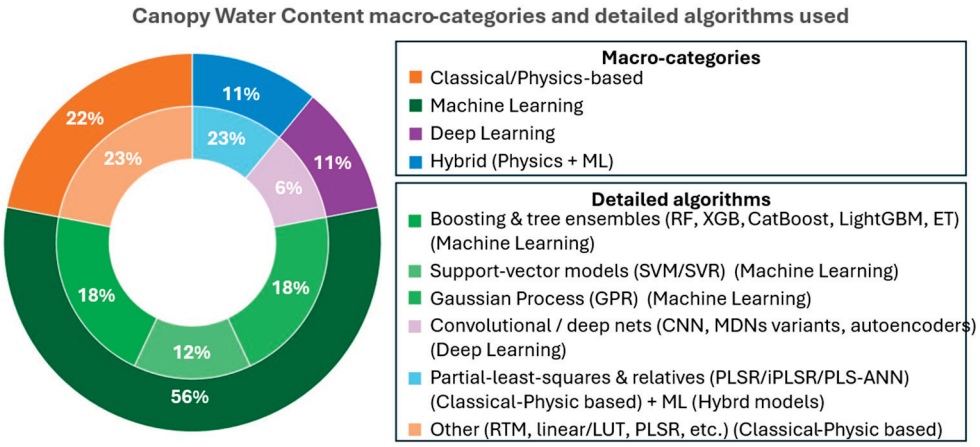


Figure 7. Most used algorithms for canopy water content retrieval: the outer circle represents the categories of algorithms (i.e., C/PB, ML, DL or hybrid models); the inner circle represents, in each category, the most used algorithms.

The following sensor platforms were used to assess CWC (Figure 8):

- Field spectroradiometers are the most commonly used instrument in studies (31%), confirming that field-calibrated data remain the gold standard for model development and validation.
- Spaceborne and airborne hyperspectral systems, such as PRISMA, AVIRIS-NG, and EnMap (15%), are now significant contributors, reflecting the maturation of satellite hyperspectral missions.
- UAV-mounted hyperspectral sensors and airborne systems each contribute 8 and 23%, respectively, indicating the rising relevance of flexible, high-resolution platforms for local and regional studies.
- The multispectral spaceborne Sentinel-2 MSI system also appears with 8%, despite its spectral limitations for precise water absorption feature detection.

These results emphasize the importance of spectral representation fidelity (hyperspectral spectroradiometer) and spatial flexibility (UAVs) in current CWC research.

Finally, the spectral ranges utilized in CWC retrieval (Figure 9) point to a preference for broad, high-information content ranges:

- Full-range VNIR-SWIR coverage (400–2500 nm) is the most exploited (~55%), confirming that key water absorption features at 970 nm, 1200 nm, 1450 nm, and 1940 nm are critical for accurate estimation.

- VNIR-only sensors (400–1000 nm) are used in about 30% of cases, primarily when SWIR coverage is unavailable due to radiometric limitations (e.g., high noise level).
- SWIR-only windows (e.g., 2100–2300 nm) and thermal infrared (TIR, 7000–12,000 nm) are minimally used, indicating that, despite being useful for stress detection, they are less informative for direct water quantification.

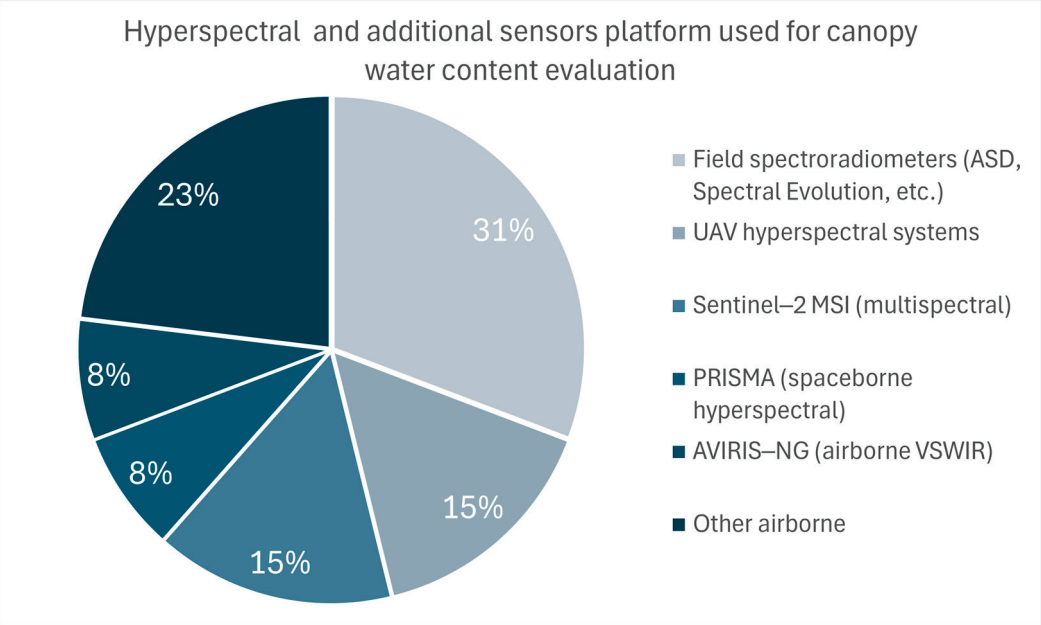


Figure 8. Hyperspectral and additional sensor platforms used for canopy water content retrieval.

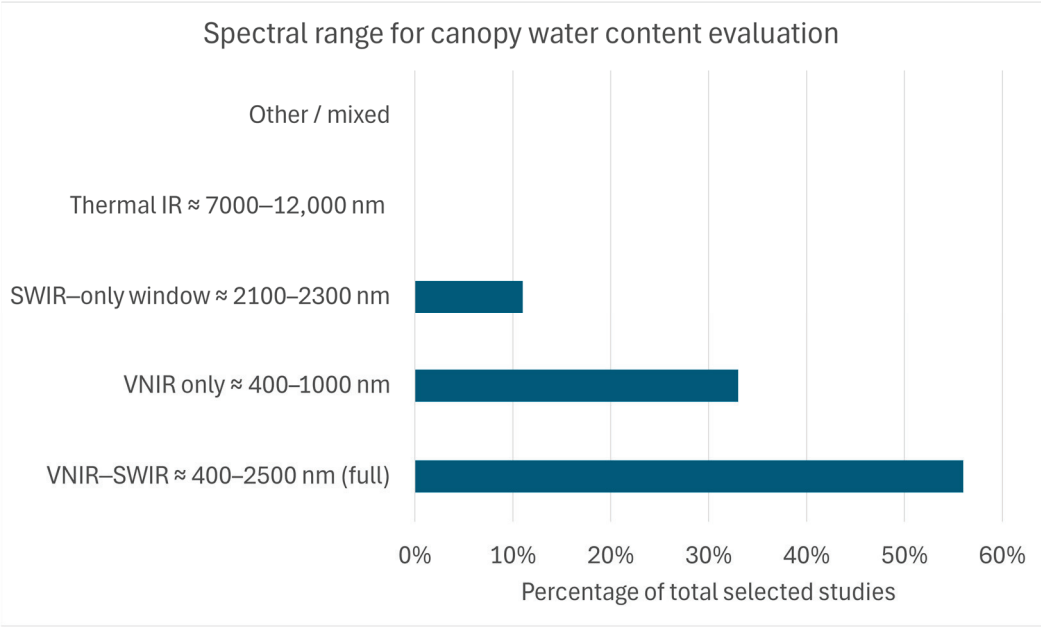


Figure 9. Spectral ranges used for the estimation of canopy water content.

This result emphasizes the importance of accessing both the VNIR and SWIR domains to fully capture the spectral signatures related to CWC.

Finally, a comparison of algorithm performance is shown in Figure 10. The accuracy metrics (R^2) indicated in the studies are the following:

- Machine learning approaches exhibit the widest R^2 range, from a minimum of 0.58 to a maximum of 0.95, suggesting high variability but also remarkable potential for excellent predictions under optimal conditions.
- For hybrid models (ML + RTM), just one paper was analyzed, presenting an R^2 value of 0.78.
- Physics-based methods show the narrowest and generally lower R^2 values (0.448–0.665), highlighting the limitations of purely physical models when confronted with complex canopy heterogeneity and non-linear water content responses.

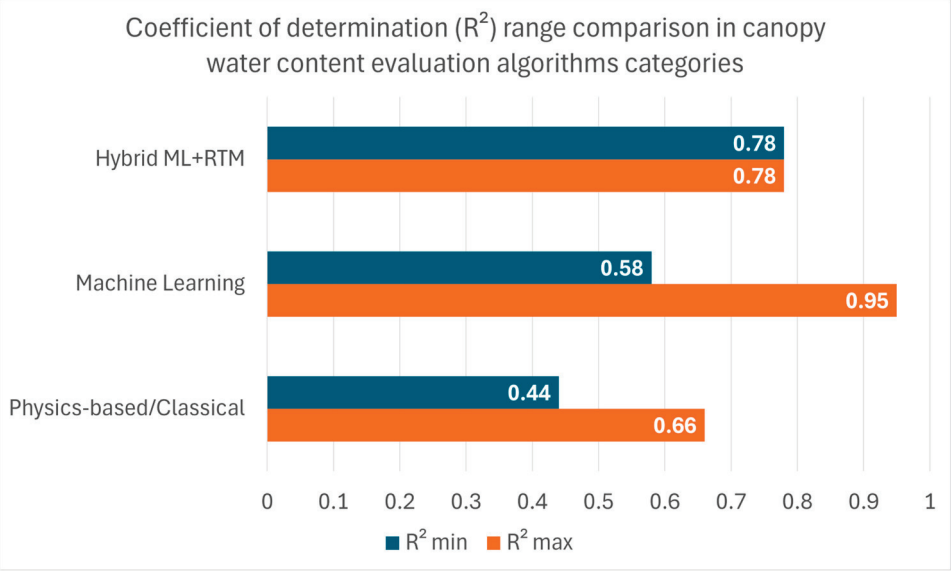


Figure 10. Coefficient of determination (R^2) range comparison in canopy water content evaluation algorithm categories. The blue bars indicate the minimum value of R^2 reported, while the orange bars indicate the maximum value.

The wide R^2 range observed for physics-based approaches reflects the heterogeneity of physical modelling strategies collected in this work (from empirical regression to RTM inversion) and experimental conditions across studies, rather than inconsistent performance of some methods themselves.

6.1.2. Chlorophyll Content (CC)

The application of ML algorithms dominates studies on CC, being employed in 59% of the reviewed studies (Figure 11). C/PB approaches, such as radiometric indices (e.g., NDVI variants, chlorophyll absorption ratios) and RTM inversions, account for 21%. Deep learning methods are employed in 12% of the analyzed studies, a slightly higher proportion than in CWC studies. Hybrid approaches (C/PB + ML) are emerging for the assessment of chlorophyll content (8%).

A breakdown of algorithmic families (Figure 11) reveals a clear preference structure:

- ML algorithms, such as Boosting and ensemble trees (Random Forest, XGBoost, CatBoost, LightGBM, Completely Random Forests), represent the most widely used group (28%), leveraging their capability to handle high-dimensional, noisy hyperspectral inputs.
- The C/PB PLSR (but also iPLSR) and its hybrid derivatives (PLS-ANN) retain a relevant role (21%).
- The ML Support Vector Machines (SVM/SVR), including semi-supervised extensions, account for 15%, typically used for cases with limited ground-truth data.

- Spectral–spatial DL architectures (CNNs, Mixture Density Networks, Deep Belief Networks) follow at 12%, indicating growing confidence in their capacity to model complex chlorophyll spectral signatures and spatial heterogeneity.
- The ML Gaussian Processes (GPRs), both homoscedastic and heteroscedastic, remain marginal (9%), again likely due to computational constraints for large datasets.
- C/PB models, including rule-based regression models such as Cubist, Multivariate Adaptive Regression Splines (MARS), logistic models, and direct Look-Up Table and Radiative Transfer Models (LUT/RTM) inversions, account for 8%, indicating that purely physics-based strategies are increasingly complemented by statistical corrections or optimizations.
- Finally, 7% of the algorithms considered are hybrid, in particular RTM combined with DL techniques, such as CNNs, or ML techniques, such as SVM.

Sensor platform analysis (Figure 12) shows, as well as for CWC, a marked preference for field-based measurements:

- Field spectroradiometers (e.g., ASD, PSR, SVC, Ocean Optics) dominate at 35%, confirming that field-level high-resolution data remain essential for calibrating and validating chlorophyll models.
- UAV-based hyperspectral platforms (Headwall, Cubert, GaiaSky, Pika) contribute a significant 25%, reflecting the requirement of high-spatial-resolution measurements at the canopy scale.
- Other platforms, including Sentinel-2 MSI, HyMap, DESIS, Hyperion, LiDAR-HS, and various CubeSat/MS solutions, represent 27%.
- Satellite and airborne sensors, such as PRISMA and AVIRIS-NG, appear at 7% and 6%, respectively.

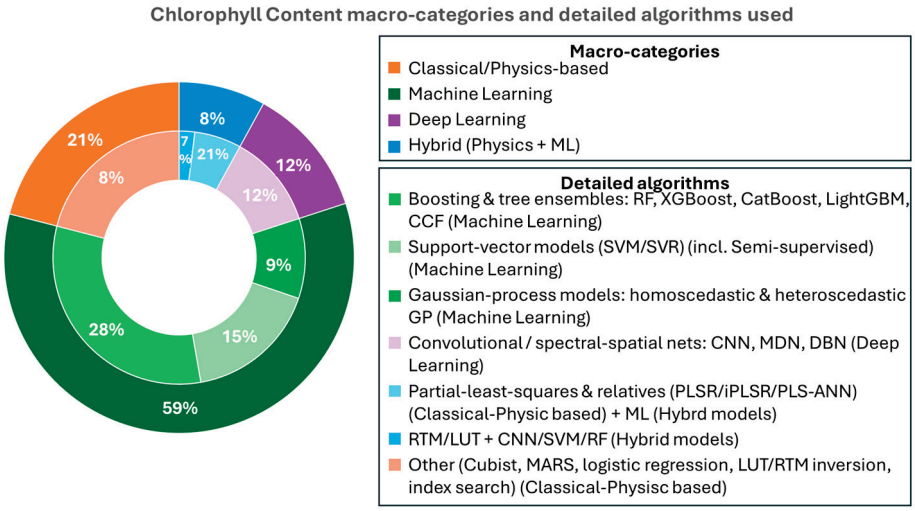


Figure 11. Most used algorithms for chlorophyll content estimation: the outer circle represents the categories of algorithms (i.e., C/PB, ML, DL or hybrid models); the inner circle represents, in each category, the most used algorithms.

It is evident from Figure 13 that the full VNIR-SWIR range (400–2500 nm) is employed in the majority (more than 85%) of studies for CC estimation. This full spectral range is crucial for capturing the multiple absorption features related to chlorophyll, as well as for the CWC.

VNIR-only (400–1000 nm) approaches are rare (~5–8%), and the utilization of SWIR-only, MWIR, or TIR ranges is low or negligible, underscoring the necessity of broad spectral coverage to accurately isolate chlorophyll signals from background variability.

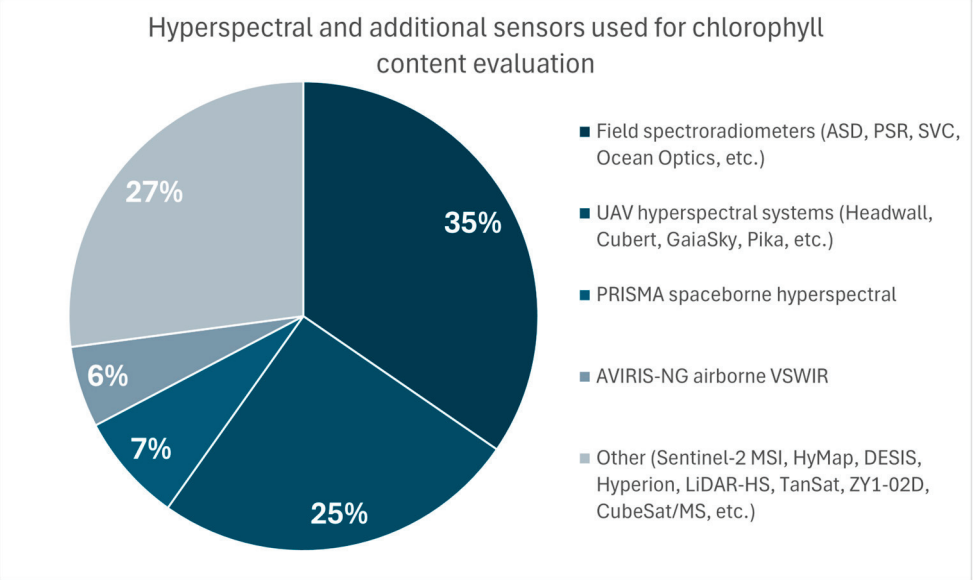


Figure 12. Hyperspectral and additional sensor platforms used for chlorophyll content estimation.

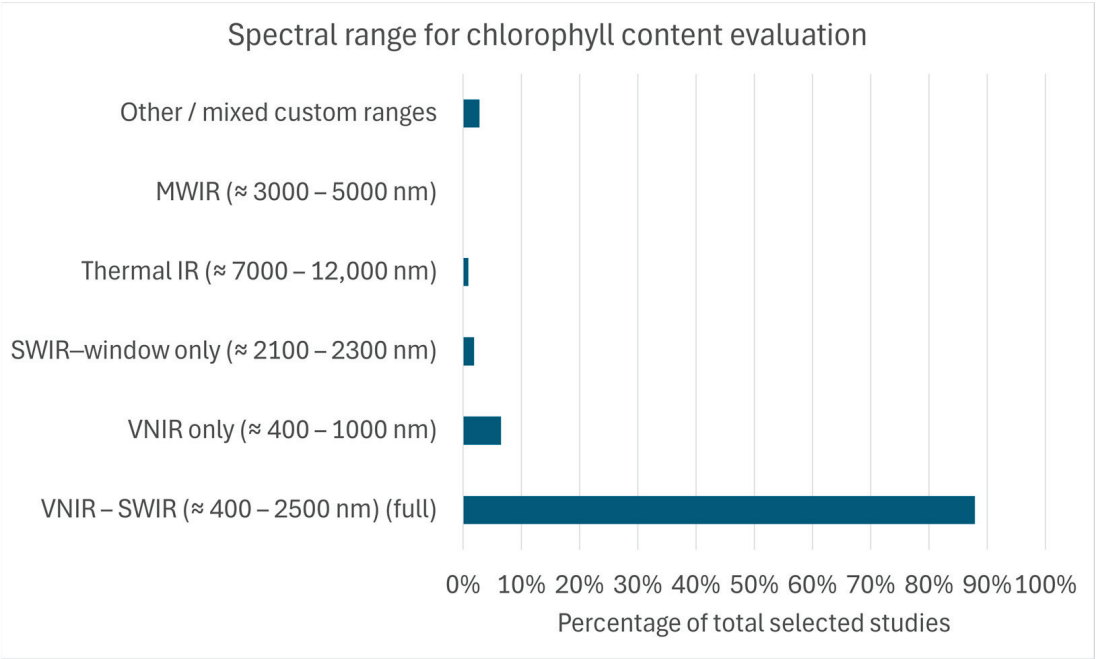


Figure 13. Spectral ranges used for the estimation of chlorophyll content.

The performance analysis based on the coefficient of determination (R^2) (Figure 14) reveals important trends:

- DL models follow closely, with R^2 ranging from ~0.8 to ~0.98, reflecting their ability to model complex non-linear spectral signatures, albeit with higher sensitivity to training data quality.
- Pure ML models show greater variability, with R^2 from ~0.44 to ~0.99, emphasizing that, without physical constraints, the models can perform extremely well but also risk underfitting or overfitting in heterogeneous conditions, according to the analyzed papers.

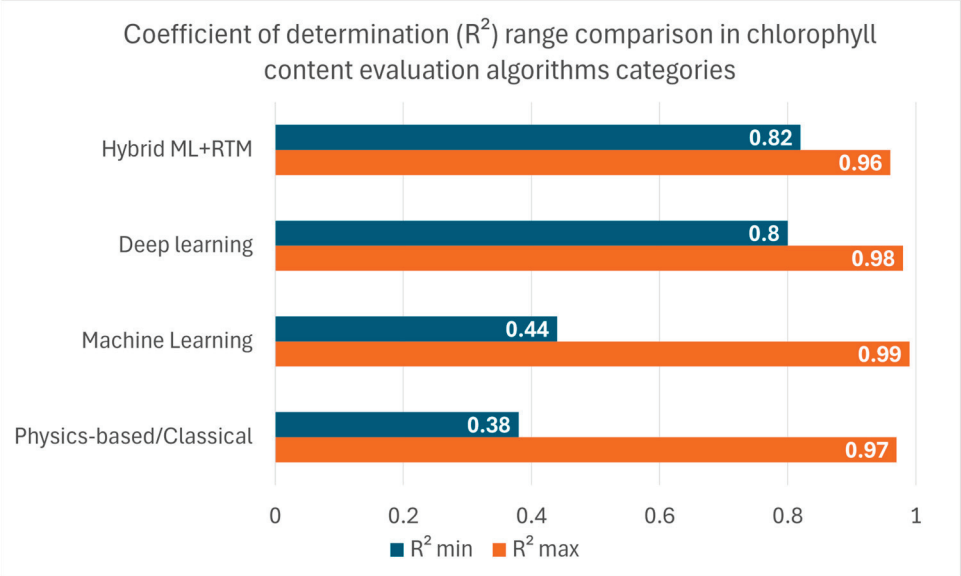


Figure 14. Coefficient of determination (R^2) range comparison in chlorophyll content evaluation algorithm categories. The blue bars indicate the minimum value of R^2 reported, while the orange bars indicate the maximum value.

6.2. Raw Materials

The following paragraphs report the findings of the reviews for the raw materials domain, divided by mineral mapping and methane mapping. Four results will be reported for each product: (1) the most commonly used algorithm, (2) the most common sensor platform, (3) the most used spectral ranges, and (4) the accuracy metric values.

6.2.1. Mineral Mapping

A detailed analysis of the studies reviewed provides key insights into algorithm usage, sensor deployment, spectral preferences, and performance metrics.

In contrast to vegetation applications, C/PB methods dominate mineral mapping studies, comprising 52% of the algorithms analyzed (Figure 15). These include techniques such as Spectral Angle Mapper (SAM), Matched Filtering (MF), Linear Spectral Unmixing (LSU), and direct reflectance-based detection.

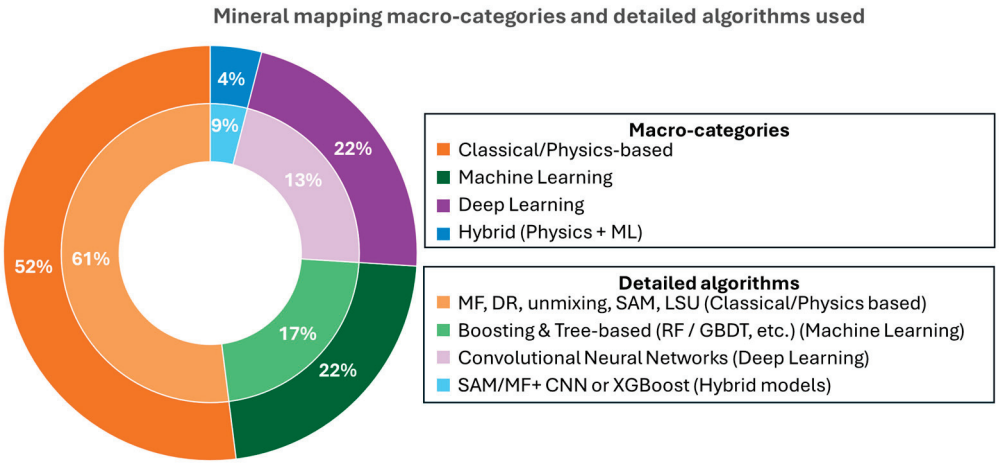


Figure 15. Most used algorithms for mineral mapping: the outer circle represents the categories of algorithms (i.e., C/PB, ML, DL or hybrid models); the inner circle represents, in each category, the most used algorithms.

Machine learning methods account for 22%, the same as deep learning methods (22%), while hybrid models (C/PB + ML) are still marginal (4%).

Analyzing the specific algorithm families (Figure 15), we find a clear traditional dominance:

- Classical/physics-based methods (SAM, LSU, MF) still dominate with 61% usage. These approaches exploit the well-defined spectral absorption characteristics of minerals, leveraging physics-based matching rather than data-driven pattern recognition.
- Boosting and tree-based ML algorithms (e.g., Random Forest, Gradient-Boosted Decision Trees) represent 17%, valued for their ability to discriminate subtle spectral mixtures and handle noisy field or airborne data.
- DL methods, such as Convolutional Neural Networks (CNNs), have started to gain traction (13%), particularly in high-complexity mapping scenarios, considering the analyzed references.
- Hybrid models are marginal (9%), mainly related to SAM/MF techniques added to an ML framework, for computation empowerment.

The analysis on the sensor and platform usage (Figure 16) highlights the dominance of airborne platforms:

- Airborne or satellite hyperspectral sensors (HyMap, DESIS, Hyperion, M3, ASTER, etc.) account for the majority (61%), reflecting the historical and ongoing importance of airborne hyperspectral mapping missions in mineral exploration.
- AVIRIS-NG, a high-performance airborne imaging spectrometer, is utilized in 13% of the studies, highlighting its high spectral resolution in the VNIR-SWIR.
- PRISMA (spaceborne hyperspectral) accounts for 17%, indicating the importance of hyperspectral satellite missions in large-scale mineral mapping.
- Thermal and longwave systems (e.g., HyTES, HySpex LWIR) are also represented (9%), important for detecting minerals with strong TIR emissivity features (e.g., carbonates, sulfates).

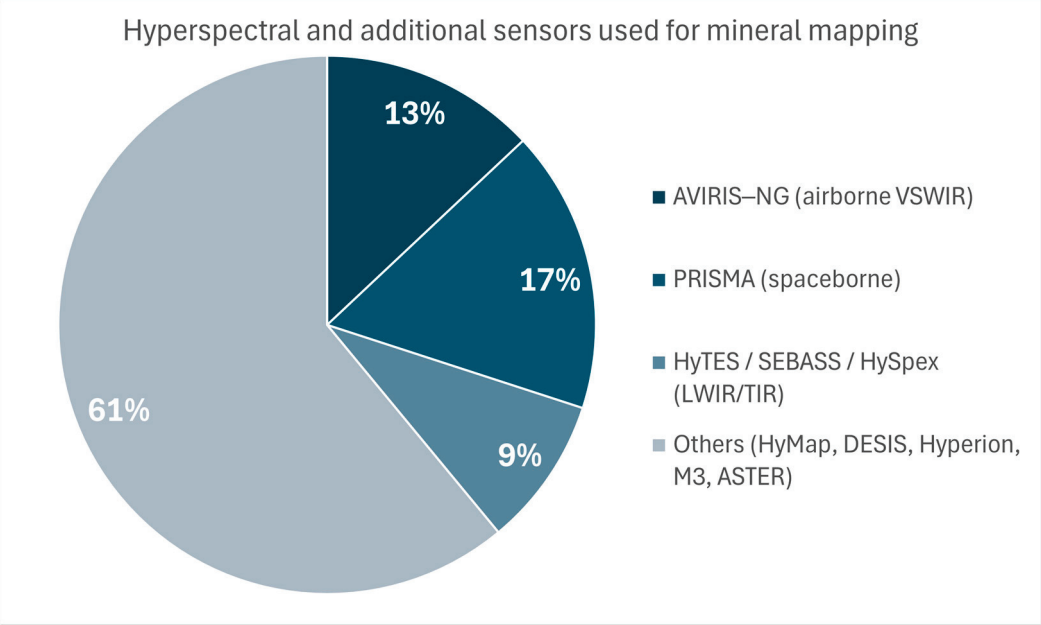


Figure 16. Hyperspectral and additional sensor platforms used for mineral mapping.

Most mineral mapping studies employ the full VNIR-SWIR range (400–2500 nm) (Figure 17), critical for capturing key mineral absorption features. Each mineral shows a different response in the electromagnetic spectrum:

- Al-rich phyllosilicates (Al-OH), such as Kaolinite, Illite, Muscovite, Pyrophyllite; iron-bearing phyllosilicates (Fe-OH), such as Nontronite, Ferruginous smectite, Goethite; magnesium-rich phyllosilicates (Mg-OH), for example, Talc, Chlorite, Saponite, Serpentine, which all have spectral features in the 2000–2500 nm SWIR region.
- Iron oxide features are pronounced in the 400 nm–1000 nm VNIR range.
- Some studies incorporate thermal infrared (TIR, 7000–12,000 nm) or MWIR (3000–5000 nm) data, but these remain minority cases (<5%), primarily used for advanced mineral classes (e.g., carbonates, sulfates).

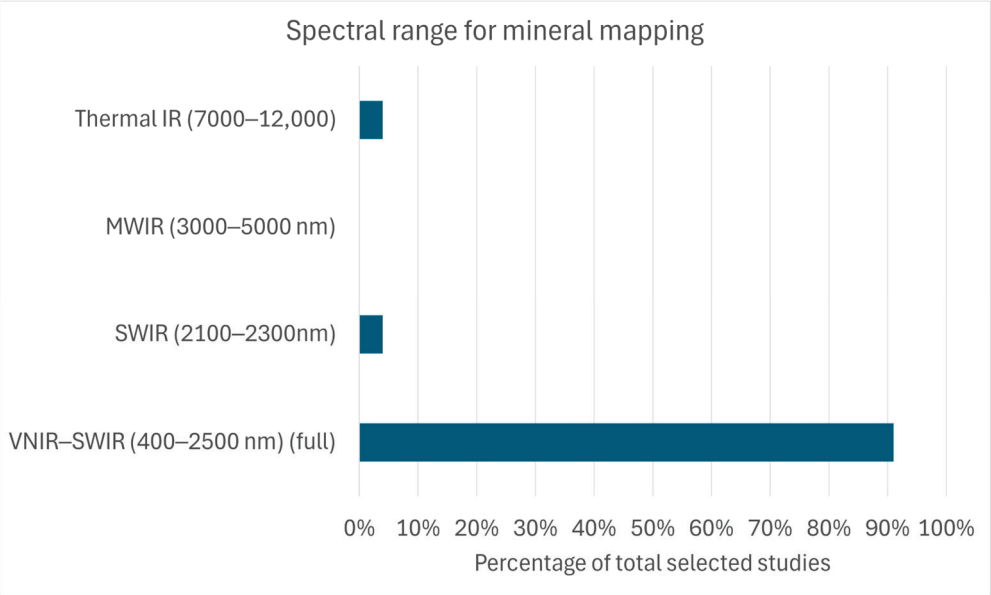


Figure 17. Spectral range used for mineral mapping.

The almost universal reliance on VNIR-SWIR confirms the fundamental importance of broad spectral coverage for accurate and discriminative mineral detection.

For this application domain, Overall Accuracy (OA) was selected as the primary comparative metric, because it was the most reported parameter across the reviewed references. Moreover, the R^2 parameter is not always applicable or reported, especially in classification tasks, such as mineral mapping. Overall Accuracy offers a standardized and interpretable measure of performance across classical, machine learning, and hybrid approaches. The comparison of Overall Accuracy (OA) across algorithm categories (Figure 18) reveals distinct performance characteristics:

- Deep learning models achieve high OA values, often ranging from 82.7% to 94.8%, reflecting their superior ability to model complex non-linear and spectral–spatial patterns, particularly in challenging or mixed-mineral scenes.
- Machine learning approaches show broader variability, with OA values from ~50% to ~90%, depending heavily on feature selection, training sample quality, and model tuning.
- C/PB methods, while highly interpretable and consistent, typically achieve slightly lower OA (ranging from ~54% to 80%), constrained by their inability to model subtle spectral variations and mixed pixel effects without manual intervention.
- Hybrid models achieve the highest OA values (80% to 96%).

While classical methods remain operationally dominant, the highest accuracies are increasingly obtained through hybrid models, especially in complex or high-resolution datasets.

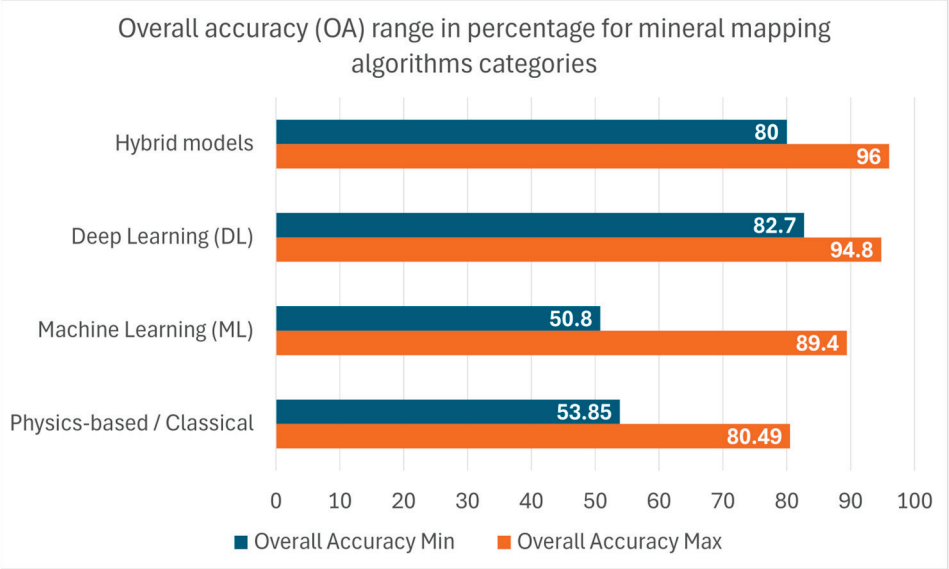


Figure 18. Overall Accuracy (OA) range in percentage for mineral mapping algorithm categories. The blue bars indicate the minimum value of R2 reported, while the orange bars indicate the maximum value.

6.2.2. Methane Mapping

The detection and quantification of methane (CH_4) by means of hyperspectral data are technically challenging, due to the weak and narrow spectral characteristics of methane compared with signals from vegetation and minerals. This literature review revealed critical patterns in the evolution of algorithmic approaches, sensor use, spectral range utilization and model performance.

Methane mapping remains heavily grounded in C/PB methods, which account for 64% of the reviewed studies (Figure 19). These methods include variants of Matched Filter (MF), Optimal Estimation, and Differential Optical Absorption Spectroscopy (DOAS). Machine learning methods are applied in 11% of the studies, while deep learning is slightly more prominent at 16%. Hybrid approaches (C/PB + ML) are still emerging, representing 9% of the cases.

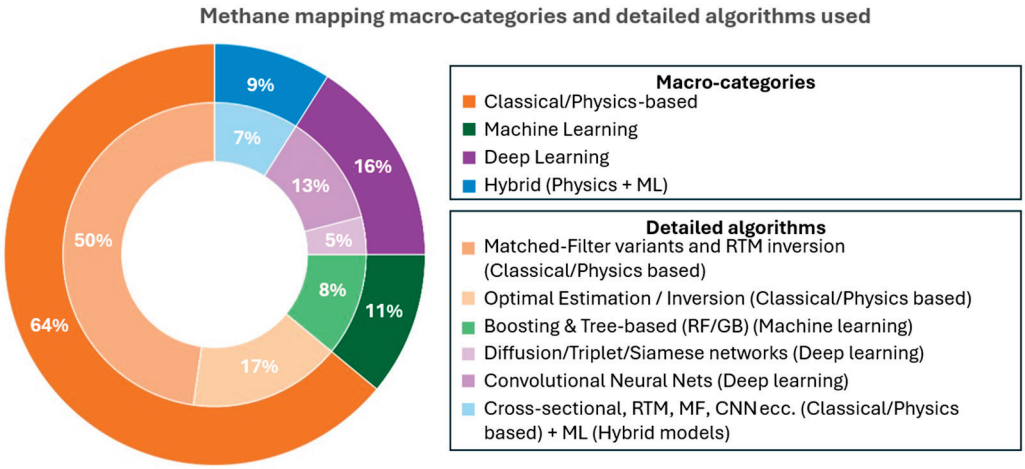


Figure 19. Methane mapping's most used algorithms: the outer circle represents the categories of algorithms (i.e., C/PB, ML, DL or hybrid models); the inner circle represents, in each category, the most used algorithms.

A breakdown of specific algorithm types (Figure 19) shows the following:

- As C/PB-based methods, Matched Filter variants (e.g., standard MF, Adaptive Matched Filter, AMF, Orthogonal Subspace Projection) dominate with 50%, exploiting known methane absorption features centered around ~2300 nm in the SWIR domain.
- Optimal estimation and inversion methods (17%) also play a central role, particularly in PB retrievals from spaceborne sounders, which is a specialized remote sensing instrument that measures the atmospheric temperature and moisture profiles by analyzing infrared (IR) radiation [60].
- DOAS and cross-sectional techniques (7%) are mostly used with hyperspectral sounders for column concentration retrievals but are included in Matched Filtering, associated with machine learning techniques such as the MLP (Multi-Layer Perceptron) as an innovative hybrid approach.
- Among machine learning-based approaches, Boosting/tree ensembles (8%) are the main techniques used for mineral mapping.
- Deep learning architectures, including diffusion models, triplet networks, Siamese nets (5%), and CNNs (13%), emerge as powerful tools to model the complex spatial structures of methane plumes and improve false-positive rejection.

Sensor platform analysis (Figure 20) highlights the diversity of methane sensing:

- AVIRIS-NG leads (31%), confirming its critical role in airborne, high-resolution methane mapping campaigns.
- PRISMA contributes to 15% of the studies, reflecting its promising capabilities for methane detection using spaceborne sensors.
- Other types of sensors, such as Sentinel-5P TROPOMI (13%), provide low-resolution methane presence products.
- EMIT (9%, ISS onboard) provides coarse resolution but large-scale methane monitoring.
- Thermal and LWIR sensors (HyTES, SEBASS, HySpex-TIR) account for 11%, critical for surface temperature corrections and, in some cases, direct methane thermal emission retrieval.
- Hyperspectral sounders (CrIS, IASI, AIRS) (7%) and other custom platforms (UAV-based, sUAS, custom TDLAS systems) together represent the remaining 14%.

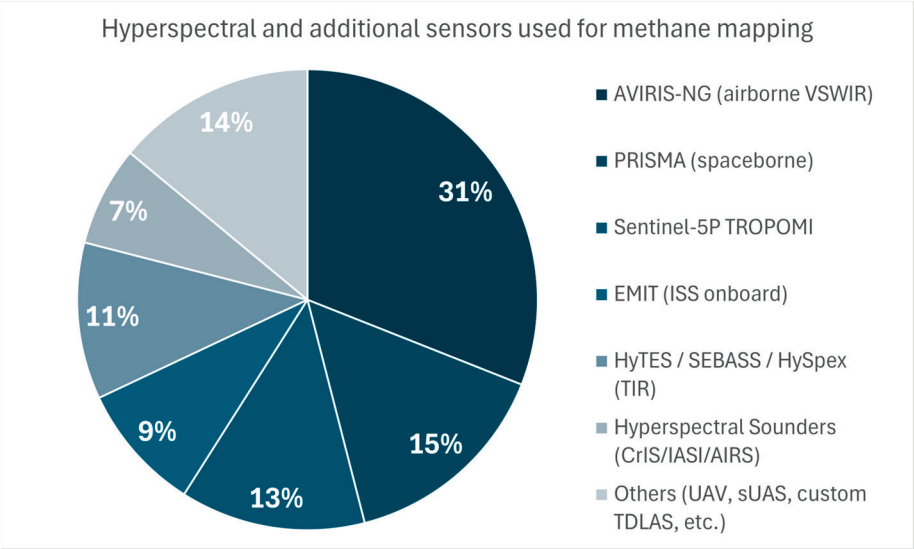


Figure 20. Hyperspectral and additional sensor platforms used for methane mapping.

Methane mapping relies overwhelmingly on the SWIR window (2000–2500 nm), as shown in Figure 21:

- Around 73% of the studies exclusively utilize the SWIR range, exploiting strong methane absorption features at ~2290–2310 nm.
- VNIR-SWIR full coverage is employed in ~20% of cases, mostly when multi-target mapping is considered (e.g., simultaneous vegetation or mineral studies).
- Use of MWIR (3000–5000 nm) (<5%) or TIR (7500–13,500 nm) (6%) remains marginal due to atmospheric absorption and weaker surface reflectance in these ranges for methane but of course related to the available hyperspectral sensor spectral range capability.

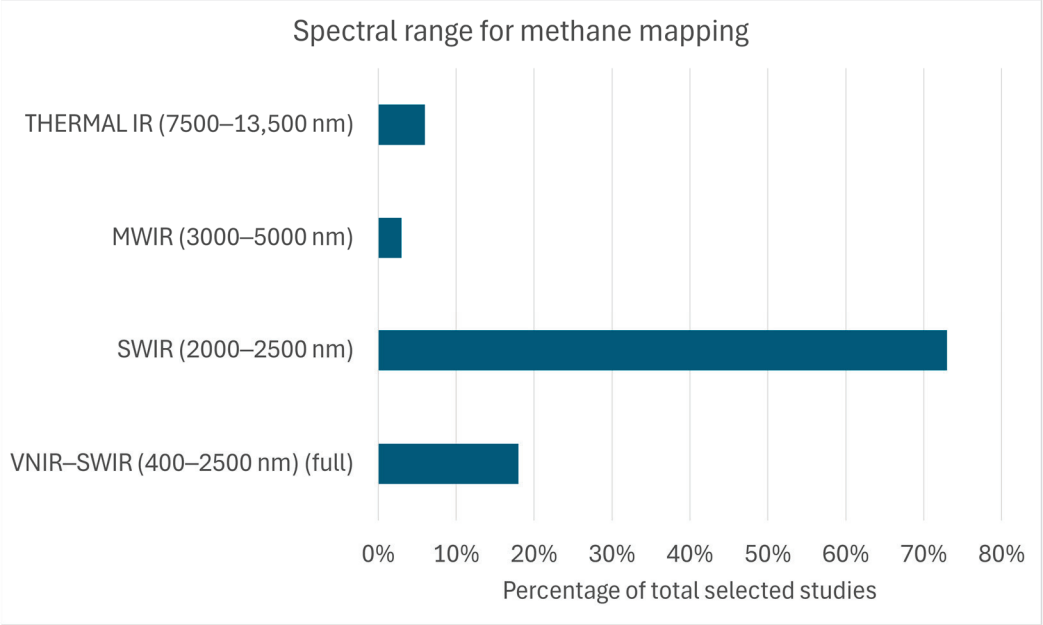


Figure 21. Spectral range used for methane mapping.

Analyzing the R^2 values reported across different categories (Figure 22), we see the following:

- Hybrid ML + RTM approaches show the highest maximum R^2 value ranges, between a minimum of 0.96 and a maximum of 0.99, confirming that integrating physics-based modeling with ML correction layers provides the most reliable methane retrievals.
- ML models, including tree-based and Boosting methods, show a constant range, from ~0.9 to ~0.92, depending heavily on feature engineering and background characterization.
- Deep learning models (CNNs, Siamese, Diffusion models), with R^2 from ~0.74 to ~0.85, show potential for plume detection but still struggle slightly with precise quantification compared to hybrid systems.
- C/PB methods explain that there is still a good option in methane mapping (~0.84 to ~0.99) but with higher variability, due to sensitivity to background spectral variability, atmospheric correction errors, and methane concentration levels.

6.3. Synthesis of Main Findings and Remaining Challenges

The results across the four products selected and across the two application domains reveal a consistent pattern about how algorithm classes are used and where their limitations arise. For vegetation products (CWC and CC), ML and DL methods dominate more than others and reach the highest R^2 values reported, but their performance strongly depends on data quality, field-based calibration and the availability of the full spectrum coverage (VNIR-SWIR). Hybrid models are used less but already show a notable balance between

accuracy and robustness, especially when radiative transfer models (RTMs) are used to regularize ML estimators.

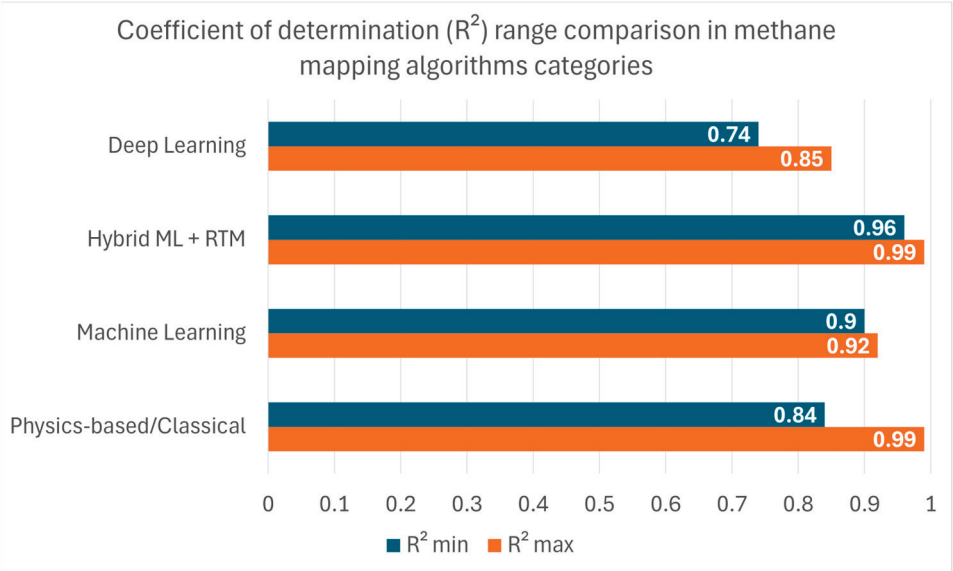


Figure 22. Coefficient of determination (R^2) range comparison in methane mapping algorithm categories. The blue bars indicate the minimum value of R2 reported, while the orange bars indicate the maximum value.

In the raw materials domain, C/PB methods remain strongly used and very effective, especially when spectral features are strong and pixels are relatively pure. ML and DL become more advantageous in heterogenous and mixed scenes. For methane, the current results confirm that C/PB approaches are still essential to exploit narrow absorption features under low-SNR conditions, while emergent methodology within DL and hybrid algorithms sections has been shown to reduce false positives and improve plume detection.

These insights highlight a set of cross-cutting challenges, data scarcity and representativeness, sensor limitations (spectral/spatial/temporal), and limited uncertainty reporting, which are repeatedly encountered across domains. In the following Discussion, we explicitly link these challenges to the algorithm groups reviewed and outline how classical, ML/DL and hybrid approaches can contribute to resolving them, also in view of upcoming missions such as CHIME, SBG and IRIDE.

7. Discussion

The analysis of 194 scientific papers on four key products (canopy water content, chlorophyll content, mineral presence maps, and methane detection) in two application domains (vegetation and raw materials) provides quantitative insight into the current state of the art of hyperspectral algorithms. The time range considered for the analysis ranges from 2015 and 2025, which coincides with the availability and consolidation of spaceborne and airborne hyperspectral sensors. This review supports the understanding of the most promising direction for the processing of hyperspectral data. Hyperspectral data analysis is constantly evolving, and machine learning (ML) and deep learning (DL) methods are increasingly central to operational applications. This review provides an understanding of the current algorithmic and sensor trends in hyperspectral data processing in Earth Observation, which are most relevant to the optics of new hyperspectral satellite missions such as CHIME (ESA), considering selected high-priority products (HPPs). As demonstrated by Berger et al. [17], the products here considered need further attention for increasing their TRLs, and this can be reached by exploring the synergy of future satellite

missions, such as CHIME, Sentinel 2 Next Generation, and Copernicus Expansion missions. Rather than a simplistic description of the performance per single methodology, the results of this work highlight deeper patterns in how the domain of application, the typology of the sensor and the quality of the training and validation data all contribute to algorithm choice and, consequently, to the retrieval accuracy. Limitations of the review process include restriction to two scientific paper databases and the inclusion of only English, open access studies. Furthermore, studies that did not report quantitative accuracy metrics (e.g., R^2, RMSE or OA) were excluded during the eligibility screening stage.

Additionally, the synthesis of trends and results also reveals how specific classes of algorithms can be used to address important products within the selected application domain. In the following, we discuss not only where C/PB, ML/DL and hybrid approaches perform well or poorly but also how they can be combined or adapted to mitigate limitations related to spectral complexity, mixed pixels, sensor constraints and data availability.

7.1. Vegetation

The maturity of algorithms in the estimation of the CWC has progressed in a significant way, with ML emerging as a relevant methodology, used in 56% of the papers analyzed on this item. This reflects a consolidated use of data-driven approaches, supported by their ability to represent this kind of variable in the high-dimensional hyperspectral data, by considering the entire VNIR-SWIR range, which was used in more than 55% of study cases about CWC assessment [61].

Considering the ML algorithms, Boosting algorithms and tree-based ensembles (e.g., Random Forest, XGBoost) are most frequently applied, as well as GPR and SVM. These kinds of methodologies demonstrate strong capability to evaluate the CWC variable. Also, the performance varies significantly, indicating high potential, but it is also important considering the input data quality. Hybrid models (e.g., ML + RTM) are represented in 11% of cases, but they consistently achieve stable performance across studies. Their ability to integrate physical constraints from Radiative Transfer Models (RTMs) with the flexibility of ML makes them particularly promising for operational-scale applications, especially in satellite-based mapping of heterogeneous, mixed-vegetation canopies [62]. Several studies have demonstrated how hybrid strategies are able to blend the physics described by methodologies such as RTMs and use the efficiency of ML to infer vegetation variables such as the CWC, which are influenced mainly by spectral feature selection balanced with the signal-to-noise ratio for each hyperspectral band [14].

This is increasingly relevant, as upcoming missions, such as IRIDE and CHIME, aim to deliver global-scale, operational vegetation products. Conversely, physics-based models alone (e.g., RTMs, LUTs) show lower and more constrained accuracy, revealing their limitations in representing canopy heterogeneity and non-linear dynamics. C/PB approaches are mainly related to the regression relationship between field observation data and vegetation indices. This method is easy to use but is significantly dependent on the availability of observational data. Other relevant physical models are based on the response of CWC to canopy reflectance, by simplifying the simulation of solar radiation and the transfer processes. Also, in this case, the application of these models is limited by model parametrization [63]. Another relevant point to consider is the scale of the data and the type of data platform. High-performance models are generally developed using small-scale, homogeneous datasets, primarily based on field data collected using a spectroradiometer and a UAV with mounted hyperspectral sensors, where the signal quality and canopy condition are also controlled [62]. However, the extension of these models to spaceborne sensors (e.g., PRISMA, EnMAP, Sentinel-2) introduces relevant

https://doi.org/10.3390/rs18050798

challenges, such as atmospheric effects, limitation of data availability for the area of interest (low revisit time) and low spatial resolution. These can impact model transferability.

These results suggest how combining RTM-based simulations or priors with ML models trained on multi-platform data (field, UAV, satellite) is relevant in the CWC analysis. RTMs can constrain the solution space and ensure physical consistency, while ML captures non-linear residuals and sensor-specific effects. This hybridization approach directly addresses the challenges of heterogeneity and sensor differences that are faced by purely physics-based or purely data-driven methods [14].

For the CC evaluation, ML is even more pronounced, used in 59% of the studies reviewed. It is underlined that chlorophyll has strong optical activity in the red-edge region [64]. In terms of algorithm performance, ML and DL models achieve the highest accuracy, improving the prediction accuracy of leaf chlorophyll content at different growth stages [65]. Tree-based ensemble models are the most applied from the ML family, followed by SVM and GPR. DL architectures (e.g., CNN) are represented in 12% of the cases, while for CWC, it is only 6%. This is due to the higher availability of reference datasets and well-characterized chlorophyll spectral responses [66]. Hybrid approaches (ML or DL combined with RTMs) are less present in the cases analyzed but demonstrate significant potential. These C/PB-informed ML models offer robustness in generalization, which is relevant in transferring models across different land cover categories and by using different sensor platforms [35,67]. They outperform purely physics-based models, which, while interpretable, often fail to capture subtle spectral variability, based on the papers selected, especially under heterogeneous conditions. Sensor platform analysis for CC reveals a predominance of field spectroradiometers and UAV-mounted hyperspectral sensors. This reflects the necessity of high-spatial-resolution data to resolve canopy-scale chlorophyll variability. The adoption of satellite sensors (e.g., PRISMA) is more frequent but still limited by requirements such as spectral resolution, limited revisit time for vegetation assessment in different seasons, and spatial resolution (30 m) and calibration uncertainties [68]. Spectral range utilization aligns closely with that of CWC, with >85% of studies employing the full VNIR-SWIR range. However, for chlorophyll specifically, the VNIR domain (400–1000 nm) is especially critical due to the concentration of diagnostic absorption features [66].

The reviewed studies indicate that physics-informed ML is effective in the stabilization of CC retrievals across sites and sensors. By using RTM simulations, red-edge band selection and chlorophyll-sensitive indices as priors or features [69], ML and DL models can reduce overfitting, improve interpretability and better handle spectral variability, thereby addressing one of the main challenges identified in the results: the limited transferability of purely empirical models [47].

7.2. Raw Materials

In the raw materials domain, mineral mapping from hyperspectral data presents a distinct methodological direction, reflecting both the complexity of mineralogical signatures and the maturity of traditional remote sensing techniques in this domain. The field still relies primarily on C/PB algorithms, particularly in scenarios with limited spatial extent and high mineral purity, such as airborne surveys targeting hydrothermal alteration zones or desert outcrops [70]. This heavy reliance on C/PB methods is justified by the strong diagnostic absorption features that minerals display in the VNIR-SWIR range, which have enabled high-accuracy mapping, without the need for complex learning models [71–73]. However, as the field shifts toward larger and more heterogeneous areas, with the presence of a high number of mixed pixels, especially with the rise of new spaceborne hyperspectral missions like PRISMA, the operational paradigm is evolving. In such contexts, C/PB methods often experience reduced performance due to spectral mixing and atmospheric

https://doi.org/10.3390/rs18050798

variability. Consequently, ML and DL are becoming increasingly relevant, offering flexibility and automation to address these new challenges. Their adoption remains constrained by the need for high-quality, labeled ground-truth data, which are often limited in mineral exploration contexts [70]. Hybrid models, combining spectral matching with ML refinement, could offer the best trade-off between accuracy and robustness, though they are still emerging and underrepresented in current applications. In order to address the limitations of classical methods in heterogeneous mineral scenes, as evidenced by the results, a two-stage hybrid workflow is proposed. The workflow comprises the following stages: (1) physics-based spectral matching (e.g., SAM, MF, unmixing) for candidate detection and feature extraction, followed by (2) ML or DL classifiers operating on a reduced set of pixels and features. This strategy is predicated on the premise that it preserves the interpretability of C/PB methods while leveraging data-driven models to improve accuracy in mixed pixels and noisy conditions [74]. This is of particular importance for large-scale mapping with CHIME-like sensors such as EnMAP [46,47,75].

For the studies analyzed related to methane mapping, it remains predominantly rooted to C/PB algorithms, mostly related to the narrow and low SNR (signal–noise ratio) in the corresponding absorption features of this gas, which requires refined knowledge of the spectral behavior of the methane and the background normalization. Many of the study cases use airborne data (e.g., AVIRIS) and over high-flux leakage sources (e.g., on gas infrastructures or landfills), with plumes very extended in space and in clear contrast with the background [76]. DL models, particularly the ones that incorporate spatial learning (such as CNNs or Siamese networks), are emerging in the scientific literature, allowing for a relevant reduction in false positives and improving plume delineation. The ability of this kind of algorithm in methane quantification remains strictly connected and limited by the lack of validated datasets. The highest performances (R^2 near 0.99) in a few cases are related to the use of hybrid models, combining spectral matched filtering, with neural networks trained on real and synthetic data [57,77,78]. This methodology is represented only in 9% of the studies considered; the studies are limited in number, but they offer a scalable and transferable path in the optics of an operational CH_4 monitoring service in orbit. Overall, C/PB methods are relevant and still necessary in methane mapping, for precise SWIR real absorption assessment, but novel learning-based models progressively model hybridization, enhancing detection sensitivity and robustness, especially in noise handling in heterogeneous scenes.

In practical terms, this review suggests that methane mapping challenges can be addressed by combining the following: (1) physics-based SWIR matched filtering for robust detection of CH_4 absorption; (2) DL models that learn spatial plume coherence and background patterns; and (3) synthetic data generation to compensate for the scarcity of well-validated training datasets. This triad directly tackles the issues of low SNR, false positives and limited labelled data highlighted in the results [53,77,79].

7.3. Overall Discussion of the Main Findings and Challenges for Hyperspectral Applications

Beyond the discussion of specific application domains and products, the findings of this review highlighted common challenges for future hyperspectral applications.

Firstly, currently available hyperspectral sensor technical characteristics represent a constraint [7]. Across all the additional and high-priority products, performance is closely linked to the following:

- Spectral coverage: More than 70% of the studies require a full-range spectrum VNIR-SWIR (400–2500 nm), especially for vegetation assessment but also for mineral detection.
- Spatial resolution: Studies based on UAV and airborne data showed higher accuracy then the satellite-based ones. High-resolution datasets derived from these platforms

https://doi.org/10.3390/rs18050798

are relevant in supervised ML approaches such as SVM and RF but also in object-based classification methods, where additional spatial details improve mapping.

- Revisit time and multi-sensor approach: The existing hyperspectral sensors are limited in the revisit time and are not operational yet. This limits the possibility of time-series analysis and a multi-sensor approach. In fact, the number of studies implementing multi-temporal or multi-sensor approaches is very limited.

Future missions such as CHIME, SBG and IRIDE will help to reduce some of these limitations. Behind these, their benefits will be fulfilled only if algorithm development evolves. In fact, the development of a robust processing chain that integrates hybrid models can reduce the impact of the technical limits of these sensors, for example, to improve spatial or temporal resolution [80,81]. Within this landscape, hybrid approaches emerge as a particularly promising direction [82–85]. Physics-based components can introduce physical consistency, constrain the solution space and support the estimation of uncertainty, while ML/DL models can capture complex spectral–spatial patterns and integrate information from multiple sensors. In addition, recent Foundation Models (FMs), large-scale hyperspectral classification frameworks such as SpectralEarth, HyperSIGMA, and HyperSL, were not systematically reviewed, as they primarily target representation learning and specific classification benchmarks rather than product-specific variable retrieval; however, they represent a promising basis for future hyperspectral data processing [82–85].

Finally, our analysis revealed that differences in accuracy are also driven by ground-truth datasets. Studies with dense and well-distributed field data, derived from field campaigns, gauging stations or airborne campaigns, show higher product accuracy and higher algorithm robustness. This is particularly relevant in DL and in hybrid models, which require diverse training data to achieve generalization of the model. On the other side, models applied at local scale, with small training datasets, are limited in performance and generalizability [85]. This highlights the need for open benchmark datasets, Cal/Val infrastructures and shared and general validation protocols, especially in view of the new hyperspectral missions such as CHIME and IRIDE [41].

Taken together, these elements suggest three key priorities for future research: (1) the development of future and upcoming hyperspectral missions, such as CHIME and IRIDE (PLATiNO), should take into account the technical requirements to improve the image availability and quality; (2) it is necessary to develop and improve the design of physics-informed ML architectures that embed physics-based spectral references; and (3) it is necessary to enhance systematic field dataset construction for algorithm training and validation of the products. These findings are in line with the conclusion of a recent study by Berger et al. [17].

8. Conclusions and Future Research Directions

This review provides a comprehensive and comparative analysis of the main algorithmic strategies and sensor characteristics in hyperspectral data processing (across two application domains and four high-level products). These product categories, selected based on their alignment with CHIME's HPPs and additional products, are deeply connected to an emerging global need in environmental monitoring, sustainable resource management and climate policy frameworks, as highlighted in the Introduction.

While each of these products presents specific challenges and signal characteristics, several cross-cutting trends emerge that are particularly relevant in view of upcoming hyperspectral missions such as CHIME, SBG, and IRIDE [7].

A core finding is that a single category of algorithms (C/PB, ML, DL) cannot universally outperform others across all use cases. Instead, performance is strictly dependent on the physical properties of the target, data quality, data availability, scale, and the intended

https://doi.org/10.3390/rs18050798

operational use. A relevant result is the growing importance of hybrid models, which combines the interpretability and transferability of the C/PB methods with the adaptability and automation of ML/DL. Although still underrepresented in the literature, these approaches are increasingly seen as the most promising for ensuring both robustness and scalability, two features that are essential for transitioning from research-grade analyses to operational services.

Across the different application domains, the use of ML/DL methods varies according to the specific study case. Applications involving high-dimensionality data and relevant spatial heterogeneity, such as the vegetation biophysical assessment, benefit more from ML/DL versatility. Instead, use cases involving well-characterized and specific spectral features, like pure mineral surfaces, are still relying on C/PB methods, which are still needed for these types of analysis. This uneven distribution of algorithmic needs reflects diverse user communities. These communities include research institutions seeking, for example, biophysical indicators. Resource management agencies need routine mineral monitoring. Climate-focused bodies are pushing for transparent, near-real-time methane assessments.

These results highlight an important shift: algorithm assessment and development are no longer purely related to the technical–scientific research field but increasingly driven by operational goals and policy frameworks. Documents, such as the European Green Deal, the Critical Raw Materials Act and the Global Methane Pledge, all point to EO, in particular to the new and high-potentiality sensors such as the hyperspectral ones, as a strategic enabler. In fact, in response to these needs, space agencies like ESA, NASA and ASI are designing new missions, such as CHIME, SBG and PLATiNO, from the IRIDE Constellation, with a relevant focus on algorithm maturity and an open access processing chain to support higher-level products based on that.

Based on the evidence gathered in this review, starting from the need for VNIR-SWIR full-spectral coverage to the impact of spatial resolution, revisit time, etc., to address these constraints, not only is an advancement in sensors and algorithms required but also a tailored and complementary muti-mission strategy.

In order to achieve success in present and future hyperspectral missions, it is essential to undertake a comprehensive evaluation of the various prospective research directions. This evaluation must encompass the development of hybrid processing chains, the enhancement of construction and the enhancement of availability of open training datasets, the establishment of a standardized accuracy assessment protocol, and the creation of interoperable architecture. These measures are imperative to ensure integration across a multitude of platforms and missions, including hyperspectral ones, thereby facilitating the development of a broad spectrum of multi-source operational services.

In this evolving landscape, it becomes clear that future success will not depend on replacing one class of methods with another. Instead, it will be achieved by bringing together the explanatory power of C/PB models and the flexibility of the ML/DL framework, guided by user needs and mission constraints. This convergence is not just a technical upgrade; it is a key pathway to unlock the full potential of hyperspectral data in support of environmental monitoring, resource security, climate action, policy implementation, scientific monitoring, and industrial innovation.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18050798/s1, Table S1: Data-extraction template, listing the column names extracted during the screening phase.

Author Contributions: Conceptualization, A.T., S.L. and E.V.; methodology, S.L. and S.S.; formal analysis, S.L., S.S. and A.N.X.; investigation, A.T., S.L., A.N.X. and E.V.; resources, A.T., A.N.X., S.M. and E.V.; data curation, S.L.; writing—original draft preparation, S.L. and A.T.; writing—review and editing, S.L., A.N.X. and S.S.; visualization, S.L., A.N.X., S.M. and S.S.; supervision, A.T., E.V. and A.N.X.; project administration, A.T. and A.N.X.; funding acquisition, A.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the CHEES Chime End to End Mission Performance Simulator ISPRA-GMV Contract No. GMVAD CHEES/ISPRA 30747/22 and Space It Up project funded by the Italian Space Agency, ASI, and the Ministry of University and Research, MUR, under contract n. 2024-5-E.0-CUP n. I53D24000060005.

Institutional Review Board Statement: This review was not registered in PROSPERO or any other database, and no formal protocol was prepared.

Data Availability Statement: No original datasets or code were generated for this review. Extracted data tables used for analysis are available from the corresponding author upon reasonable request.

Acknowledgments: This paper and related research have been conducted during and with the support of the Italian national in-ter-university PhD course in Earth Observation of the University of Rome "Sapienza", co-funded by Scuola Universitaria Superiore (IUSS) of Pavia and by Aresys s.r.l.

Conflicts of Interest: Author Simone Mancon was employed by the company Aresys S.r.l. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Schodlok, M.C.; Frei, M.; Segl, K. Implications of new hyperspectral satellites for raw materials exploration. *Miner. Econ.* **2022**, *35*, 495–502. [CrossRef]
2. Inoue, Y.; Guérif, M.; Baret, F.; Skidmore, A.; Gitelson, A.; Schlerf, M.; Olioso, A. *Simple and Robust Methods for Remote Sensing of Canopy Chlorophyll Content: A Comparative Analysis of Hyperspectral Data for Different Types of Vegetation*; Wiley Online Library: Hoboken, NJ, USA, 2016. [CrossRef]
3. Owalekar, K.; Italia, U.; Vijeta; Deshpande, S. Comparative assessment of methane leak detection using hyperspectral data. In *Proceedings of the International Conference on Data Management, Analytics & Innovation*; Springer: Berlin, Germany, 2023; pp. 783–815.
4. Manolakis, D.; Truslow, E.; Pieper, M.; Cooley, T.; Brueggeman, M. Detection Algorithms in Hyperspectral Imaging Systems: An Overview of Practical Algorithms. *IEEE Signal Process. Mag.* **2014**, *31*, 24–33. [CrossRef]
5. Stanzione, V.; Contini, C.; Mafficini, A.; Novi, S.; Ficco, G.; Meini, M.; Mastrandrea, M.; Fidanzati, P.; Pulcino, V.; Ansalone, L. PLATiNO multimission platform for ESA (Italian PNRR) IRIDE hyperspectral optical mission. In *Proceedings of the Small Satellites Systems and Services Symposium (4S 2024)*; Petrozzi-Ilstad, M., Ed.; SPIE: Bellingham, WA, USA, 2025; p. 37. [CrossRef]
6. Ghamisi, P.; Yokoya, N.; Li, J.; Liao, W.; Liu, S.; Plaza, J.; Rasti, B.; Plaza, A. Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 37–78. [CrossRef]
7. Schiavon, E.; Taramelli, A.; Tornato, A.; Lee, C.M.; Luvall, J.C.; Schollaert Uz, S.; Townsend, P.A.; Cima, V.; Geraldini, S.; Nguyen Xuan, A.; et al. Maximizing Societal Benefit Across Multiple Hyperspectral Earth Observation Missions: A User Needs Approach. *J. Geophys. Res. Biogeosci.* **2023**, *128*, e2023JG007569. [CrossRef]
8. Cawse-Nicholson, K.; Townsend, P.A.; Schimel, D.; Assiri, A.M.; Blake, P.L.; Buongiorno, M.F.; Campbell, P.; Carmon, N.; Casey, K.A.; Correa Pabon, R.E.; et al. NASA's surface biology and geology designated observable: A perspective on surface imaging algorithms. *Remote Sens. Environ.* **2021**, *257*, 112349. [CrossRef]
9. Chakravarty, S.; Paikaray, B.K.; Mishra, R.; Dash, S. Hyperspectral Image Classification using Spectral Angle Mapper. In *Proceedings of the 2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*; IEEE: New York, NY, USA, 2021; pp. 87–90. [CrossRef]
10. Valentini, E.; Taramelli, A.; Marinelli, C.; Martin, L.P.; Fassari, M.; Troffa, S.; Mzid, N.; Casa, R.; Pignatti, S. Hyperspectral Mixture Models in the CHIME Mission Implementation for Topsoil Texture Retrieval. *J. Geophys. Res. Biogeosci.* **2023**, *128*, e2022JG007272. [CrossRef]
11. Shafri, H.Z.M. Machine Learning in Hyperspectral and Multispectral Remote Sensing Data Analysis. In *Artificial Intelligence Science and Technology*; World Scientific: Singapore, 2017; pp. 3–9. [CrossRef]
12. Hool, A.; Helbig, C.; Wierink, G. Challenges and opportunities of the European Critical Raw Materials Act. *Miner. Econ.* **2024**, *37*, 661–668. [CrossRef]

13. Malley, C.S.; Borgford-Parnell, N.; Haeussling, S.; Howard, I.C.; Lefèvre, E.N.; Kuylenstierna, J.C.I. A roadmap to achieve the global methane pledge. *Environ. Res. Clim.* **2023**, *2*, 011003. [CrossRef]
14. Pascual-Venteo, A.B.; Portalés, E.; Berger, K.; Tagliabue, G.; Garcia, J.L.; Pérez-Suay, A.; Rivera-Caceido, J.P.; Verrelst, J. Prototyping Crop Traits Retrieval Models for CHIME: Dimensionality Reduction Strategies Applied to PRISMA Data. *Remote Sens.* **2022**, *14*, 2448. [CrossRef]
15. European Space Agency and Copernicus. Copernicus Hyperspectral Imaging Mission for the Environment Mission Requirements Document, ESA-EOPSM-CHIM-MRD-3216, Version 3.0. 2021. Available online: https://sentiwiki.copernicus.eu/__attachments/1678167/ESA-EOPSM-CHIM-MRD-3216%20-%20Copernicus%20Hyperspectral%20Imaging%20Mission%20for%20the%20Environment%20MRD%202021%20-%203.0.pdf?inst-v=19224347-b78a-4e98-8878-0bb2ef5b4589 (accessed on 15 January 2024).
16. Taramelli, A.; Tornato, A.; Magliozzi, M.L.; Mariani, S.; Valentini, E.; Zavagli, M.; Costantini, M.; Nieke, J.; Adams, J.; Rast, M. An interaction methodology to collect and assess user-driven requirements to define potential opportunities of future hyperspectral imaging sentinel mission. *Remote Sens.* **2020**, *12*, 1286. [CrossRef]
17. Berger, K.; Hostert, P.; Schlerf, M.; Immitzer, M.; Szantoi, Z.; Okujeni, A.; Foerster, S.; Colditz, R.; Giardino, C.; Machwitz, M.; et al. Advancing optical earth observation for EU policies: Needs, opportunities, recommendations. *Environ. Sci. Eur.* **2026**, *38*, 52. [CrossRef]
18. Matemilola, S.; Fadeyi, O.; Sijuade, T. Paris agreement. In *Encyclopedia of Sustainable Management*; Springer International Publishing: Cham, Switzerland, 2023; pp. 2555–2559. [CrossRef]
19. Chiu, I.H.Y. The EU sustainable finance agenda: Developing governance for double materiality in sustainability metrics. *Eur. Bus. Organ. Law Rev.* **2022**, *23*, 87–123. [CrossRef]
20. Shaik, R.U.; Periasamy, S.; Zeng, W.P. Potential Assessment of PRISMA Hyperspectral Imagery for Remote Sensing Applications. *Remote Sens.* **2023**, *15*, 1378. [CrossRef]
21. Van der Meer, F.D.; Van der Werff, H.M.; Van Ruitenbeek, F.J.; Hecker, C.A.; Bakker, W.H.; Noomen, M.F.; Van der Meijde, M.; Carranza, E.J.M.; De Smeth, J.B.; Woldai, T. Multi-and hyperspectral geologic remote sensing: A review. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *14*, 112–128. [CrossRef]
22. Kumar, S.; Arevalo, I.; Manjunath, B.S. Methane Mapper: Spectral Absorption Aware Hyperspectral Transformer for Methane Detection. Available online: https://github.com/UCSB-VRL/MethaneMapper-Spectral-Absorption-aware-Hyperspectral-Transformer-for-Methane-Detection (accessed on 5 November 2024).
23. Yang, G.; Huang, K.; Sun, W.; Meng, X.; Mao, D.; Ge, Y. Enhanced mangrove vegetation index based on hyperspectral images for mapping mangrove. *ISPRS J. Photogramm. Remote Sens.* **2022**, *189*, 236–254. [CrossRef]
24. Qian, S.-E. Hyperspectral Satellites, Evolution, and Development History. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7032–7056. [CrossRef]
25. Jia, J.; Wang, Y.; Chen, J.; Guo, R.; Shu, R.; Wang, J. Status and application of advanced airborne hyperspectral imaging technology: A review. *Infrared Phys. Technol.* **2020**, *104*, 103115. [CrossRef]
26. Adão, T.; Hruška, J.; Pádua, L.; Bessa, J.; Peres, E.; Morais, R.; Sousa, J.J. Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sens.* **2017**, *9*, 1110. [CrossRef]
27. Benelli, A.; Cevoli, C.; Fabbri, A. In-field hyperspectral imaging: An overview on the ground-based applications in agriculture. *J. Agric. Eng.* **2020**, *51*, 129–139. [CrossRef]
28. Aburaed, N.; Alkhatib, M.Q.; Marshall, S.; Zabalza, J.; Al Ahmad, H. A Review of Spatial Enhancement of Hyperspectral Remote Sensing Imaging Techniques. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 2275–2300. [CrossRef]
29. Neinavaz, E.; Schlerf, M.; Darvishzadeh, R.; Gerhards, M.; Skidmore, A.K. Thermal infrared remote sensing of vegetation: Current status and perspectives. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102415. [CrossRef]
30. Lee, C.M.; Glenn, N.F.; Stavros, E.N.; Luvall, J.; Yuen, K.; Hain, C.; Schollaert Uz, S. Systematic integration of applications into the Surface Biology and Geology (SBG) Earth mission architecture study. *J. Geophys. Res. Biogeosci.* **2022**, *127*, e2021JG006720. [CrossRef]
31. Rao, D.A.; Guha, A. Potential Utility of Spectral Angle Mapper and Spectral Information Divergence Methods for mapping lower Vindhyan Rocks and Their Accuracy Assessment with Respect to Conventional Lithological Map in Jharkhand, India. *J. Indian Soc. Remote Sens.* **2018**, *46*, 737–747. [CrossRef]
32. Gomez, C.; Lagacherie, P.; Coulouma, G. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma* **2008**, *148*, 141–148. [CrossRef]
33. Liu, H.; Li, W.; Xia, X.-G.; Zhang, M.; Guo, Z.; Song, L. SegHSI: Semantic Segmentation of Hyperspectral Images with Limited Labeled Pixels. *IEEE Trans. Image Process.* **2024**, *33*, 6469–6482. [CrossRef] [PubMed]
34. Kumar, V.; Venkatachalaperumal, S.R.; Persello, C. Synergistic fusion of spaceborne polarimetric SAR and hyperspectral data for land cover classification. In *Spaceborne Synthetic Aperture Radar Remote Sensing*; CRC Press: Boca Raton, FL, USA, 2023; pp. 169–209.

35. Righini, M.; Valentini, E.; Sapio, S.; Marinelli, C.; Gatti, I.; Jimenez, M.J.; Bresciani, M.; Giardino, C.; Pinardi, M.; Boschetti, M.; et al. Dynamic land cover mapping exploiting hyperspectral PRISMA data. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2023)*; IEEE: New York, NY, USA, 2023; pp. 1497–1500.
36. Rast, M.; Nieke, J.; Adams, J.; Isola, C.; Gascon, F. Copernicus Hyperspectral Imaging Mission for the Environment (CHIME). In *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2021; pp. 108–111. [CrossRef]
37. Nieke, J.; Despoisse, L.; Gabriele, A.; Weber, H.; Strese, H.; Ghasemi, N.; Gascon, F.; Alonso, K.; Boccia, V.; Tsonevska, B.; et al. The Copernicus Hyperspectral Imaging Mission for the Environment (CHIME): An overview of its mission, system and planning status. In *Sensors, Systems, and Next-Generation Satellites XXVII*; SPIE: Bellingham, WA, USA, 2023; Volume 12729, pp. 21–40. [CrossRef]
38. Im, J.; Jensen, J.R. Hyperspectral Remote Sensing of Vegetation. *Geogr. Compass* **2008**, *2*, 1943–1961. [CrossRef]
39. Räsänen, A.; Aurela, M.; Juutinen, S.; Kumpula, T.; Lohila, A.; Penttilä, T.; Virtanen, T. Detecting northern peatland vegetation patterns at ultra-high spatial resolution. *Remote Sens. Ecol. Conserv.* **2020**, *6*, 457–471. [CrossRef]
40. Bian, L.; Wang, Z.; Zhang, Y.; Li, L.; Zhang, Y.; Yang, C.; Fang, W.; Zhao, J.; Zhu, C.; Meng, Q.; et al. A broadband hyperspectral image sensor with high spatio-temporal resolution. *Nature* **2024**, *635*, 73–81. [CrossRef] [PubMed]
41. Boccia, V.; Adams, J.; Thome, K.J.; Turpie, K.R.; Kokaly, R.; Bouvet, M.; Green, R.O.; Rast, M. NASA–ESA cooperation on the SBG and CHIME hyperspectral satellite missions: A roadmap for the Joint Working Group on Cal/Val activities. In *EGU General Assembly 2021, Conference Abstracts*; European Geosciences Union: Vienna, Austria, 2021; p. EGU21-15166. [CrossRef]
42. Pearlshtien, D.H.; Pignatti, S.; Ben-Dor, E. Vicarious CAL/VAL Approach for Orbital Hyperspectral Sensors Using Multiple Sites. *Remote Sens.* **2023**, *15*, 771. [CrossRef]
43. Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral Remote Sensing Data Analysis and Future Challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36. [CrossRef]
44. Habashi, J.; Moghadam, H.J.; Oskouei, M.M.; Pour, A.B.; Hashim, M. PRISMA Hyperspectral Remote Sensing Data for Mapping Alteration Minerals in Sar-e-Châh-e-Shur Region, Birjand, Iran. *Remote Sens.* **2024**, *16*, 1277. [CrossRef]
45. Green, R.O.; Mahowald, N.; Ung, C.; Thompson, D.R.; Bator, L.; Bennett, M.; Bernas, M.; Blackway, N.; Bradley, C.; Cha, J.; et al. The Earth Surface Mineral Dust Source Investigation: An Earth science imaging spectroscopy mission. In *Proceedings of the 2020 IEEE Aerospace Conference, Big Sky, MT, USA, 7-14 March 2020*; IEEE: New York, NY, USA, 2020; pp. 1–15. [CrossRef]
46. Pearlshtien, D.H.; Pignatti, S.; Greisman-Ran, U.; Ben-Dor, E. PRISMA sensor evaluation: A case study of mineral mapping performance over Makhtesh Ramon, Israel. *Int. J. Remote Sens.* **2021**, *42*, 5884–5916. [CrossRef]
47. Peyghambari, S.; Zhang, Y. Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: An updated review. *J. Appl. Remote Sens.* **2021**, *15*, 031501. [CrossRef]
48. Asadzadeh, S.; Zhou, X.D.; Chabrillat, S. Assessment of the spaceborne EnMAP hyperspectral data for alteration mineral mapping: A case study of the Reko Diq porphyry Cu-Au deposit, Pakistan. *Remote Sens. Environ.* **2024**, *314*, 114389. [CrossRef]
49. Hajaj, S.; El Harti, A.; Pour, A.B.; Jellouli, A.; Adiri, Z.; Hashim, M. A review on hyperspectral imagery application for lithological mapping and mineral prospecting: Machine learning techniques and future prospects. *Remote Sens. Appl.* **2024**, *35*, 101218. [CrossRef]
50. Udelhoven, T.; Schlerf, M.; Segl, K.; Mallick, K.; Bossung, C.; Retzlaff, R.; Rock, G.; Fischer, P.; Müller, A.; Storch, T.; et al. A satellite-based imaging instrumentation concept for hyperspectral thermal remote sensing. *Sensors* **2017**, *17*, 1542. [CrossRef]
51. Vicente, L.E.; de Souza, C.R. Identification of mineral components in tropical soils using reflectance spectroscopy and advanced spaceborne thermal emission and reflection radiometer (ASTER) data. *Remote Sens. Environ.* **2011**, *115*, 1824–1836. [CrossRef]
52. Gålfalk, M.; Olofsson, G.; Bastviken, D. Approaches for hyperspectral remote flux quantification and visualization of GHGs in the environment. *Remote Sens. Environ.* **2017**, *191*, 81–94. [CrossRef]
53. Jiang, Y.; Zhang, L.; Zhang, X.; Cao, X. Methane Retrieval Algorithms Based on Satellite: A Review. *Atmosphere* **2024**, *15*, 449. [CrossRef]
54. Xiao, C.; Fu, B.; Shui, H.; Guo, Z.; Zhu, J. Detecting the Sources of Methane Emission from Oil Shale Mining and Processing Using Airborne Hyperspectral Data. *Remote Sens.* **2020**, *12*, 537. [CrossRef]
55. Dennison, P.E.; Thorpe, A.K.; Pardyjak, E.R.; Roberts, D.A.; Qi, Y.; Green, R.O.; Bradley, E.S.; Funk, C.C. High spatial resolution mapping of elevated atmospheric carbon dioxide using airborne imaging spectroscopy: Radiative transfer modeling and power plant plume detection. *Remote Sens. Environ.* **2013**, *139*, 116–129. [CrossRef]
56. Xu, D.Q.; Ni, G.Q.; Jiang, L.L.; Shen, Y.T.; Li, T.; Ge, S.L.; Shu, X.B. Exploring for natural gas using reflectance spectra of surface soils. *Adv. Space Res.* **2008**, *41*, 1800–1817. [CrossRef]
57. Ayasse, A.K.; Dennison, P.E.; Foote, M.; Thorpe, A.K.; Joshi, S.; Green, R.O.; Duren, R.M.; Thompson, D.R.; Roberts, D.A. Methane mapping with future satellite imaging spectrometers. *Remote Sens.* **2019**, *11*, 3054. [CrossRef]

58. Thorpe, A.K.; Roberts, D.A.; Bradley, E.S.; Funk, C.C.; Dennison, P.E.; Leifer, I. High resolution mapping of methane emissions from marine and terrestrial sources using a Cluster-Tuned Matched Filter technique and imaging spectrometry. *Remote Sens. Environ.* **2013**, *134*, 305–318. [CrossRef]
59. Gewali, U.B.; Monteiro, S.T.; Saber, E. Machine learning based hyperspectral image analysis: A survey. *arXiv* **2018**, arXiv:1802.08701.
60. Sr, W.L.S.; Revercomb, H.E.; Zhou, D.K.; Huang, H.-L.A. Hyperspectral sounding: A revolutionary advance in atmospheric remote sensing. In *Multispectral and Hyperspectral Remote Sensing Instruments and Applications II*; SPIE: Bellingham, WA, USA, 2005; pp. 1–11.
61. Wijesingha, J.; Dayananda, S.; Wachendorf, M.; Astor, T. Comparison of Spaceborne and UAV-Borne Remote Sensing Spectral Data for Estimating Monsoon Crop Vegetation Parameters. *Sensors* **2021**, *21*, 2886. [CrossRef] [PubMed]
62. Zhu, J.; Lu, J.; Li, W.; Wang, Y.; Jiang, J.; Cheng, T.; Zhu, Y.; Cao, W.; Yao, X. Estimation of canopy water content for wheat through combining radiative transfer model and machine learning. *Field Crops Res.* **2023**, *302*, 109077. [CrossRef]
63. Liu, L.; Li, S.; Yang, W.; Wang, X.; Luo, X.; Ran, P.; Zhang, H. Forest canopy water content monitoring using radiative transfer models and machine learning. *Forests* **2023**, *14*, 1418. [CrossRef]
64. Yang, X.; Yu, Y.; Fan, W. Chlorophyll content retrieval from hyperspectral remote sensing imagery. *Environ. Monit. Assess.* **2015**, *187*, 456. [CrossRef]
65. Ta, N.; Chang, Q.; Zhang, Y. Estimation of Apple Tree Leaf Chlorophyll Content Based on Machine Learning Methods. *Remote Sens.* **2021**, *13*, 3902. [CrossRef]
66. Clevers, J.G.; Kooistra, L. Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *5*, 574–583. [CrossRef]
67. Valentini, E.; Nghiem, S.V.; Thy, P.T.M.; Sapio, S.; Bresciani, M.; Giardino, C.; Boschetti, M.; Nutini, F.; Pellegrino, A.; Le Trung, C.; et al. Multisensor mapping of land use/land cover pattern in mangrove ecosystems: A case study in Cần Giờ Biosphere Reserve, Vietnam. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 20776–20789. [CrossRef]
68. De Luca, G.; Carotenuto, F.; Genesio, L.; Pepe, M.; Toscano, P.; Boschetti, M.; Miglietta, F.; Gioli, B. Improving PRISMA hyperspectral spatial resolution and geolocation by using Sentinel-2: Development and test of an operational procedure in urban and rural areas. *ISPRS J. Photogramm. Remote Sens.* **2024**, *215*, 112–135. [CrossRef]
69. Wocher, M.; Berger, K.; Danner, M.; Mauser, W.; Hank, T. RTM-based dynamic absorption integrals for the retrieval of biochemical vegetation traits. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *93*, 102219.
70. Chen, X.; Warner, T.A.; Campagna, D.J. Integrating visible, near-infrared and short-wave infrared hyperspectral and multispectral thermal imagery for geological mapping at Cuprite, Nevada: A rule-based system. *Int. J. Remote Sens.* **2010**, *31*, 1733–1752. [CrossRef]
71. Kopačková, V.; Koucká, L. Integration of Absorption Feature Information from Visible to Longwave Infrared Spectral Ranges for Mineral Mapping. *Remote Sens.* **2017**, *9*, 1006. [CrossRef]
72. Johnson, K.E.; Koperski, K. VNIR-SWIR Superspectral Mineral Mapping: An Example from Cuprite, Nevada. *Photogramm. Eng. Remote Sens.* **2020**, *86*, 695–700. [CrossRef]
73. Koirala, B.; Khodadadzadeh, M.; Contreras, C.; Zahiri, Z.; Gloaguen, R.; Scheunders, P. A Supervised Method for Nonlinear Hyperspectral Unmixing. *Remote Sens.* **2019**, *11*, 2458. [CrossRef]
74. Adep, R.N.; Vijayan, A.P.; Shetty, A.; Ramesh, H. Performance evaluation of hyperspectral classification algorithms on AVIRIS mineral data. *Perspect. Sci.* **2016**, *8*, 722–726. [CrossRef]
75. Frankenberg, C.; Thorpe, A.K.; Thompson, D.R.; Hulley, G.; Kort, E.A.; Vance, N.; Borchardt, J.; Krings, T.; Gerilowski, K.; Sweeney, C.; et al. Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 9734–9739. [CrossRef]
76. Pieper, M.L.; Lockwood, R.; Chrisp, M. Hyperspectral SWIR sensor parameterization for optimal methane detection. In *Imaging Spectrometry XXVI: Applications, Sensors, and Processing*; SPIE: Bellingham, WA, USA, 2023; pp. 112–127.
77. Masiello, G.; Mastro, P.; Serio, C.; Falabella, F.; Pasquariello, P. Methane profile retrieval from IASI: A deep learning inversion approach based on feed-forward neural networks. In *Remote Sensing of Clouds and the Atmosphere XXVII*; Comerón, A., Kassianov, E.I., Schäfer, K., Picard, R.H., Weber, K., Singh, U.N., Eds.; SPIE: Bellingham, WA, USA, 2022; p. 10. [CrossRef]
78. Guanter, L.; Irakulis-Loitxate, I.; Gorroño, J.; Sánchez-García, E.; Cusworth, D.H.; Varon, D.J.; Cogliati, S.; Colombo, R. Mapping methane point emissions with the PRISMA spaceborne imaging spectrometer. *Remote Sens. Environ.* **2021**, *265*, 112671. [CrossRef]
79. Li, T.; Liu, T.; Wang, Y.; Li, X.; Gu, Y. Spectral Reconstruction Network from Multispectral Images to Hyperspectral Images: A Multitemporal Case. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5535016. [CrossRef]
80. Muhammad, U.; Laaksonen, J. Hybrid Deep Learning for Hyperspectral Single-Image Super-Resolution. *IEEE Geosci. Remote Sens. Lett.* **2025**, *22*, 5509905. [CrossRef]
81. Yang, J.; Wu, C.; Du, B.; Zhang, L. Enhanced Multiscale Feature Fusion Network for HSI Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10328–10347. [CrossRef]

82. Wang, D.; Hu, M.; Jin, Y.; Miao, Y.; Yang, J.; Xu, Y.; Qin, X.; Ma, J.; Sun, L.; Li, C.; et al. HyperSIGMA: Hyperspectral Intelligence Comprehension Foundation Model. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 6427–6444. [CrossRef]
83. Li, J.; Liu, Y.; Wang, X.; Peng, Y.; Sun, C.; Wang, S.; Sun, Z.; Ke, T.; Jiang, X.; Lu, T.; et al. HyperFree: A channel-adaptive and tuning-free foundation model for hyperspectral remote sensing imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2025)*; IEEE: New York, NY, USA, 2025; pp. 23048–23058.
84. Yang, J.; Du, B.; Zhang, L. From center to surrounding: An interactive learning framework for hyperspectral image classification. *ISPRS J. Photogramm. Remote Sens.* **2023**, *197*, 145–166. [CrossRef]
85. Li, S.; Hao, Q.; Gao, G.; Kang, X. The Effect of Ground Truth on Performance Evaluation of Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 7195–7206. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

remote sensing

MDPI

Article

Hyperspectral Image Classification Using SIFANet: A Dual-Branch Structure Combining CNN and Transformer

Yuannan Gui [1], Lu Xu [1,2,*], Dongping Ming [1,2,3,4], Yanfei Wei [5,6,7] and Ming Huang [6,8]

[1] School of Artificial Intelligence, China University of Geosciences (Beijing), Beijing 100083, China; guiyn@email.cugb.edu.cn (Y.G.); mingdp@cugb.edu.cn (D.M.)
[2] Hebei Key Laboratory of Geospatial Digital Twin and Collaborative Optimization, Beijing 100083, China
[3] Frontier Science Center for Deep-Time Digital Earth, China University of Geosciences (Beijing), Beijing 100083, China
[4] State Key Laboratory of Geological Processes and Mineral Resources, Beijing 100083, China
[5] School of Natural Resources and Geomatics, Nanning Normal University, Nanning 530100, China; yanfeiwei@nnnu.edu.cn
[6] University Engineering Research Center of "Satellite +" Space AI Intelligent Governance of Natural Resources, Nanning 530100, China; huangmingtyz@163.com
[7] Guangxi Engineering Research Center for Smart Monitoring and Governance of Agricultural Land, Nanning 530100, China
[8] Guangxi Zhuang Autonomous Region Institute of Geographic Information Surveying and Mapping, Liuzhou 545006, China
* Correspondence: xlirs@email.cugb.edu.cn

Highlights

What are the main findings?

- A novel dual-branch network, SIFANet, was developed, which synergistically integrates CNN and Transformers for enhanced HSI classification.
- The architecture successfully captures both local spectral–spatial features via the CNN branch and long-range global dependencies through the Transformer branch, overcoming the limitations of single-backbone models.

What are the implications of the main findings?

- SIFANet provides a robust solution for HSI classification in complex scenarios, significantly improving accuracy by effectively balancing local texture details and global contextual information.
- The dual-branch fusion strategy offers a scalable framework for hybrid deep learning models, potentially advancing the processing of high-dimensional remote sensing data with limited labeled samples.

Abstract

The hyperspectral image (HSI) is rich in spectral information and has important applications in the field of ground objects classification. However, HSI data have high dimensions and variable spatial–spectral features, which make it difficult for some models to adequately extract the effective features. Recent studies have shown that fusing spatial and spectral features can significantly improve accuracy by exploiting multi-dimensional correlations. Based on this, this article proposes a spectral integration and focused attention network (SIFANet) with a two-branch structure. SIFANet captures the local spatial features and global spectral dependencies through the parallel-designed spatial feature extractor (SFE) and spectral sequence Transformer (SST), respectively. A cross-module attention fusion (CMAF) mechanism dynamically integrates features from both branches before final classification. Experiments on the Salinas dataset and Xiong'an hyperspectral dataset show

that the overall accuracy on these two datasets is 99.89% and 99.79%, which is higher than the other models compared. The proposed method also had the lowest standard deviation of category accuracy and optimal computational efficiency metrics, demonstrating robust spatial–spectral feature integration for improved classification.

Keywords: hyperspectral image classification; convolutional neural network; transformer; attention mechanism; spatial-spectral feature fusion

1. Introduction

The hyperspectral image (HSI) is a three-dimensional data cube acquired by remote sensing platforms (such as aircrafts or satellites) using push-broom or staring imaging techniques [1,2]. It contains two-dimensional spatial images corresponding to multiple spectral bands. Unlike traditional multispectral images, HSIs collect continuous spectral information from hundreds of narrow bands. These bands span the visible, near-infrared, and shortwave infrared regions (400–2500 nm) of the electromagnetic spectrum. As a result, each pixel in the HSI contains reflectance value across hundreds of spectral channels, forming a continuous spectral curve [3]. This enables HSIs to provide rich spatial–spectral information [4]. In recent years, the rapid development of computer technology has greatly advanced HSI classification tasks. HSIs have been widely applied in agricultural production [5,6], city planning [7,8], environmental science [9,10], and other fields.

HSI classification faces challenges from three aspects: data, model training, and application. At the data level, the typical 200–300 spectral bands of HSIs lead to high dimensionality, raising computational complexity [11]. Moreover, the limited availability of ground truth training samples results in data sparsity in the high-dimensional feature space, violating statistical learning assumptions. This leads to the Hughes phenomenon. Studies have shown that when the ratio of feature dimensionality to sample size exceeds a critical threshold, the generalization performance of classifiers such as Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) will sharply decline due to overfitting. In addition, strong inter-band correlations will introduce substantial redundant information to models, which hinders effective learning and training [12]. At the model training level, large training sample sizes demand substantial computational resources, significantly increasing training costs [13]. In practical applications, HSI preprocessing involves complex procedures such as radiometric and atmospheric correction, geometric registration, and manual sample labeling [14]. Additionally, to further improve the quality of hyperspectral data, research on enhanced deep image prior has been applied to unsupervised hyperspectral image super-resolution [15], which effectively enhances the spatial resolution and reduces noise without the need for large-scale training datasets. Furthermore, cross-scene model transfer issues caused by various imaging angles and sensor inconsistencies present significant challenges [16]. These factors will greatly affect the HSI classification performance in practical applications. In response, researchers have proposed and refined various classification algorithms to address these challenges, including traditional machine learning methods, deep learning approaches, and multi-model fusion methods.

Traditional machine learning methods depend on manual feature engineering. The SVM uses kernel functions to map features into high-dimensional space and performs well especially in small-sample scenarios. Random forest (RF), using bagging and random subspaces, effectively handles high-dimensional nonlinear data. The K-Nearest Neighbor (K-NN) algorithm classifies samples by distance-based majority voting. Despite the

https://doi.org/10.3390/rs18030398

K-NN being simple, it faces limitations in high-dimensional spaces due to the "curse of dimensionality", which makes it difficult to achieve satisfactory classification results. Additionally, K-NN is sensitive to parameters and has lower computational efficiency. Champa et al. [17] proposed a hybrid technique which combined tree-based classifiers with feature dimensionality reduction. The results show that decision trees (DTs), extra trees (ETs), and RF perform well in HSI classification.

Compared to traditional methods, deep learning approaches offer better adaptability for HSI classification tasks. Convolutional neural networks (CNNs) automatically extract spatial–spectral features, mimicking biological visual mechanisms [18]. Residual neural networks (ResNet) incorporate residual structures for cross-layer information transfer. This structure effectively mitigates the vanishing gradient problem [19]. Li et al. [20] introduced the deep belief network (DBN) for feature extraction and image classification. Zhao et al. [21] proposed a feature-level fusion classification framework combining CNN with texture features, utilizing HSI and LiDAR data for high classification accuracy. As for sequential spectral data, the recurrent neural network (RNN) captures spectral sequence correlations via recursive feedback mechanisms [22]. Long short-term memory (LSTM) networks use gating units to alleviate long-term dependency issues [23]. Recently, the Transformer has gained popularity in natural language processing (NLP), which improves semantic extraction through global dependency modeling [24]. Inspired by this, researchers have integrated attention mechanisms into HSI classification to enhance accuracy by strengthening feature interactions. Hong et al. [25] proposed the SpectralFormer network, which uses self-attention to model long-range spectral dependencies. However, its unimodal (spectral-only) structure neglects critical spatial information of HSIs, limiting effectiveness. To address this issue, Wang et al. [26] introduced a 3D attention mechanism to capture joint spectral–spatial features. Ding et al. [27] developed a global–local Transformer network (GLT-Net) for joint classification using multi-scale feature fusion. However, inherent architectural constraints limit the performance of single-structure models. Specifically, CNNs rely on local receptive fields, which make them effective at capturing local textures but less capable of modeling long-range spatial dependencies or the continuous sequential correlation of spectral bands. Conversely, while Transformers excel at global dependency modeling via self-attention, they lack the spatial inductive bias (such as translation invariance and locality) inherent in CNNs. This often leads to difficulties in capturing fine-grained local spatial structures and results in quadratic computational complexity when dealing with high-dimensional spectral sequences. Consequently, deep learning models still face two key limitations: they require large amounts of annotated data, which is costly and time-consuming [28], and they involve high computational complexity, hindering practical deployment [29]. To alleviate this, transudative few-shot learning with enhanced spectral–spatial embedding [30] has been proposed to achieve robust classification with minimal labeled samples.

In recent years, multi-model fusion methods have shown great promise in pattern recognition and image classification by leveraging the complementary strengths of different structures. Researchers optimize feature extraction and processing by combining various models. For example, Wei et al. [31] proposed a strategy that uses CNNs for feature extraction, followed by SVM classification. This approach not only improves accuracy but also reduces overfitting and alleviates reliance on large sample datasets. Combining multi-scale feature fusion with gradient boosting decision trees (GBDTs) has also been proven effective. This ensemble approach integrates multiple weak learners to enhance classification stability and accuracy, especially in complex land cover scenarios [32]. To better exploit the complementary nature of spatial and spectral information, researchers have increasingly turned to dual-branch architectures. For example, the Double-Branch Dual-Attention (DBDA)

https://doi.org/10.3390/rs18030398

network [33] utilizes two independent branches to capture spatial and spectral features, respectively, enhanced by dual-attention mechanisms. Similarly, the Multiscale Neighborhood Attention Transformer (MSNAT) [34] further refines this by extracting multi-scale features to adapt to varying land cover sizes. To address feature extraction from a frequency perspective, the Dual Frequency Transformer Network (DFTN) [35] was developed. It utilizes a dual-branch frequency domain feature extraction block to simultaneously capture high-frequency local details and low-frequency global variations, demonstrating the effectiveness of frequency domain information in enhancing HSI classification. Alkhatib et al. [36] proposed an attention-based dual-branch network that fuses features from a real-valued neural network (RVNN) and a complex-valued neural network (CVNN). By using the Fourier transformation to extract frequency information, this model enhances HSI classification performance. While these models have significantly improved classification accuracy, they often face challenges such as high parameter redundancy and suboptimal fusion of heterogeneous features from the two branches. Many researchers have focused on efficient spatial–spectral feature extraction. Yang et al. [3] introduced the multi-scale hybrid CNN–attention (MS-Hybrid-A) network, which uses 3D convolutions for spectral–spatial features extraction. Additionally, it supplementary extracts spatial details with 2D convolutions and incorporates a convolutional block attention module (CBAM) to improve classification performance. Liang et al. [37], inspired by the Transformer, proposed HSI-Mixer, which utilizes a hybrid measurement-based linear projection (HMLP) module for deep spectral–spatial feature fusion. Kong et al. [38] developed a co-feature extraction framework that integrates graph embedding with deep learning. They constructed a supervised within-class/between-class hypergraph (SWBH) for spectral feature learning and introduced a random zero-masking strategy to generate augmented labeled samples. It could facilitate CNN-based spatial feature extraction and mitigate overfitting in small-sample settings. Ahmad et al. [39] proposed the spatial morphological Mamba (SMM) and spatial–spectral morphological Mamba (SSMM) networks, which employ depth-wise separable convolutions to implement morphological operations, like erosion and dilation. By leveraging State Space Models (SSMs), these Mamba-based approaches achieve linear computational complexity while effectively modeling long-range dependencies, offering a promising, efficient alternative to standard Transformers for processing high-dimensional spectral sequences. To address the challenge of cross-layer information loss, Chen et al. [40] designed a hybrid pooling attention (HPA) module and a cross-layer feature fusion (CFF) module to preserve crucial information during the propagation process. Gao et al. [41] introduced a plug-and-play adaptive feature fusion (AFF) module that processes multi-layer networks to better utilize spatial and spectral features. Guo et al. [42] proposed an adaptive score-weighting method to fuse features from spatial and spectral branches. Similarly, the concept of reference-based adaptive modulation has been explored in multi-style fusion tasks [43], providing insights into the dynamic adjustment of features across different domains.

In conclusion, traditional machine learning methods such as SVM and RF perform well on high-dimensional data when only a few labeled samples are available. However, they have limited capacity to model nonlinear relationships and thus struggle to process large-scale, high-dimensional data with complex noise. Deep learning methods, by contrast, greatly improve feature representation through hierarchical abstraction. Nevertheless, they require large quantities of high-quality labeled samples and bring drawbacks such as long training times and high computational cost. Furthermore, multi-model fusion methods can mitigate overfitting effectively and reduce dependence on labeled samples. However, they pose challenges in designing compatible modules and incur heavy computational overhead.

https://doi.org/10.3390/rs18030398

The existing research still lacks effective methods to dynamically fuse spatial and spectral features for HSI classification. Furthermore, the number of model parameters and computational overhead remain too high. In the broader field of remote sensing, highly efficient architectures like SFEARNet [44] have successfully combined semantic flow and edge-aware refinement for tasks such as change detection, demonstrating the importance of structural information in efficient network design. Furthermore, the hybrid CNN–Transformer architecture has been widely validated in diverse remote sensing applications, including cloud detection using Landsat/Sentinel data [45], wind speed sensing using Global Navigation Satellite System Reflectometry (GNSS-R) [46], and chlorophyll concentration inversion using SeaWiFS data [47]. These successes in handling complex, multi-modal remote sensing data provide a strong rationale for adopting a hybrid design to capture both local spatial details and global spectral dependencies in HSI classification. Building on the need for efficiency, this article proposes a dual-branch network called the spectral integration and focused attention network (SIFANet). SIFANet is built on a hybrid CNN–Transformer structure and incorporates a channel attention mechanism to emphasize important spectral bands. By optimizing feature extraction and fusion, SIFANet could improve HSI classification performance and enhance classification accuracy to a certain extent.

The main contributions of this article are as follows:

1. Efficient feature extraction structure
   This article designs a dual-branch network composed of a spatial feature extractor (SFE) and a spectral sequence Transformer (SST). The SFE is enhanced with residual blocks (RBs) to alleviate the vanishing gradient problem and accelerate convergence. Simultaneously, the SST incorporates a Conv-Former module to improve spectral feature extraction, enabling the efficient and parallel extraction of spatial–spectral features.
2. Cross-Module Attention Fusion (CMAF)
   This article introduces a channel attention-based CMAF mechanism to dynamically and adaptively fuse features from different branches, which significantly reduces information loss during the feature integration step.
3. Comprehensive HSI classification accuracy assessment indices
   This article develops a novel computation accuracy parameter efficiency (CAPE) index to quantify the computational efficiency of different models. In addition, the proposed evaluation index system (EIS) also includes classification accuracy metrics, confusion category performance, and computational efficiency index, enabling a comprehensive, multidimensional assessment of the model performance.

2. Methodology

The research workflow is illustrated in Figure 1. First, principal component analysis (PCA) was applied to reduce the dimensionality of the HSI data. To retain most of the information while significantly compressing the data, we retained the first few principal components (PCs) based on a cumulative variance contribution rate threshold of >99%. Specifically, for the Salinas dataset (originally 204 bands), the first 7 PCs were retained, preserving 99.12% of the spectral information. For the Xiong'an hyperspectral dataset (originally 256 bands), the first 6 PCs were retained, preserving 99.09% of the information. Subsequently, spatial and spectral feature tensors were generated via a masking operation and then were input into SIFANet. Within SIFANet, spatial and spectral features were captured through SFE and SST branches, respectively. CMAF was employed to achieve the dynamic weighted fusion. Finally, the fused features were fed into the classifier to

https://doi.org/10.3390/rs18030398

produce the result. During the validation step, SIFANet was comprehensively evaluated and compared against other comparable models using the EIS.

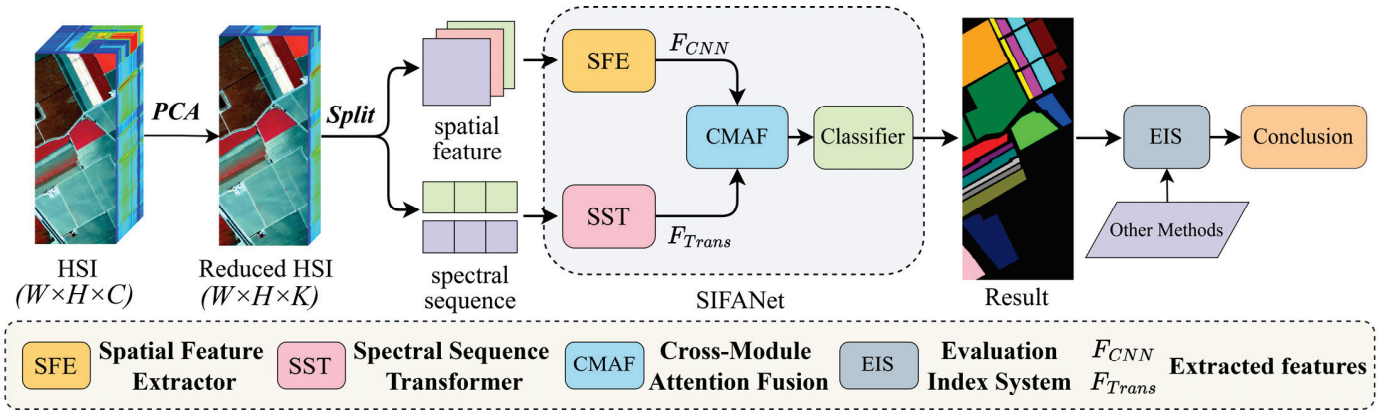


Figure 1. Workflow of the SIFANet.

2.1. SIFANet Structure

The SIFANet adopts a parallel dual-branch structure rather than a serial one to decouple the extraction of spatial and spectral features. While serial structures often lead to the loss of original information during sequential processing, the parallel design allows the SFE and SST to independently capture local textures and global dependencies. This simultaneous processing ensures that the distinctive characteristics of both domains are fully preserved before being dynamically integrated through the CMAF module, thereby maximizing feature complementarity.

2.1.1. Spatial Feature Extraction

SFE branch adopts a "Conv + RB" CNN variant structure. It applies residual learning to improve gradient flow and mitigate the vanishing gradient problem. Additionally, max-pooling downsampling is incorporated to reduce computational complexity and accelerate training.

Figure 2 illustrates the detailed structure of the SFE module. It begins by applying a 3×3 convolutional kernel to extract preliminary spatial features from local neighborhoods. To further reduce spatial dimensionality, a max-pooling layer is subsequently applied, formulated as

$$\boldsymbol{X}_{3d} = f_{MP}(f_{Conv}(\boldsymbol{X})) \tag{1}$$

where $\boldsymbol{X}$ is the input tensor, $\boldsymbol{X}_{3d}$ is the processing result, f_{Conv} and f_{MP} represent the 3×3 convolution layer and the max-pooling layer, respectively. f_{MP} uses a 2×2 window and a stride of 2 for pooling.

The RB module adopts a two-branch structure to facilitate feature reuse. The main branch consists of two consecutive "Conv + BatchNorm (BN)" layers, which are used to extract deeper spatial features. BN standardizes intermediate feature distributions across mini batches to mitigate the vanishing gradient problem and accelerate model convergence. The residual output is computed as follows:

$$\boldsymbol{X}_{BN} = \frac{\boldsymbol{X} - \overline{\boldsymbol{X}}}{\sqrt{\sigma_X^2 + \epsilon}} \cdot \gamma + \beta \tag{2}$$

where $\boldsymbol{X}$ is the residual block input tensor, $\boldsymbol{X}_{BN}$ is the output of the residual block, $\overline{\boldsymbol{X}}$ is the mean of the input tensor, σ_X^2 is the variance of the input data, ϵ is a small constant added for numerical stability, γ and β are learnable parameters.

https://doi.org/10.3390/rs18030398

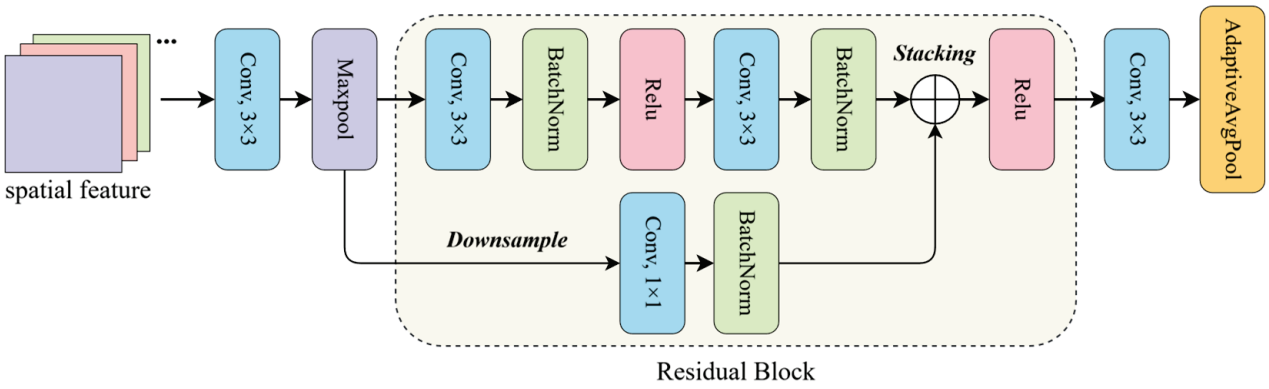


Figure 2. Spatial feature extractor.

Nonlinearity is introduced through the ReLU activation function, The operation is expressed as follows:

$$X^{'} = f_{Max}(0, X) \tag{3}$$

where X denotes the input tensor, $X^{'}$ is the output tensor, and f_{Max} represents the element-wise maximum function used ReLU.

The skip connection employs cross-layer identity mapping to directly propagate the shallow features X_{3d} to deeper layers, thereby preserving important spatial information. The fused output is obtained through element-wise feature summation. Denoting the sequence of operations in the main path as f_{RB}, and using $\oplus$ to represent summation, the operation can be defined as follows:

$$X_{RB} = X_{3d} \oplus f_{RB}(X_{3d}) \tag{4}$$

Finally, at the output stage, an adaptive average pooling layer is applied to adjust the spatial resolution of the feature maps to a fixed size, and the result is denoted as F_{CNN}. This operation ensures consistent output feature dimensions across both processing branches.

2.1.2. Spectral Sequence Transformer

Some land cover categories are difficult to distinguish when only relying on spatial features due to their high spatial similarity. To address this issue, this article proposes an SST branch, which is illustrated in Figure 3.

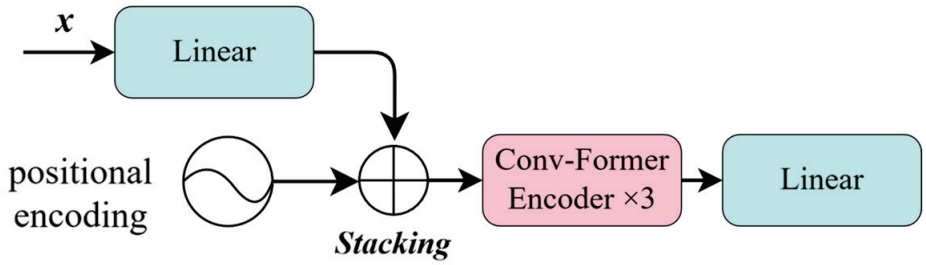


Figure 3. Spectral sequence Transformer.

A fully connected layer processes the input X, converting the spectral vector of each pixel into a token embedding. This transformation constructs a sequence of tokens X_{token}, forming the primary input sequence for the Transformer. Concurrently, a lower branch generates a learnable positional embedding P designed to encode the position. P is added element-wise to the token sequence to obtain the position-augmented input representation $X_{stacked}$, i.e.,

$$X_{stacked} = X_{token} \oplus P \tag{5}$$

https://doi.org/10.3390/rs18030398

The position-augmented spectral sequence $\boldsymbol{X_{stacked}}$ is then fed into the improved Conv-Former encoder module for feature encoding, producing $\boldsymbol{Z_{out}}$. Finally, a linear layer projects $\boldsymbol{Z_{out}}$ to match the output dimensions of the SFE branch. The final spectral feature representation is given by

$$\boldsymbol{F_{Trans}} = \boldsymbol{W_p Z_{out}} + \boldsymbol{b_p} \tag{6}$$

where $\boldsymbol{W_p}$ is the weight matrix of the linear layer, $\boldsymbol{b_p}$ is the bias vector of the linear layer, and $\boldsymbol{F_{Trans}}$ is the spectral feature information output by the SST branch.

This article combined convolution with a Transformer encoder to leverage the advantages of convolution operations and self-attention mechanisms. Inspired by Lin et al. [48], the improved "Conv-Former" structure is shown in Figure 4. We have significantly adapted the module for HSI-specific characteristics. Specifically, the conventional feed-forward network is replaced with a convolutional feed-forward network based on 1×1 convolutions, which efficiently reorganizes and compresses high-dimensional spectral channels while preserving spatial information. Batch normalization is employed throughout the module to improve training stability under the small-sample regime commonly encountered in HSI applications. Together, these modifications allow the proposed module to better capture discriminative spatial–spectral representations and make it more suitable for HSI analysis.

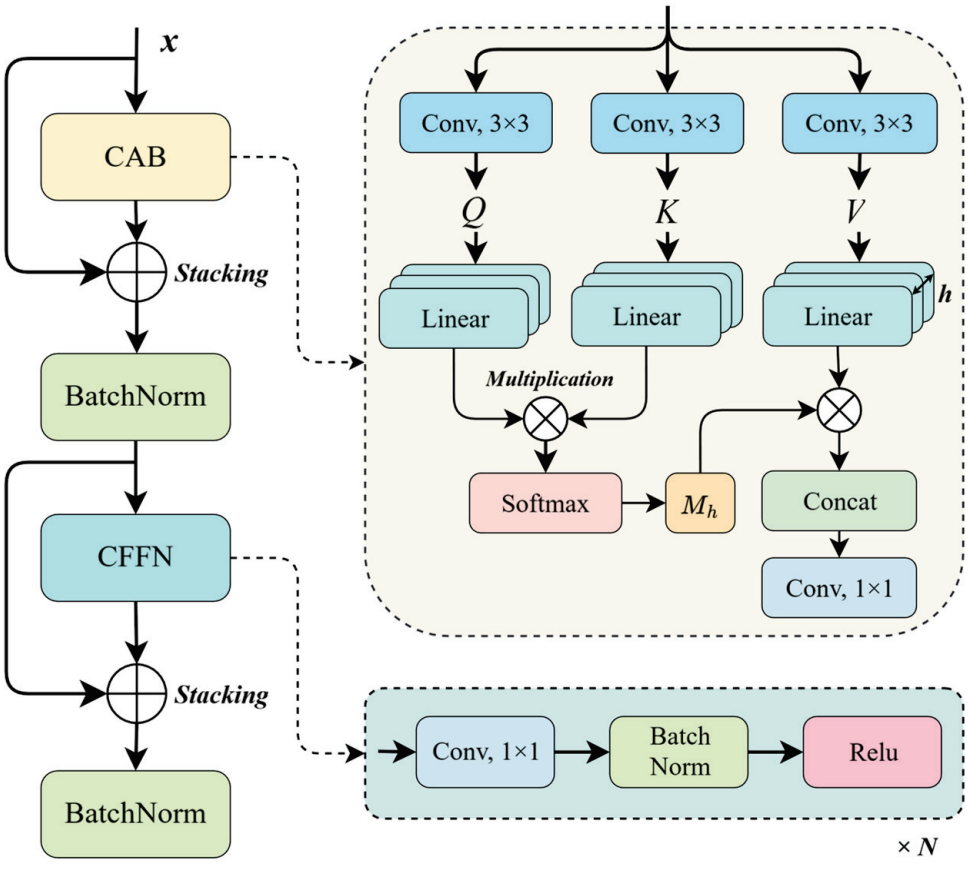


Figure 4. Improved Conv-Former encoder structure.

The convolution-attention residual module processes features through its dual-branch structure. The main branch extracts features with a convolution-attention block (CAB), while the auxiliary branch preserves the original input features. The output of the module is computed as

$$\boldsymbol{X_{CA}} = \boldsymbol{X} \oplus f_{CAB}(\boldsymbol{X}) \tag{7}$$

where $\boldsymbol{X}$ is the input tensor, $\boldsymbol{X_{CA}}$ is the module's output, and f_{CAB} denotes the operations within the CAB.

The CAB adopts a "Conv + Self-Attention" processing method. It first applies a 3×3 convolution to extract local spectral features, then constructs the query ($\boldsymbol{Q}$), key ($\boldsymbol{K}$), and value ($\boldsymbol{V}$) matrices for self-attention (SA) computation. A multi-head self-attention (MHSA) mechanism with $h = 4$ is employed to model global contextual relationships, balancing computational costs and model representational capacity. The SA process is defined as

$$\boldsymbol{M_h} = f_{SM}\left(\frac{\boldsymbol{Q} \otimes \boldsymbol{K^T}}{\sqrt{d_k}}\right) \tag{8}$$

$$SA(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \boldsymbol{M_h} \otimes \boldsymbol{V} \tag{9}$$

where d_k denotes the input sequence dimension, $\boldsymbol{M_h}$ represents the attention weight matrix, $\boldsymbol{K^T}$ is the transpose of $\boldsymbol{K}$, $\otimes$ indicates matrix multiplication, and f_{SM} is the SoftMax function.

The MHSA mechanism captures diverse dependencies and hierarchical relationships by integrating multiple self-attention heads [24]. The overall MHSA output is expressed as

$$MHSA(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = f_{CAT}(SA_1,\ SA_2, SA_3, \ldots, SA_h) \tag{10}$$

where $SA_i (i = 1, 2, 3, \ldots, h)$ represents the self-attention result of the i-th part. f_{CAT} is the Concat function.

Finally, a 1×1 convolution layer is applied to integrate the features obtained from global dependencies, resulting in the refined output $\boldsymbol{X_{SA}}$.

The current feed-forward network (CFFN) refines features from the upper layer through the following operation:

$$\boldsymbol{X_{CFFN}} = \boldsymbol{X_{CA}} \oplus f_{CFFN}(\boldsymbol{X_{CA}}) \tag{11}$$

where $\boldsymbol{X_{CFFN}}$ is the output and f_{CFFN} refers to the transformation process within the current feed-forward network (CFFN).

2.1.3. Cross-Module Attention Fusion

To facilitate cross-modal feature fusion, this article designs the CMAF module, as shown in Figure 5, inspired by the convolutional block attention module (CBAM) [49].

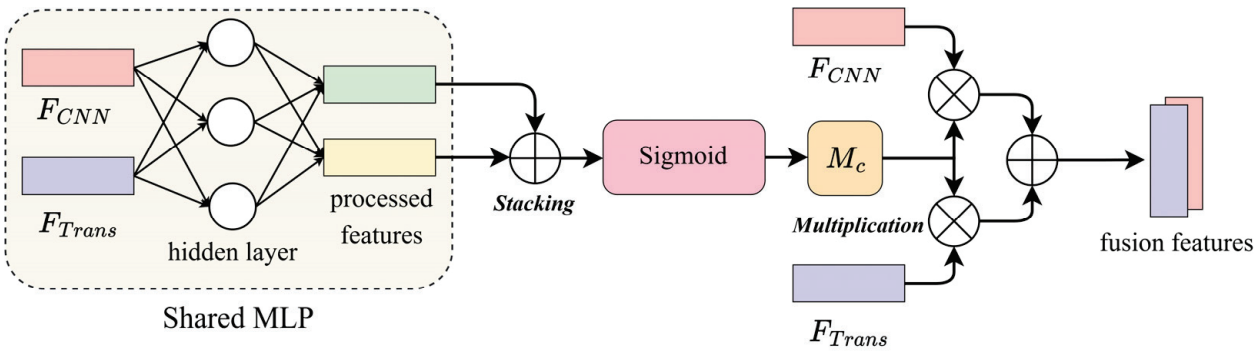


Figure 5. Cross-module attention fusion.

CMAF first constructs a channel attention weight matrix with a shared multi-layer perception (Shared MLP). Specifically, the features $\boldsymbol{F_{CNN}}$ and $\boldsymbol{F_{Trans}}$ are fed into the Shared MLP, where the input and hidden layers are connected via fully connected networks to enable feature interaction. After being processed by the hidden layer and activated by the ReLU function, the compressed and expanded feature vectors $\boldsymbol{F_{CNN}^{'}}$ and $\boldsymbol{F_{Trans}^{'}}$ are output

https://doi.org/10.3390/rs18030398

through another fully connected layer. Then, $\boldsymbol{F}_{\boldsymbol{CNN}}^{'}$ and $\boldsymbol{F}_{\boldsymbol{Trans}}^{'}$ are concatenated along the channel dimension, followed by Sigmoid activation to produce the channel attention weight matrix $\boldsymbol{M_c}$. The process is defined as follows:

$$\boldsymbol{s_{CNN}} = f_{Relu}\left(\boldsymbol{W}_0 \otimes \boldsymbol{F}_{\boldsymbol{CNN}}^{'}\right) \tag{12}$$

$$\boldsymbol{s_{Trans}} = f_{Relu}\left(\boldsymbol{W}_0 \otimes \boldsymbol{F}_{\boldsymbol{Trans}}^{'}\right) \tag{13}$$

$$\boldsymbol{M_c} = \sigma(\boldsymbol{W}_1 \otimes \boldsymbol{s_{CNN}} + \boldsymbol{W}_1 \otimes \boldsymbol{s_{Trans}}) \tag{14}$$

The above Equations (12) and (13) denote the hidden layer processing and ReLU activation, while (14) describes the steps of output and Sigmoid activation. σ is the Sigmoid function; $\boldsymbol{W}_0$ and $\boldsymbol{W}_1$ correspond to the weight matrices of the input layer and output layer, respectively. $\boldsymbol{s_{CNN}}$ is the spatial feature information after hidden layer processing, $\boldsymbol{s_{Trans}}$ is the spectral feature information after hidden layer processing, and f_{Relu} is the ReLU function.

After calculating $\boldsymbol{M_c}$, the original features $\boldsymbol{F_{CNN}}$ and $\boldsymbol{F_{Trans}}$ are, respectively, channel-weighted with $\boldsymbol{M_c}$ through matrix multiplication and the weighted results are stacked to obtain the fused feature $\boldsymbol{F_{fused}}$.

$$\boldsymbol{F_{fused}} = (\boldsymbol{M_c} \otimes \boldsymbol{F_{CNN}}) \oplus (\boldsymbol{M_c} \otimes \boldsymbol{F_{Trans}}) \tag{15}$$

Although the proposed CMAF module is inspired by classical channel attention mechanisms such as the Squeeze-and-Excitation Network (SE-Net) [50] and CBAM, it differs fundamentally in both design objective and information flow. SE-Net and CBAM generate channel attention weights based on the statistical information of a single feature branch, focusing on self-channel recalibration. In contrast, CMAF jointly models CNN-based features and Transformer-based features through a shared MLP to generate a unified channel attention matrix.

By leveraging cross-module feature interaction, the attention weights in CMAF dynamically reflect the relative contribution of different channels from heterogeneous representations, enabling adaptive fusion of local spatial features and global contextual features. Therefore, CMAF acts as a dynamic cross-module fusion mechanism rather than a conventional self-attention-based feature enhancement module.

2.2. Space–Spectral Feature Tensor Generation

SIFANet is for multi-modal tensor inputs, with the SFE and SST branches requiring spatial feature tensors and spectral feature tensors, respectively.

For the SFE branch, the input tensor is constructed from multiple image blocks, with each input tensor having the shape $\boldsymbol{X} \in \mathbb{R}^{B \times P \times P}$, where B is the batch size and P is the image block size. As shown in Figure 6, the specific operation for obtaining image blocks is as follows:

First, the original image is padded with a margin of $\left\lfloor \frac{(P-1)}{2} \right\rfloor$ pixels on all sides to ensure that edge pixels are adequately included. Then, a sliding window (also referred to as a mask) of size $P \times P$ is applied, starting from the first valid pixel in the top-left corner. The window moves across the entire image with a stride of 1 pixel, excluding padded and unclassified pixels. For each valid central pixel, a corresponding local neighborhood is extracted to form an image block. Each block is represented as $\boldsymbol{X_{Patch}} \in \mathbb{R}^{P \times P}$, and multiple such blocks are batched together to construct the spatial input tensor.

https://doi.org/10.3390/rs18030398

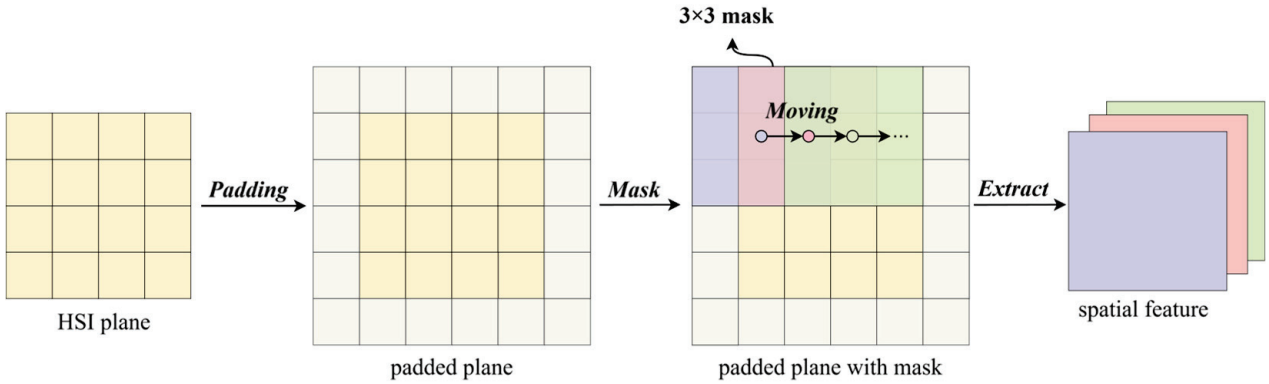


Figure 6. Spatial feature tensor generation.

For the SST branch, as shown in Figure 7, the multi-band spectral reflectance data of each pixel are directly extracted and assembled in batches to construct the spectral feature input tensor $\boldsymbol{X} \in \mathbb{R}^{B\times 1\times M}$, where $\boldsymbol{M}$ is the number of bands. It retains the spatial location of pixels while highlighting the discriminative features, avoiding spatial neighborhood interference, and enhancing the sensitivity to subtle spectral differences.

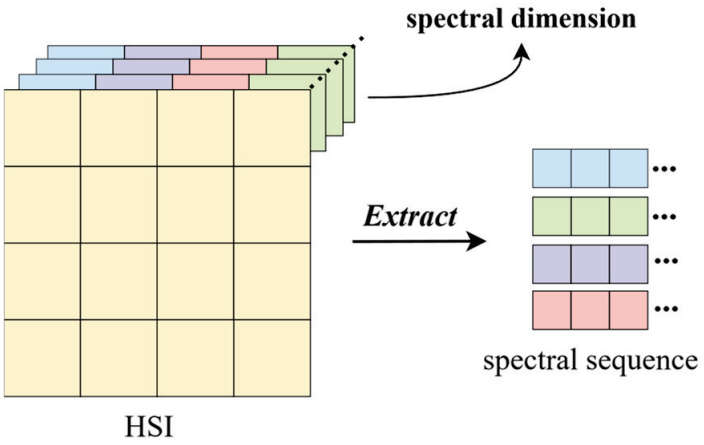


Figure 7. Spectral feature tensor generation.

2.3. Evaluation Index System

As illustrated in Figure 8, the proposed EIS incorporates three core modules. The first module utilizes classification accuracy indexes, including class accuracy (CA), overall accuracy (OA), the Kappa coefficient, and standard deviation (S.D.). These indices jointly evaluate the model's recognition capability of individual categories and its overall classified consistency and stability. The second module specifically addresses the performance evaluation of confusing classes by integrating the precision–recall curve (PRC) and F1-Score into a unified PR–F comprehensive evaluation model. It focuses on comparative assessment of a model's discriminatory capacity for challenging, spectrally similar categories. The third module evaluates computational efficiency through the novel CAPE index, which provides a quantitative assessment by systematically integrating key factors including the model's computational demands, parameter count, achieved classification accuracy, and measured stability performance.

https://doi.org/10.3390/rs18030398

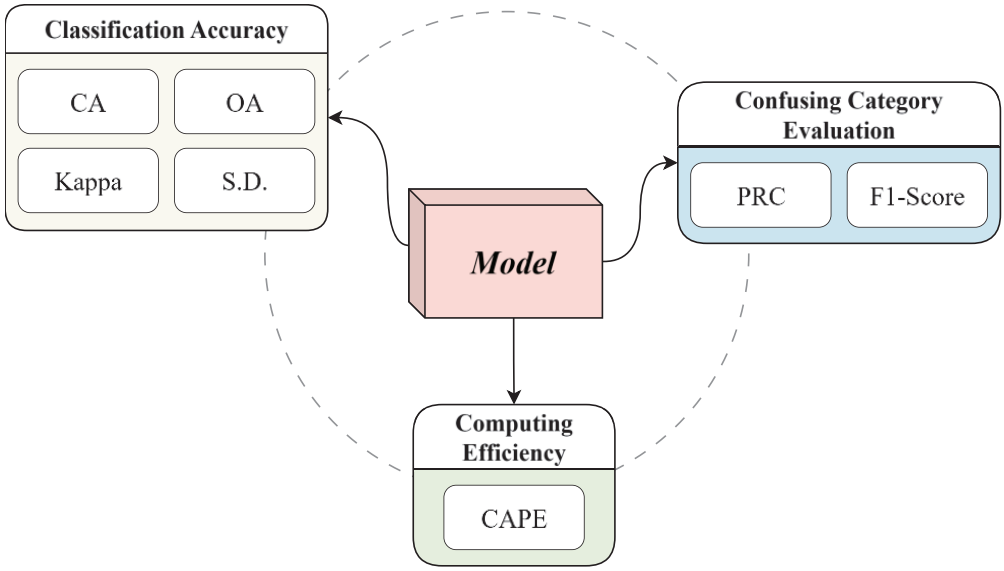


Figure 8. Evaluation index system.

2.3.1. Overall Accuracy

OA represents the proportion of correctly classified samples, calculated as

$$OA = \frac{T_c}{N} \tag{16}$$

where T_c denotes the number of correctly classified samples and N denotes the total number of samples.

2.3.2. Kappa Coefficient

The Kappa coefficient measures classification effectiveness while accounting for class imbalance. The calculation formula is as follows:

$$p_e = \frac{\sum_{i=1}^{m}(R_i \times C_i)}{N^2} \tag{17}$$

$$Kappa = \frac{OA - p_e}{1 - p_e} \tag{18}$$

where p_e is the chance agreement probability under random classification, m denotes the number of categories, R_i denotes the sum of row i in the confusion matrix, C_i denotes the sum of the column i in the confusion matrix, N denotes the total number of samples.

2.3.3. PR–F Comprehensive Evaluation Model

For the easily confusable ground object types, this article utilizes the PR–F comprehensive evaluation model, draws PRC, and statistically reports the F1-Score and AP values for auxiliary evaluation.

The F1-Score is a comprehensive metric used to evaluate a model's performance by balancing precision and recall. Precision refers to the proportion of correctly predicted positive samples among all samples predicted as positive. Recall measures the proportion of correctly predicted positive samples among all actual positive samples, indicating the model's ability to identify all relevant cases. These metrics are derived from the confusion matrix: true positives (TPs) represent the number of samples correctly predicted as positive; false positives (FPs) are the samples incorrectly predicted as positive; and false negatives

https://doi.org/10.3390/rs18030398

(FNs) are the actual positive samples that the model failed to identify. Precision, recall, and F1-Score are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{21}$$

The precision–recall curve (PRC) is used to evaluate the performance of a model by illustrating how precision and recall change at different classification thresholds. The closer the model's PRC is to the upper right of the coordinate system, the better the model's performance. To quantify overall model performance, the average precision (AP) is calculated as the area under the PRC. The closer the AP value is to 1, the better the model balances precision and recall.

2.3.4. Computational Accuracy Parameter Efficiency

Traditional single metrics such as OA focus solely on performance, overlooking computational burden and offering limited insight into model applicability [51]. To address this limitation, the CAPE index integrates multiple factors: the number of parameters, the number of floating-point operations (FLOPs), the S.D., and OA. It provides a comprehensive quantitative evaluation of a model's computational efficiency, classification accuracy, and prediction stability. A higher CAPE value indicates superior model performance—simultaneously reflecting higher accuracy, greater stability, optimized efficiency, and balanced computational design.

To unify measurement scales and introduce a penalization mechanism, the number of parameters (P) and FLOPs (F) are logarithmically transformed using $y = log_{10}x$. Penalization is implemented via the reciprocal:

$$N_P = \frac{1}{log_{10}(1 + P)} \tag{22}$$

$$N_F = \frac{1}{log_{10}(1 + F)} \tag{23}$$

here, N_P and N_F represent the normalized values of parameters and FLOPs. When either P or F is large, their corresponding normalized values decrease significantly, penalizing models with excessive computational costs.

The intermediate metric E, representing computational burden, is then calculated as the geometric mean of N_P and N_F. A lower E score indicates higher computational resource consumption:

$$E = \sqrt{N_P \times N_F} \tag{24}$$

The final CAPE score is computed by E, OA, and the inverse of the S.D. ($\frac{1}{\sigma}$):

$$CAPE = E \times OA \times \frac{1}{\sigma} \tag{25}$$

This multiplicative formulation highlights performance differences among models and balances high accuracy with prediction stability. Notably, even models with high OA could receive a low CAPE score if they have a high S.D.

https://doi.org/10.3390/rs18030398

3. Experimental Design and Results

3.1. Dataset

3.1.1. Salinas Dataset

The Salinas dataset (denoted as SL), acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) (developed by NASA's Jet Propulsion Laboratory, Pasadena, CA, USA) covers a portion of the Salinas Valley in California, USA. The image is 217 pixels × 512 pixels, with a spatial resolution of 3.7 m. It originally comprised 224 spectral bands. After removing bands affected by water absorption, 204 spectral bands remain, spanning the 400–1000 nm wavelength range. This dataset primarily includes 16 categories of crops. The corresponding ground truth map is shown in Figure 9.

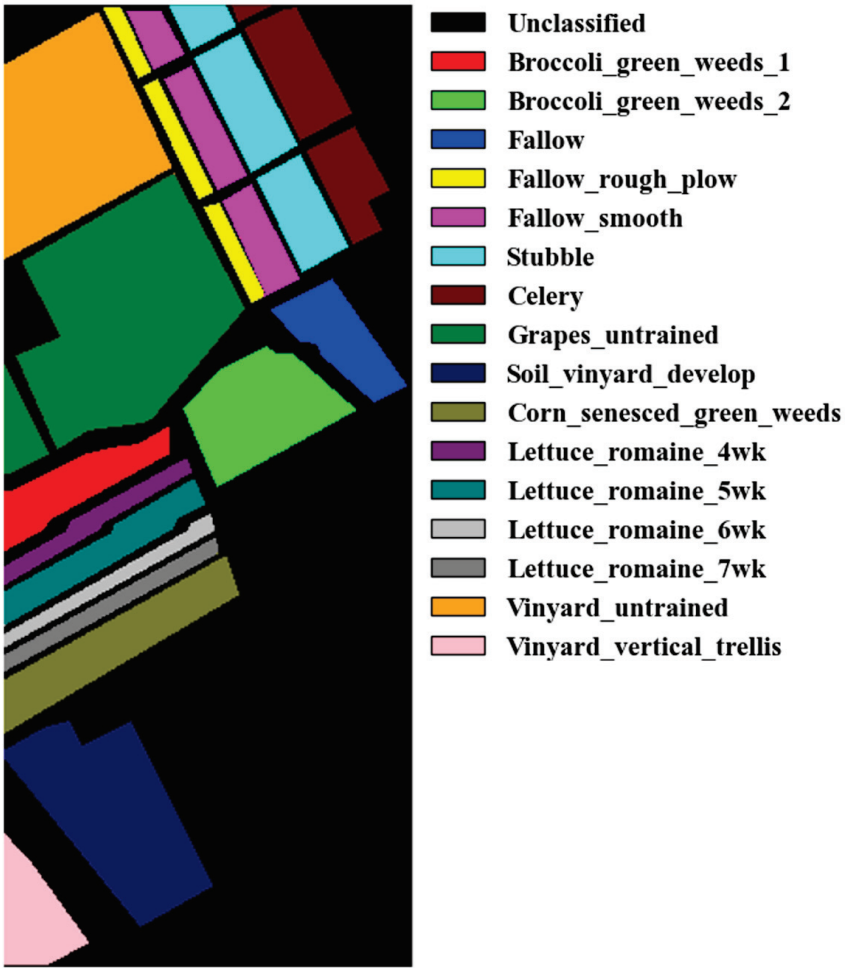


Figure 9. Salinas dataset ground truth.

3.1.2. Xiong'an Hyperspectral Dataset

The Xiong'an dataset (denoted as XA) was acquired using the Full-spectrum Multimodal Imaging Spectrometer (manufactured by Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China) [52]. It covers an area of approximately 1320 km^2, including Xiong County, Anxin County, Rongcheng County, and the main water area of Baiyangdian Lake. This article uses a subset of Matiwan Village, sized 3750 × 1580 pixels at 0.5 m spatial resolution, with 256 spectral bands (400–1000 nm). It includes 20 land cover classes, mainly cash crops (Figure 10).

https://doi.org/10.3390/rs18030398

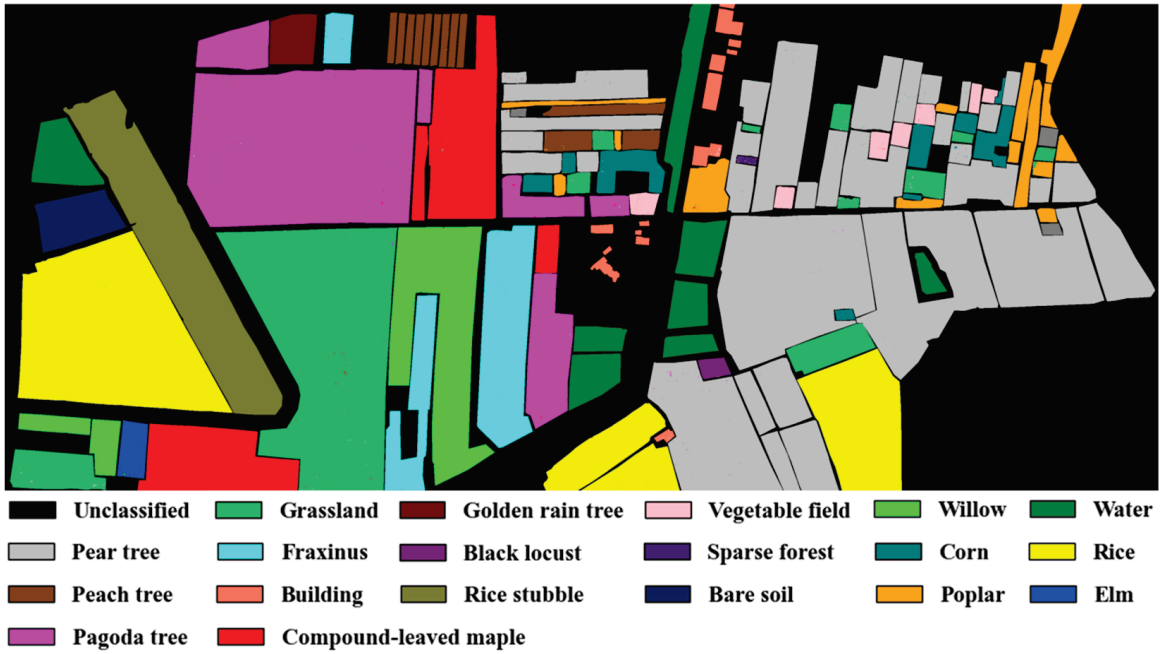


Figure 10. Xiong'an hyperspectral dataset ground truth.

The key characteristics of the two datasets are summarized in Table 1 for a clear comparison.

Table 1. Key characteristics of the two datasets.

Dataset	Image Size (pixel)	Spatial Resolution (m)	Wavelength Range (nm)	Bands	Classes
Salinas	217 × 512	3.7	400–1000	204	16
Xiong'an	3750 × 1580	0.5	400–1000	256	20

The Salinas and Xiong'an datasets were selected for their complementary traits: the Salinas dataset, being small-scale with limited samples, tests small-sample learning and performance under scarcity. In contrast, the large-scale Xiong'an dataset assesses generalization ability and computational efficiency. The Salinas dataset, as a benchmark, enables direct comparison with prior works, while Xiong'an, with complex land cover and higher resolution, reflects real-world diversity and tests model adaptability.

To evaluate the performance and robustness of SIFANet under various data constraints, we first conducted a sensitivity analysis by varying the training sample proportions from 1% to 20%. The results, summarized in Table 2, reveal that SIFANet achieves competitive accuracy even with extremely limited training data (e.g., 1% or 5%). As the proportion increases to 20%, the classification performance tends to saturate, indicating that the model has effectively captured the discriminative spectral–spatial features.

Table 2. OA of SIFANet under different training sample proportions. (Unit: %).

Dataset	1%	5%	10%	15%	20%
Salinas	91.02	97.45	98.92	99.34	99.89
Xiong'an	85.30	94.18	97.05	98.42	99.79

The rationale for selecting the 20% ratio as our baseline is twofold. First, high-spectral data typically exhibit strong spatial correlation; using higher ratios (e.g., 30% or 40%) may lead to overfitting on local spatial patterns rather than learning generalized features [53].

Second, the 20% ratio strikes a balance between ensuring deep learning model convergence and simulating real-world scenarios where labeled samples are scarce and expensive to obtain. Based on this analysis, the detailed partitioning of training and testing samples for each dataset is provided in Table 3.

Table 3. Sample sizes for training and testing in the dataset.

	Salinas Dataset			Xiong'an Hyperspectral Dataset		
NO.	**Class Name**	**Training**	**Test**	**Class Name**	**Training**	**Test**
1	Broccoli_green_weeds_1	399	1610	Compound-leaved maple	45,153	180,764
2	Broccoli_green_weeds_2	784	2942	Willow	35,829	144,124
3	Fallow	381	1595	Elm	3067	12,222
4	Fallow_rough_plow	284	1110	Rice	90,871	361,574
5	Fallow_smooth	535	2143	Pagoda tree	94,755	380,562
6	Stubble	798	3161	Fraxinus	34,056	135,125
7	Celery	730	2849	Golden rain tree	4629	18,695
8	Grapes_untrained	2222	9049	Water	33,125	132,610
9	Soil_vinyard_develop	1226	4977	Bare soil	7800	30,611
10	Corn_senesced_green_weeds	626	2652	Rice stubble	38,763	155,348
11	Lettuce_romaine_4wk	229	839	Black locust	1093	4471
12	Lettuce_romaine_5wk	390	1537	Corn	11,675	47,331
13	Lettuce_romaine_6wk	174	742	Pear tree	205,019	820,983
14	Lettuce_romaine_7wk	230	840	Soybean	1437	5711
15	Vinyard_untrained	1463	5805	Poplar	18,290	73,019
16	Vinyard_vertical_trellis	354	1453	Vegetable field	5875	23,138
17				Sparse forest	297	1218
18				Grassland	84,691	337,850
19				Peach tree	13,070	52,552
20				Building	5926	23,781
Total		10,825	43,304		735,421	2,941,689

3.2. Experimental Design

This article employs a controlled variable experimental design, varying only the model type to reduce systematic errors and to enhance comparability. All experiments were implemented using the PyTorch (version 2.6.0) framework. For clarity, the specific structural hyperparameters of SIFANet—including kernel sizes, embedding dimensions, and fusion settings—are listed in Table 4. Specifically, the training process consisted of 200 epochs with a batch size of 256 and an initial learning rate of 0.0001. All experimental results reported are an average of 50 independent experiments.

Table 4. Detailed configuration of SIFANet architecture.

Module	**Parameter/Component**	**Value/Configuration**
SFE Branch	Structure	Conv + MaxPool + Residual Block
	Conv Kernel Size	3×3
	Max Pooling	Window 2×2, Stride 2
	Residual Block	2 Layers (3×3 Conv + BN + ReLU)
SST Branch	Structure	Conv-Former
	Encoder Layers	3
	Attention Heads	4
	Feed-Forward Network	CFFN (1×1 Conv)
	Embedding Dimension	64

https://doi.org/10.3390/rs18030398

Table 4. *Cont.*

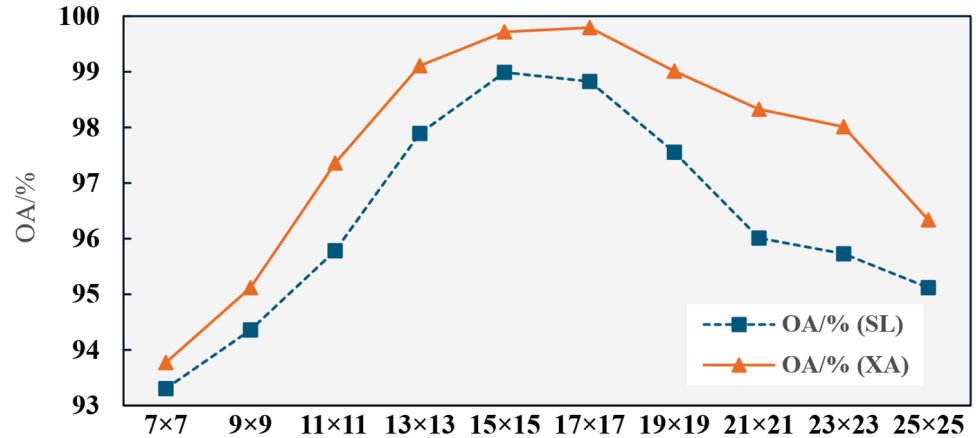
Module	Parameter/Component	Value/Configuration
CMAF	Interaction Mechanism	Shared MLP
	Activation Function	Sigmoid
	Hidden Layer Units	32

The EIS proposed in this article is used to assess and compare the classification performance of SIFANet with that of SVM, 3D-CNN, and LSTM. Here, SVM is a typical traditional machine learning model, while 3D-CNN and LSTM are deep learning models. Specifically, 3D-CNN can capture local spatial features through convolution. LSTM excels at processing sequence data. By comparing these diverse benchmarks, this article validates the effectiveness of SIFANet for HSI classification tasks.

3.3. Key Parameter Setting Experiments

3.3.1. Image Block Size

Due to constraints in image width and computing resources, the original HSI is partitioned into multiple blocks of identical size, each containing an odd number of pixels. Since block size significantly affects model performance [54], the optimal size was determined experimentally. Holding other parameters constant, this study tested block sizes from 7×7 to 25×25 using a 3D-CNN on both the Salinas and Xiong'an datasets. OA was used for evaluation, with results averaging over 10 runs, as shown on Figure 11.



	7×7	9×9	11×11	13×13	15×15	17×17	19×19	21×21	23×23	25×25
OA/% (SL)	93.3	94.36	95.78	97.89	98.99	98.83	97.55	96.01	95.73	95.12
OA/% (XA)	93.77	95.12	97.36	99.11	99.72	99.79	99.01	98.32	98.01	96.34

Figure 11. The impact of selected image block size on OA.

Block size nonlinearly influences model performance. Both datasets show a consistent trend: OA first rises with increasing block size, peaks, and then declines. Specifically, smaller blocks (e.g., 7×7, 9×9) lead to insufficient spatial feature extraction due to limited receptive fields, yielding lower OA. At 15×15, OA reaches 98.99% (SL) and 99.72% (XA)—gains of 6.1% and 6.35% over 7×7. Although XA peaks at 99.79% with 17×17, the 0.07% difference from 15×15 is statistically insignificant. Beyond 17×17, performance drops due to boundary effects and local feature dilution, likely caused by increased background noise, reduced focus on discriminative features, and edge information loss.

Considering both marginal gains and computational cost, 15×15 was selected as optimal. It achieves peak accuracy on the Salinas dataset and near-peak on Xiong'an dataset, with high computational efficiency, and served as the input size for all subsequent experiments.

3.3.2. Optimizer and Loss Function

The loss function quantifies the error between the predictions and ground truth, guiding parameter updates. The optimizer's role is to iteratively adjust parameters to minimize the loss, thereby enhancing predictive capability.

Common optimizers include stochastic gradient descent (SGD) and adaptive moment estimation (Adam). SGD uses random data subsets for gradient updates, offering higher efficiency but potentially causing oscillations or slow convergence. Adam estimates gradient moments to combine momentum and adaptive learning rates, enabling faster and more stable convergence. For loss functions, squared hinge loss (SHL) strengthens misclassification penalties via a squared term to maximize margins, while categorical cross-entropy (CCE) minimizes the Kullback–Leibler divergence between predicted and true distributions, aligning the model with inherent class distributions.

This article experimentally evaluated combinations of these optimizers and loss functions, with results averaged over 10 runs (Figure 12).

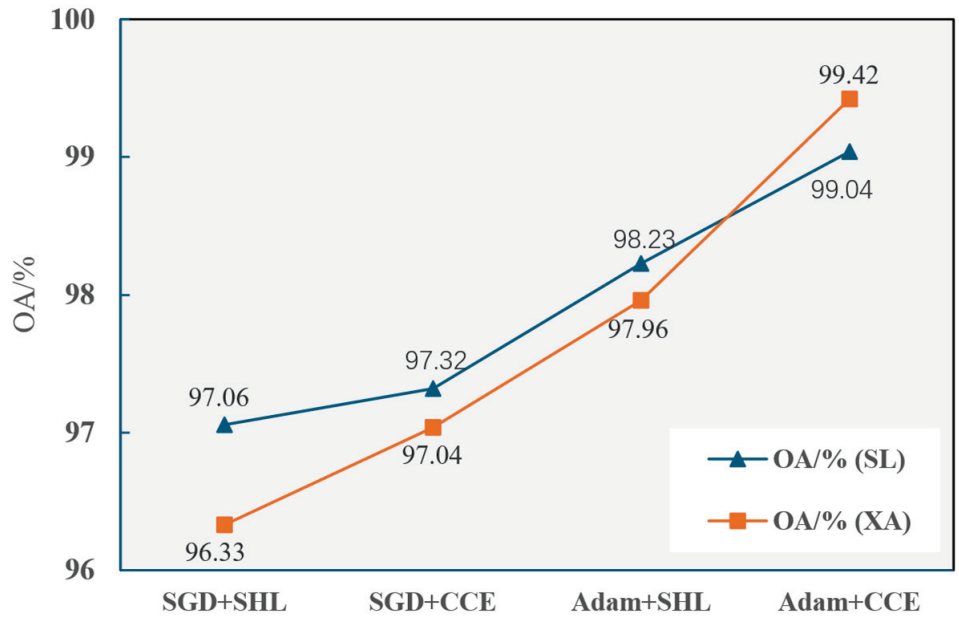


Figure 12. The impact of optimizers and loss functions on OA.

"Adam + CCE" achieved the best OA on both Salinas (99.04%) and Xiong'an (99.42%). For a fixed optimizer, CCE consistently outperformed SHL; for a fixed loss, Adam surpassed SGD. Thus, "Adam + CCE" was selected for subsequent experiments.

3.4. Experiments Results

Based on the experimental setup discussed above, this article compares the classification performance of SVM, 3D-CNN, LSTM, and SIFANet on the Salinas and the Xiong'an hyperspectral datasets. The results are shown in Tables 5 and 6. These tables present the classification accuracy in the format of "accuracy ± difference from SIFANet". Different colors are used to distinguish comparison results: orange indicates accuracy lower than SIFANet, blue indicates accuracy better than SIFANet, and black indicates parity. Red highlights easily confused categories. Box plots of classification accuracy are also provided for enhanced visualization (Figure 13).

https://doi.org/10.3390/rs18030398

Table 5. Results of Salinas dataset. (Unit: %).

No.	SVM		3D-CNN		LSTM		SIFANet
1	96.56	−3.44	100.00	0.00	100.00	0.00	100.00
2	91.74	−8.16	100.00	0.10	99.50	−0.40	99.90
3	88.84	−10.97	99.07	−0.74	97.12	−2.69	99.81
4	98.18	−0.56	98.93	0.19	99.04	0.30	98.74
5	83.31	−16.60	98.82	−1.09	99.02	−0.89	99.91
6	99.97	−0.03	99.87	−0.13	100.00	0.00	100.00
7	92.92	−6.94	100.00	0.14	99.93	0.07	99.86
8	64.16	−35.83	99.04	−0.95	80.80	−19.19	99.99
9	86.58	−13.42	99.88	−0.12	99.13	−0.87	100.00
10	69.74	−30.18	99.62	−0.30	96.39	−3.53	99.92
11	0.00	−99.41	100.00	0.59	98.25	−1.16	99.41
12	66.84	−32.70	98.93	−0.61	97.25	−2.29	99.54
13	91.30	−8.56	100.00	0.14	98.30	−1.56	99.86
14	94.24	−5.41	97.89	−1.76	96.40	−3.25	99.65
15	58.87	−41.13	96.35	−3.65	78.02	−21.98	100.00
16	99.01	−0.85	100.00	0.14	99.86	0.00	99.86
OA	79.18	−20.71	99.05	−0.84	92.36	−7.53	99.89
Kappa	76.55	−23.33	98.95	−0.93	91.48	−8.40	99.88
S.D.	25.18	−24.85	1.00	−0.67	6.68	−6.35	0.33

Table 6. Results of Xiong'an hyperspectral dataset. (Unit: %).

No.	SVM		3D-CNN		LSTM		SIFANet
1	33.20	−66.52	99.53	−0.19	78.48	−21.24	99.72
2	51.81	−48.10	99.84	−0.07	77.48	−22.43	99.91
3	0.00	−99.92	99.61	−0.31	78.27	−21.65	99.92
4	93.80	−6.19	99.97	−0.02	98.11	−1.88	99.99
5	40.49	−59.40	99.31	−0.58	72.22	−27.67	99.89
6	52.00	−47.94	99.72	−0.22	77.28	−22.66	99.94
7	85.99	−13.99	99.87	−0.11	88.03	−11.95	99.98
8	89.17	−10.82	99.93	−0.06	97.06	−2.93	99.99
9	75.15	−24.84	100.00	0.01	98.61	−1.38	99.99
10	86.08	−13.92	100.00	0.00	98.34	−1.66	100.00
11	0.00	−99.62	97.05	−2.57	54.95	−44.67	99.62
12	1.19	−98.39	98.46	−1.12	69.14	−30.44	99.58
13	53.59	−46.07	99.05	−0.61	74.35	−25.31	99.66
14	0.00	−99.25	99.37	0.12	73.71	−25.54	99.25
15	58.06	−41.50	99.03	−0.53	72.60	−26.96	99.56
16	82.67	−14.93	94.82	−2.78	70.46	−27.14	97.60
17	0.00	−99.32	93.33	−5.99	25.00	−74.32	99.32
18	45.36	−54.45	99.54	−0.27	74.37	−25.44	99.81
19	0.00	−99.78	99.46	−0.32	74.68	−25.10	99.78
20	97.84	−1.60	99.60	0.16	95.87	−3.57	99.44
OA	59.11	40.68	99.42	−0.37	80.19	−19.60	99.79
Kappa	50.80	−48.96	99.33	−0.43	76.83	−22.93	99.76
S.D.	37.09	−36.51	1.95	−1.37	18.54	−17.96	0.58

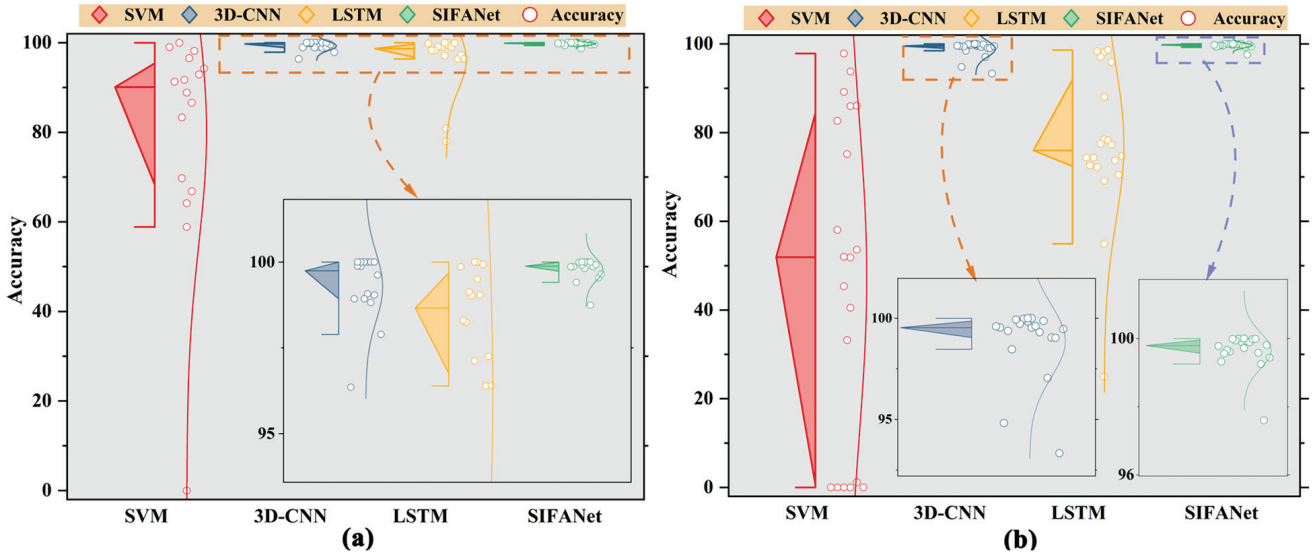


Figure 13. Box plots of classification accuracy. (**a**) Salinas dataset. (**b**) Xiong'an hyperspectral dataset.

SIFANet demonstrates superior performance with the OA of 99.89% (SL)/99.79% (XA), outperforming the second-best 3D-CNN model by 0.84% and 0.37%, respectively. The Kappa coefficient of the proposed SIFANet reaches 99.88% (SL)/99.76% (XA), significantly exceeding other models. The standard deviation remains low at 0.33% (SL)/0.58% (XA), indicated strong adaptability and robustness across land cover categories, whereas other models exhibit substantial fluctuations.

Figure 13 shows pronounced performance differences among the models. Ranked by median accuracy, the order is consistently SIFANet > 3D-CNN > LSTM > SVM. SVM exhibits high inter-category variability, with interquartile ranges of 40% (SL) and 100% (XA), and numerous outliers. LSTM outperforms SVM but still shows limitations (e.g., 45% interquartile range on XA). The 3D-CNN maintains relatively compact quartile ranges and high accuracy but has minor outliers. In contrast, SIFANet shows minimal variance, with narrow quartile ranges and peaked distributions, indicating exceptional stability.

Across all land cover categories, SIFANet consistently maintains over 97% accuracy. The SVM model, constrained by linear classification mechanisms, drops to 0% in several categories with 25.18% (SL)/37.09% (XA) standard deviations. Additionally, both LSTM and 3D-CNN fail to fully capture the spatial–spectral information, causing significant decline in specific categories' accuracy such as classes 8, 14, and 15 (SL) and 11, 16, and 17 (XA), where performance substantially lags behind the proposed SIFANet. Focusing on poorly classified categories, as for class 8 (Grapes_untrained) in Salinas, SVM's accuracy is merely 64.16%, LSTM reaches only 80.80%, while SIFANet achieves 99.99%. Similarly for class 17 of Xiong'an hyperspectral dataset (Sparse Forest), LSTM accuracy drops to 25.00% and 3D-CNN to 93.33%, whereas SIFANet maintains 99.32% accuracy.

In summary, SIFANet not only sustains high accuracy across all categories but also exhibits minimal fluctuations and stable performance, conclusively demonstrating exceptional hyperspectral feature extraction and discrimination capabilities.

4. Discussion

4.1. Classification Performance and Accuracy Consistency Verification

Some models, although performing well in quantitative metrics such as OA, still have significant discrepancies between these metrics and the actual classification results. This article analyses the classification result maps of all models and selected typical regions with multi-category boundaries for locally enlarged comparisons, which are shown in Figures 14 and 15.

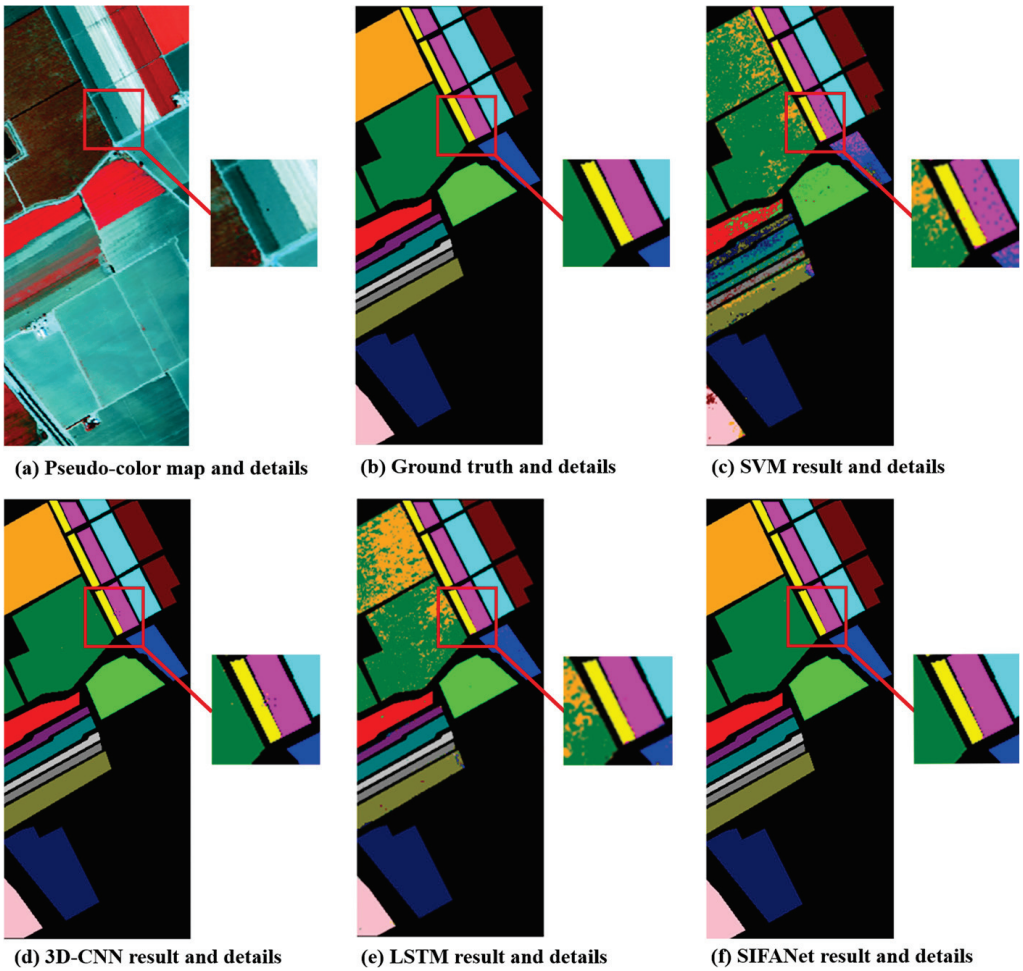


Figure 14. Classification results and details of each model on the Salinas dataset.

SVM and LSTM, lacking spatial contextual modeling capabilities, display diffusely distributed misclassification. Severe boundary penetration occurs between spectrally similar categories like "Grapes_untrained" and "Vinyard_untrained", accompanied by noticeable intra-class noise and boundary blurring.

Although 3D-CNN achieves high accuracy (OA > 99.05%), salt-and-pepper noise appears in the detailed views of both datasets, manifesting as single-pixel-scale categorical mutations. This primarily stems from excessive smoothing of local details in deep networks—an irreversible spatial resolution loss during hierarchical abstraction.

In contrast, SIFANet maintains continuous, clear category boundaries and smooth homogeneous regions in the detailed views. The results validate the effectiveness of the model's structure design. The spatial–spectral dual-branch could extract complementary features, where residual blocks in the SFE branch effectively alleviate vanishing gradients in deep networks. Concurrently, the Conv-Former module enhances spectral sequence modeling capability. Ultimately, the CMAF module achieves the dynamic adaptive fusion of dual-branch features, minimizing information loss during feature transmission.

https://doi.org/10.3390/rs18030398



Figure 15. Classification results and details of each model on the Xiong'an hyperspectral dataset.

4.2. Confusing Classes Classification Performance Comparison

Based on the experimental results in Tables 5 and 6, this article identifies land cover categories where multiple models perform poorly (i.e., confusing categories) and assesses them using the PR–F comprehensive evaluation model. The experimental results are shown in Figures 16 and 17.

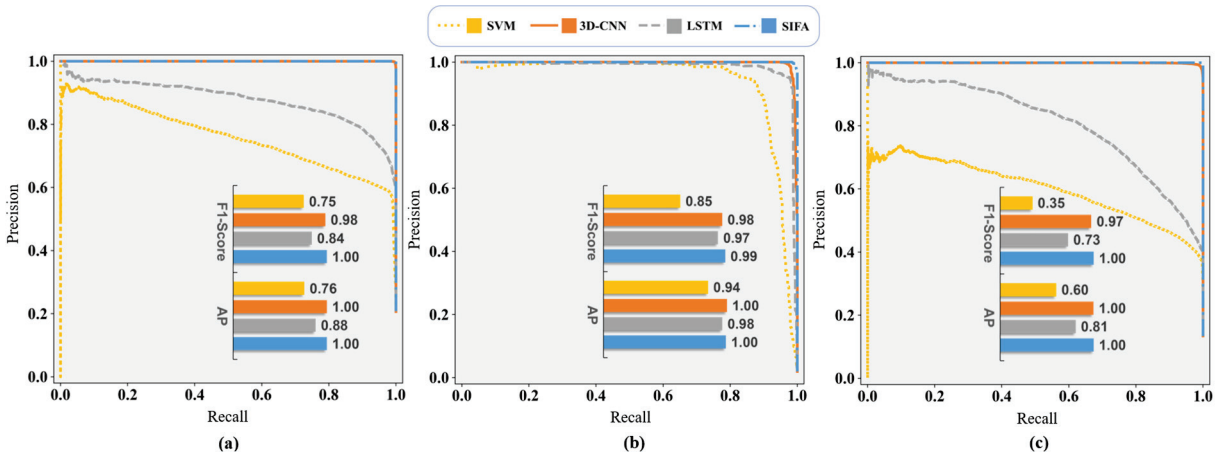


Figure 16. PR–F evaluation results on Salinas dataset. (**a**) Class 8 (Grapes_untrained) performance. (**b**) Class 14 (Lettuce_ romaine_7wk) performance. (**c**) Class 15 (Vinyard_untrained) performance.

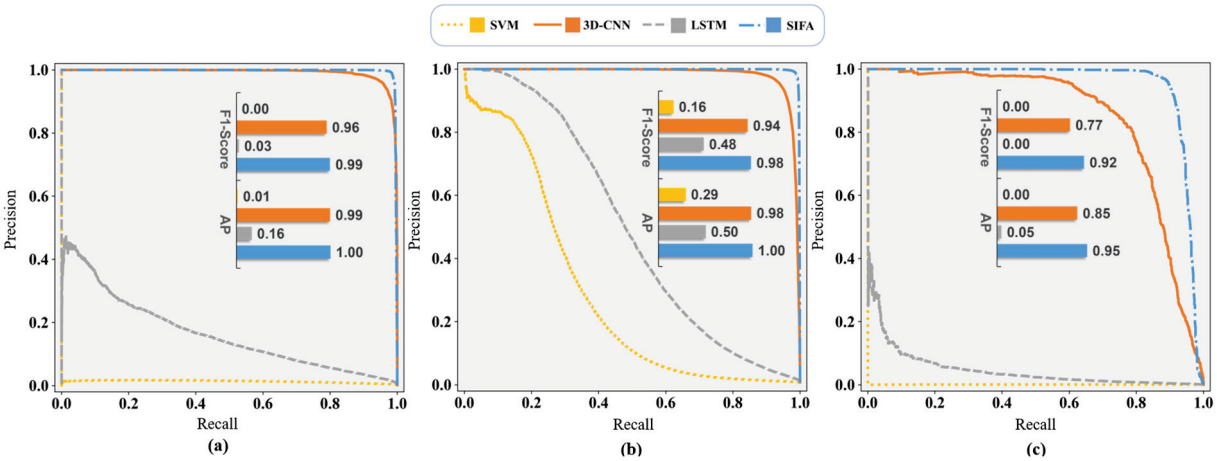


Figure 17. PR–F evaluation results on Xiong'an hyperspectral dataset. (**a**) Class 11 (Black locust) performance. (**b**) Class 16 (Vegetable field) performance. (**c**) Class 17 (Sparse Forest) performance.

The PRCs reveal significant fluctuations in SVM's accuracy at varying recall thresholds, characterized by pronounced sawtooth patterns. On the large-scale Xiong'an hyperspectral dataset, LSTM exhibits similar abrupt curve variations. Conversely, 3D-CNN and SIFANet display nearly overlapping trajectories on the Salinas dataset. However, as data complexity increases in Xiong'an, SIFANet maintains superior accuracy through its spatial–spectral fusion mechanism, while 3D-CNN shows significant deviation and accuracy degradation—confirming SIFANet's efficacy in high-difficulty tasks.

F1-Score and AP values validate PRC observations. SIFANet consistently achieves optimal performance in imbalanced datasets, demonstrating robust results: F1-Score > 0.99 and saturated AP values across all Salinas categories, while maintaining > 0.92 F1-Score and >0.95 AP in Xiong'an dataset. Though 3D-CNN performs second best in most scenes, its F1-Score and AP decrease markedly (20.3% and 14.5%, respectively) when processing challenging features like "Sparse Forest" compared to other confusing categories. SVM and LSTM perform notably worse in the Xiong'an dataset with the F1-Score and AP dropping below 0.5 and even reaching zero for categories like "Black locust" and "Sparse Forest".

It is worth noting that the Salinas dataset suffers from class imbalance, where several land cover categories contain only a limited number of training samples. Such imbalance often leads to classification bias toward majority classes, particularly for models that rely on single-branch feature representations. The proposed CMAF module helps mitigate this issue by performing dynamic cross-module feature fusion between the spatial-oriented SFE branch and the spectral sequence modeling SST branch. By jointly learning channel-wise attention weights from heterogeneous features, CMAF adaptively enhances discriminative cues that are critical for minority classes, which are often characterized by subtle spectral differences or limited spatial support. This dynamic fusion mechanism contributes to more stable precision–recall behavior and the consistently high F1-Score and AP values across confusing and low-sample categories in the Salinas dataset, effectively reducing the model's bias toward majority classes.

4.3. Evaluation of Model Calculation Efficiency

Pursuing higher accuracy and stability under the same hardware conditions has always been an inevitable trend in model development. Table 7 summarizes the number of parameters and FLOPs of each model. Inputting the model efficiency parameter data from Table 7 into the CAPE value calculation formula yields the results shown in Figure 18.

Table 7. Efficiency parameters of each model.

Index		SVM	3D-CNN	LSTM	SIFANet
Parameters		160	58,992	56,080	269,873
FLOPs ($\times 10^3$)		0.12	2005.248	54.016	7979.392
OA	SL (%)	79.18	99.05	92.36	99.89
	XA (%)	59.11	99.42	80.19	99.79
S.D.	SL (%)	25.18	1.00	6.68	0.33
	XA (%)	37.09	1.95	18.54	0.58

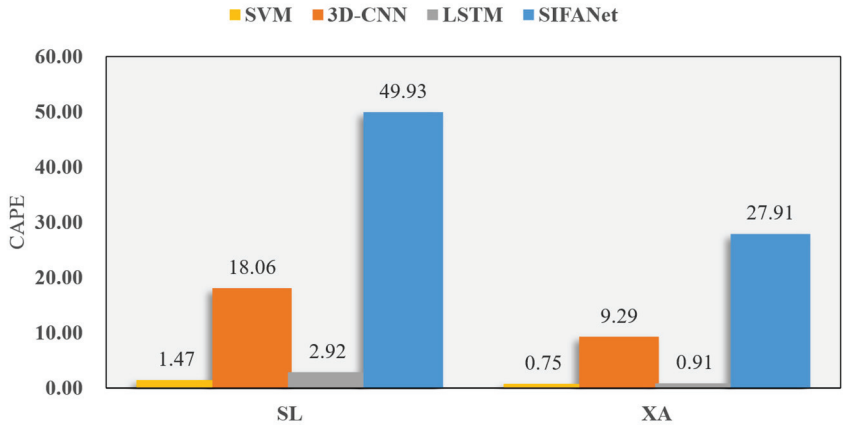


Figure 18. Comparison of CAPE values of each model.

SIFANet's CAPE values of 49.93 (SL) and 27.91 (XA) are significantly higher than those of 3D-CNN (SL: 18.06/XA: 9.29), LSTM (SL: 2.91/XA: 0.91), and SVM (SL: 1.47/XA: 0.75). The results indicate that SIFANet excels in comprehensive performance across multiple dimensions, validating its adaptability advantage in complex hyperspectral data scenarios. Furthermore, the CAPE index provides a pragmatic benchmark for model selection in real-world deployment. In resource-constrained environments such as drones or edge devices, researchers can leverage CAPE to quantify the "accuracy gain per unit of computational cost". This facilitates the identification of models that strike an optimal balance between inference speed and diagnostic precision.

4.4. Ablation Experiment

It is necessary to conduct ablation experiments to validate the effectiveness of the three core modules—SFE, SST, and CMAF. The results are shown in Table 8. Different colors in the table are used to distinguish accuracy comparison results: orange-marked values indicate relative accuracy lower than the SIFANet benchmark. "√" indicates that this structure is enabled in the experiment, while "×" indicates that it is disabled. Experiment 1 served as the baseline control group, incorporating all SFE, SST, and CMAF modules. The experimental design followed a progressive decoupling approach to analyze module interactions.

Validation of SFE/SST independent effectiveness used Experiment 1 (SFE + SST + CMAF) as the baseline. Performance comparisons with Experiment 2 (SST + CMAF only) and Experiment 3 (SFE + CMAF only) confirmed that adding either SFE or SST significantly enhanced classification accuracy. Notably, the proposed SFE module outperformed the conventional 3D-CNN due to its multi-scale convolution kernels and residual connections that effectively mitigate deep feature degradation.

Comparing Experiments 2/3 ("single module + CMAF") against Experiments 5/6 ("single module only") shows higher OA and Kappa coefficients for the former, demonstrating CMAF's ability to extract valuable information from SFE/SST branches while reducing

interference from ineffective features. Contrasting Experiment 1 (CMAF) with Experiment 4 (simply Add fusion) confirms that it could achieve complementary spatial–spectral information, improving OA by 4.77% (SL)/4.88% (XA) and Kappa by 5.29% (SL)/5.72% (XA). This superiority stems from CMAF's attention-driven gating mechanism. Unlike simple element-wise addition, which treats all features with equal importance and may lead to "feature dilution" or noise propagation, the attention-based strategy dynamically reweights the feature maps. It prioritizes discriminative spatial–spectral components while effectively suppressing redundant or non-informative features from individual branches. This adaptive fusion ensures a more precise alignment and integration of heterogeneous information.

Table 8. Ablation experiment results.

No.	SFE	SST	CMAF	SL				XA			
				OA		Kappa		OA		Kappa	
1	√	√	√	99.89		99.88		99.79		99.76	
2	×	√	√	94.33	−5.56	93.12	−6.76	90.61	−9.18	88.23	−11.53
3	√	×	√	99.26	−0.63	99.33	−0.55	99.54	−0.25	99.41	−0.35
4	√	√	×	95.12	−4.77	94.59	−5.29	94.91	−4.88	94.04	−5.72
5	√	×	×	99.15	−0.74	99.05	−0.83	99.51	−0.28	99.41	−0.35
6	×	√	×	91.74	−8.15	90.01	−9.87	89.84	−9.95	88.26	−11.5

In summary, each SIFANet module demonstrates scientifically sound design with significant synergistic effects and collective enhancing classification accuracy.

5. Conclusions

This article proposes a dual-branch collaborative network for HSI classification. The SIFANet is based on spectral integration and focused attention mechanisms. Specifically, the SFE and SST modules extract spatial and spectral features parallelly and efficiently. Subsequently, the proposed CMAF module dynamically fuses the feature information from both branches to enhance image detail representation and class discriminability. The proposed network significantly reduces feature loss and effectively relieves the insufficient feature extraction capability, which could learn the correlation of ground objects from multi-dimensional image feature spaces. As a result, SIFANet significantly outperforms the comparative models in accuracy assessment metrics such as CA, S.D., and CAPE.

However, the dual-branch architecture and attention mechanisms increase the computational complexity, which may limit its efficiency in real-time applications. Future work will focus on developing lightweight modules to reduce computational costs and exploring self-supervised learning to improve the model's robustness in scenarios with limited labeled samples.

Author Contributions: Conceptualization, Y.G. and L.X.; methodology, Y.G. and L.X.; validation, Y.G.; writing—original draft preparation, Y.G.; writing—review and editing, L.X., D.M., Y.W., and M.H.; visualization, Y.G.; funding acquisition, L.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was jointly supported by the Beijing Natural Science Foundation under grant L251045, the Deep Earth Probe and Mineral Resources Exploration—National Science and Technology Major Project under grant 2024ZD1002100, and the Fundamental Research Funds in China University of Geoscience (Beijing) under grant 590125018.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

1. Rast, M.; Painter, T.H. Earth Observation Imaging Spectroscopy for Terrestrial Systems: An Overview of Its History, Techniques, and Applications of Its Missions. *Surv. Geophys.* **2019**, *40*, 303–331. [CrossRef]
2. Liu, T.; Gu, Y.; Jia, X. Class-guided coupled dictionary learning for multispectral-hyperspectral remote sensing image collaborative classification. *Sci. China Technol. Sci.* **2022**, *65*, 744–758. [CrossRef]
3. Yang, Y.; Zhou, Y.; Chen, J.-N. Hyperspectral image classification based on multi-scale hybrid convolutional network. *Chin. J. Liq. Cryst. Disp.* **2023**, *38*, 368–377. [CrossRef]
4. Sun, Y.; Liu, B.; Wang, R.; Zhang, P.; Dai, M. Spectral–Spatial MLP-like Network with Reciprocal Points Learning for Open-Set Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5513218. [CrossRef]
5. Zhang, B.; Chen, Y.; Li, Z.; Xiong, S.; Lu, X. SANet: A Self-Attention Network for Agricultural Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5501315. [CrossRef]
6. Zhang, H.; Feng, S.; Wu, D.; Zhao, C.; Liu, X.; Zhou, Y.; Wang, S.; Deng, H.; Zheng, S. Hyperspectral Image Classification on Large-Scale Agricultural Crops: The Heilongjiang Benchmark Dataset, Validation Procedure, and Baseline Results. *Remote Sens.* **2024**, *16*, 478. [CrossRef]
7. Yang, B.; Wang, S.; Li, S.; Zhou, B.; Zhao, F.; Ali, F.; He, H. Research and application of UAV-based hyperspectral remote sensing for smart city construction. *Cogn. Robot.* **2022**, *2*, 255–266. [CrossRef]
8. Duan, P.; Xie, Z.; Kang, X.; Li, S. Self-supervised learning-based oil spill detection of hyperspectral images. *Sci. China Technol. Sci.* **2022**, *65*, 793–801. [CrossRef]
9. Leung, J.-H.; Tsao, Y.-M.; Karmakar, R.; Mukundan, A.; Lu, S.-C.; Huang, S.-Y.; Saenprasarn, P.; Lo, C.-H.; Wang, H.-C. Water pollution classification and detection by hyperspectral imaging. *Opt. Express* **2024**, *32*, 23956–23965. [CrossRef]
10. Xu, L.; Chen, Y.; Feng, A.; Shi, X.; Feng, Y.; Yang, Y.; Wang, Y.; Wu, Z.; Zou, Z.; Ma, W.; et al. Study on detection method of microplastics in farmland soil based on hyperspectral imaging technology. *Environ. Res.* **2023**, *232*, 116389. [CrossRef]
11. Arya, R.K.; Peddi, R.; Srivastava, R. Hyperspectral image classification using hybrid convolutional-based cross-patch retentive network. *Comput. Vis. Image Underst.* **2025**, *257*, 104382. [CrossRef]
12. Hong, D.; He, W.; Yokoya, N.; Yao, J.; Gao, L.; Zhang, L.; Chanussot, J.; Zhu, X. Interpretable Hyperspectral Artificial Intelligence: When nonconvex modeling meets hyperspectral remote sensing. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 52–87. [CrossRef]
13. Zheng, X.; Jia, J.; Dong, S.; Wang, Y.; Lu, R.; Chen, Y.; Wang, Y. Training and inference Time Efficiency Assessment Framework for machine learning algorithms: A case study for hyperspectral image classification. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *141*, 104591. [CrossRef]
14. Zhang, Q.; Yuan, Q.; Zeng, C.; Li, X.; Wei, Y. Missing Data Reconstruction in Remote Sensing Image with a Unified Spatial–Temporal–Spectral Deep Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4274–4288. [CrossRef]
15. Li, J.; Zheng, K.; Gao, L.; Han, Z.; Li, Z.; Chanussot, J. Enhanced Deep Image Prior for Unsupervised Hyperspectral Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5504218. [CrossRef]
16. Bi, K.; Li, Z.; Chen, Y.; Du, Q.; Ma, L.; Wang, Y.; Fang, Z.; Qi, M. Open-Set Domain Adaptation for Hyperspectral Image Classification Based on Weighted Generative Adversarial Networks and Dynamic Thresholding. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5507717. [CrossRef]
17. Champa, A.I.; Rabbi, M.F.; Hasan, S.M.M.; Zaman, A.; Kabir, M.H. Tree-Based Classifier for Hyperspectral Image Classification via Hybrid Technique of Feature Reduction. In Proceedings of the 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, 27–28 February 2021; pp. 115–119.
18. Yu, H.; Yang, T.; Zhou, L.; Wang, Y. PDNet: A Lightweight Deep Convolutional Neural Network for InSAR Phase Denoising. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5239309. [CrossRef]
19. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
20. Li, T.; Zhang, J.; Zhang, Y. Classification of hyperspectral image based on deep belief networks. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 5132–5136.
21. Zhao, W.; Li, S.; Li, A.; Zhang, B.; Chen, J. Deep fusion of hyperspectral images and multi-source remote sensing data for classification with convolutional neural network. *Natl. Remote Sens. Bull.* **2021**, *25*, 1489–1502. [CrossRef]
22. Li, H.C.; Li, S.S.; Hu, W.S.; Feng, J.H.; Sun, W.W.; Du, Q. Recurrent Feedback Convolutional Neural Network for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5504405. [CrossRef]
23. Yin, J.; Qi, C.; Chen, Q.; Qu, J. Spatial-Spectral Network for Hyperspectral Image Classification: A 3-D CNN and Bi-LSTM Framework. *Remote Sens.* **2021**, *13*, 2353. [CrossRef]

24. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
25. Hong, D.; Han, Z.; Yao, J.; Gao, L.; Zhang, B.; Plaza, A.; Chanussot, J. SpectralFormer: Rethinking Hyperspectral Image Classification with Transformers. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5518615. [CrossRef]
26. Wang, Y.; Zhang, Y.; Duan, K. Hyperspectral Image Classification Using 3D Attention Mechanism in Collaboration with Transformer. In Proceedings of the 2023 6th International Conference on Artificial Intelligence and Pattern Recognition, Xiamen, China, 22–24 September 2023; pp. 165–172.
27. Ding, K.; Lu, T.; Fu, W.; Li, S.; Ma, F. Global–Local Transformer Network for HSI and LiDAR Data Joint Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5541213. [CrossRef]
28. Wang, Y.; Liu, M.; Yang, Y.; Li, Z.; Du, Q.; Chen, Y.; Li, F.; Yang, H. Heterogeneous Few-Shot Learning for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5510405. [CrossRef]
29. Xi, B.; Zhang, Y.; Li, J.; Zheng, T.; Zhao, X.; Xu, H.; Xue, C.; Li, Y.; Chanussot, J. MCTGCL: Mixed CNN–Transformer for Mars Hyperspectral Image Classification with Graph Contrastive Learning. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5503214. [CrossRef]
30. Xi, B.; Zhang, Y.; Li, J.; Huang, Y.; Li, Y.; Li, Z.; Chanussot, J. Transductive Few-Shot Learning with Enhanced Spectral-Spatial Embedding for Hyperspectral Image Classification. *IEEE Trans. Image Process.* **2025**, *34*, 854–868. [CrossRef]
31. Wei, Z.; Taoyang, M.; Dan, L. Classification of hyperspectral images based on two-channel convolutional neural network combined with support vector machine algorithm. *J. Appl. Remote Sens.* **2020**, *14*, 024514. [CrossRef]
32. Li, S.; Sun, L.; Tian, Y.; Lu, X.; Fu, Z.; Lv, G.; Zhang, L.; Xu, Y.; Che, W. Research on non-destructive identification technology of rice varieties based on HSI and GBDT. *Infrared Phys. Technol.* **2024**, *142*, 105511. [CrossRef]
33. Li, R.; Zheng, S.; Duan, C.; Yang, Y.; Wang, X. Classification of Hyperspectral Image Based on Double-Branch Dual-Attention Mechanism Network. *Remote Sens.* **2020**, *12*, 582. [CrossRef]
34. Qiao, X.; Roy, S.K.; Huang, W. Multiscale Neighborhood Attention Transformer with Optimized Spatial Pattern for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5523815. [CrossRef]
35. Qiao, X.; Huang, W. A Dual Frequency Transformer Network for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 10344–10358. [CrossRef]
36. Alkhatib, M.Q.; Al-Saad, M.; Aburaed, N.; Zitouni, M.S.; Ahmad, H.A. Attention Based Dual-Branch Complex Feature Fusion Network for Hyperspectral Image Classification. In Proceedings of the 2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Athens, Greece, 31 October–2 November 2023; pp. 1–5.
37. Liang, H.; Bao, W.; Shen, X.; Zhang, X. HSI-Mixer: Hyperspectral Image Classification Using the Spectral–Spatial Mixer Representation from Convolutions. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6013005. [CrossRef]
38. Kong, Y.; Wang, X.; Cheng, Y. Spectral–Spatial Feature Extraction for HSI Classification Based on Supervised Hypergraph and Sample Expanded CNN. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4128–4140. [CrossRef]
39. Ahmad, M.; Butt, M.H.F.; Khan, A.M.; Mazzara, M.; Distefano, S.; Usama, M.; Roy, S.K.; Chanussot, J.; Hong, D. Spatial–spectral morphological mamba for hyperspectral image classification. *Neurocomputing* **2025**, *636*, 129995. [CrossRef]
40. Chen, P.; He, W.; Qian, F.; Shi, G.; Yan, J. A synergistic CNN-transformer network with pooling attention fusion for hyperspectral image classification. *Digit. Signal Process.* **2025**, *160*, 105070. [CrossRef]
41. Gao, H.; Chen, Z.; Xu, F. Adaptive spectral-spatial feature fusion network for hyperspectral image classification using limited training samples. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102687. [CrossRef]
42. Guo, T.; Wang, R.; Luo, F.; Gong, X.; Zhang, L.; Gao, X. Dual-View Spectral and Global Spatial Feature Fusion Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5512913. [CrossRef]
43. Liu, X.; Lu, Y.; Wang, X.; Wu, X. Training-Free Multi-style Fusion Through Reference-Based Adaptive Modulation. In Proceedings of the Pattern Recognition and Computer Vision: 8th Asian Conference on Pattern Recognition, ACPR 2025, Proceedings, Part II, Gold Coast, QLD, Australia, 10–13 November 2025; pp. 149–163.
44. Li, M.; Ming, D.; Xu, L.; Dong, D.; Zhang, Y. SFEARNet: A Network Combining Semantic Flow and Edge-Aware Refinement for Highly Efficient Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5402518. [CrossRef]
45. Wu, P.; Pan, Z.; Tang, H.; Hu, Y. Cloudformer: A Cloud-Removal Network Combining Self-Attention Mechanism and Convolution. *Remote Sens.* **2022**, *14*, 6132.
46. Qiao, X.; Huang, W. A Hybrid CNN-Transformer Network for Global Ocean Wind Speed Retrieval from GNSS-R Data. In Proceedings of the OCEANS 2024—Halifax, Halifax, NS, Canada, 23–26 September 2024; pp. 1–4.
47. Cui, Y.; Xie, T.; Li, J.; Zhang, X.; Bai, S.; Wang, C.; Liu, H. A Chlorophyll Concentration Inversion Method Based on OWTs and 1D CNN-Transformer Feature Extraction. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2026**, *19*, 2006–2032. [CrossRef]

48. Lin, X.; Yan, Z.; Deng, X.; Zheng, C.; Yu, L. ConvFormer: Plug-and-Play CNN-Style Transformers for Improving Medical Image Segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2023, Cham, Switzerland, 8–12 October 2023; pp. 642–651.
49. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the Computer Vision—ECCV 2018, Cham, Switzerland, 8–14 September 2018; pp. 3–19.
50. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
51. Mu, Y.-Z.; Wang, Z.; Chen, X.; Chen, J.-J.; Zhao, J.-K.; Wang, J.-M. Deep Learning Test Optimization Method Using Multi-objective Optimization. *Int. J. Softw. Inform.* **2022**, *33*, 2499–2524. [CrossRef]
52. Cen, Y.; Zhang, L.; Zhang, X.; Wang, Y.; Qi, W.; Tang, S.; Zhang, P. Aerial hyperspectral remote sensing classification dataset of Xiongan New Area (Matiwan Village). *Natl. Remote Sens. Bull.* **2020**, *24*, 1299–1306. [CrossRef]
53. Chen, Y.; Zhao, X.; Jia, X. Spectral–Spatial Classification of Hyperspectral Data Based on Deep Belief Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2381–2392. [CrossRef]
54. Wang, D.; Du, B.; Zhang, L.; Xu, Y. Adaptive Spectral–Spatial Multiscale Contextual Feature Extraction for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 2461–2477. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

remote sensing

MDPI

Article

DNMF-AG: A Sparse Deep NMF Model with Adversarial Graph Regularization for Hyperspectral Unmixing

Kewen Qu [1,2,*], Xiaojuan Luo [1,2] and Wenxing Bao [1,2]

[1] The School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China; 20237481@stu.nmu.edu.cn (X.L.); baowenxing@nun.edu.cn (W.B.)
[2] The Image and Intelligence Information Processing Innovation Team of the National Ethnic Affairs Commission of China, Yinchuan 750021, China
* Correspondence: kewen.qu@nmu.edu.cn

Highlights

What are the main findings?

- DNMF-AG integrates sparse deep decomposition with adversarial graph regularization, achieving stronger spatial structural modeling and improved endmember–abundance separation.
- The combination of adversarial graphs, sparsity constraints, and a truncated activation function enhances noise robustness and yields more accurate abundance estimates.

What are the implications of the main findings?

- The proposed method addresses key limitations of deep NMF in noisy environments, including insufficient structural exploitation and weak sparsity, resulting in a more robust and interpretable unmixing framework.
- The improved reliability of endmember and abundance estimation provides a more accurate foundation for subsequent hyperspectral tasks such as target detection, classification, and change detection.

Abstract

Hyperspectral unmixing (HU) aims to extract constituent information from mixed pixels and is a fundamental task in hyperspectral remote sensing. Deep non-negative matrix factorization (DNMF) has recently attracted attention for HU due to its hierarchical representation capability. However, existing DNMF-based methods are often sensitive to noise and outliers, and face limitations in incorporating prior knowledge, modeling feature structures, and enforcing sparsity constraints, which restrict their robustness, accuracy, and interpretability. To address these challenges, we propose a sparse deep NMF model with adversarial graph regularization for hyperspectral unmixing, termed DNMF-AG. Specifically, we design an adversarial graph regularizer that integrates local similarity and dissimilarity graphs to promote intraclass consistency and interclass separability in the spatial domain, thereby enhancing structural modeling and robustness. In addition, a Gram-based sparsity constraint is introduced to encourage sparse abundance representations by penalizing inner product correlations. To further improve robustness and computational efficiency, a truncated activation function is incorporated into the iterative update process, suppressing low-amplitude components and promoting zero entries in the abundance matrix. The overall model is optimized using the alternating direction method of multipliers (ADMM). Experimental results on multiple synthetic and real datasets demonstrate that the proposed method outperforms state-of-the-art approaches in terms of estimation accuracy and robustness.

Keywords: spectral image; hyperspectral unmixing; deep non-negative matrix factorization; adversarial graph learning; gram sparsity

1. Introduction

With the rapid development of hyperspectral imaging technology, hyperspectral sensors can capture narrow and continuous spectral information from ground objects, which not only yields improved ground observation accuracy but also broadens the application potential of hyperspectral remote sensing images (HSI) in fields such as agriculture, medical analysis, and environmental monitoring [1]. However, owing to the limited spatial resolutions of sensors, the phenomenon of "mixed pixels" is common in HSI, where a single pixel often contains mixed spectral information derived from numerous distinct materials. This phenomenon severely limits the precise applicability of HSI in tasks such as clustering [2], fine classification [3], and target detection [4]. Therefore, HU, whose main purpose is to identify pure objects (i.e., endmembers) and their corresponding proportions (i.e., abundances) [5] from mixed pixels, has become an important step for processing HSI.

To date, researchers have proposed many methods for solving the HU problem. These methods can be divided into two main categories: linear mixture models (LMM) [6] and nonlinear mixture models (NLMM) [7]. An LMM assumes that the mixing of ground materials occurs at a macroscopic scale, where each incident solar photon interacts with only a single type of material. In contrast, an NLMM holds that after the incident solar light reaches the ground, it refracts with various ground objects and then reflects to the sensor. Owing to the advantages of LMMs, such as their simple expression forms, strong operability, and clear physical meaning, they are still some of the most widely used models in HU research.

In recent years, many methods based on LMMs have been proposed; these approaches can be classified into geometric, statistical, sparse regression, tensor, and neural network (NN) methods. Among them, the sparse regression (SR) method [8] is a semisupervised method based on the theory of compressed sensing. It assumes that the observed hyperspectral data can be represented by a linear combination of atoms in a preknown spectral library, thereby transforming the unmixing problem into the process of finding the optimal endmember subset in the spectral library for modeling the hyperspectral data. Since the SR method does not require endmember extraction or endmember quantity estimation and can efficiently utilize the rich atomic information contained in the spectral library, it has received extensive attention in practical applications [9]. Ref. [10] first integrated SR into HU (SUnSAL) and used $L_{1,1}$ regularization to constrain the abundance matrix, obtaining ideal abundance estimates. On this basis, a total variation (TV) regularization term was incorporated into SUnSAL, resulting in piecewise smooth abundance maps and significantly improving the robustness of the constructed model [11]. Inspired by this, many SR methods with manually set abundance regularization have emerged, and all of them have produced ideal results [12–14]. However, these methods have problems such as high spectral library coherence and high computational costs. Ref. [15] proposed a two-stage optimization algorithm, NeSU-LP, that combines adaptive spectral library pruning and the Nesterov acceleration strategy, providing significantly improved unmixing efficiency. Overall, although the SR method avoids the complexity of endmember extraction, its performance is limited by the large scale and high coherence of the employed spectral library.

Geometric methods are based on convex geometry theory, which holds that the vertices of the simplex formed by hyperspectral data in the given feature space are the

endmembers. Therefore, the process of endmember extraction is transformed into the process of identifying the vertices of the simplex. This category mainly includes pure pixel methods and nonpure pixel methods. Among them, a pure pixel algorithm assumes that each endmember has at least one pure pixel in the associated hyperspectral image. Typical algorithms include pixel purity index (PPI) [16], N-finder algorithm (N-FINDR) [17], vertex component analysis (VCA) [18], etc. Nonpure pixel methods are based on the minimum simplex theory [19] and generate endmembers by estimating the vertices of the simplex; thus, they can handle highly mixed data.

Statistical methods are rooted in probability theory and mathematical statistics theory, and they have the advantage of not relying on pure pixel assumptions. Among them, non-negative matrix factorization (NMF) is a typical representative [20]. NMF decomposes a non-negative data matrix into the product of two non-negative low-rank submatrices that correspond to the basis matrix and the coefficient matrix [21–23]. However, owing to the nonconvexity of its objective function, the algorithm easily falls into local minima during the optimization process. To address this issue, researchers have integrated many prior constraints into NMF to enhance the physical interpretability and robustness of the model; these constraints include abundance sparsity constraints [24], piecewise smooth constraints [25], and manifold regularization [26]. On the basis of the sparse prior, ref. [27] proposed sparse-constraint NMF by enhancing the abundance sparsity of the model through $L_{1/2}$ regularization and proved that $L_{1/2}$ regularization is easier to solve than L_0 regularization is and has stronger sparsity than L_1 regularization does, making it a better alternative to L_p $(0 < p < 1)$ regularization. In addition, Yuan and Li et al. proposed a region-space adaptive total variation (RSATV) model [28] to address the problem that TV regularization is prone to introducing false edges in flat image regions. By combining regional k-means clustering and dual filtering mechanisms, the spatial consistency and robustness of the constructed abundance map can be improved while preserving edge information. Furthermore, Qu and Bao [29] proposed a multiprior set constraint-based NMF unmixing method, which integrates the minimum volume constraint (MVC), a reweighted $L_{1/2}$ norm, and TV regularization into the NMF framework and imposes constraints on the endmember and abundance matrices in the NMF model, thereby overcoming the limitations of traditional single-matrix constraint methods.

Since NMF usually expands three-dimensional data into a two-dimensional matrix when processing HSIs, this expansion process inevitably leads to the loss of the original spatial structure information. To solve this problem, researchers have proposed non-negative tensor factorization (NTF). NTF methods are based mainly on tensor algebra theory and can be roughly classified into four categories according to their different decomposition strategies. First, canonical polyadic decomposition (CPD) [30], which represents a tensor as the sum of several rank-one tensors, is suitable for depicting simple structures. Second, Tucker decomposition [31], which uses a core tensor and multiple factor matrices to achieve low-rank approximation, has stronger representational capabilities. Third, block tensor decomposition (BTD) [32] retains the local structure of the target tensor while enhancing the interpretability of the model. Fourth, matrix-vector NTF (MV-NTF) [33] effectively improves the computational efficiency and adaptability of the model by combining tensor decomposition and matrix decomposition strategies. Compared with matrix decomposition, NTF can more effectively preserve the original intrinsic spatial structure of HSIs. However, how to reasonably determine the rank of the target tensor is still an important challenge for this method.

With the rapid development of deep learning, especially the outstanding achievements of convolutional NNs in tasks such as hyperspectral image classification and HU [34,35], researchers have begun to integrate multilayer deep structures into the NMF framework.

https://doi.org/10.3390/rs18010155

They aim to construct NMF unmixing methods with multiple levels and high representational capabilities. Ref. [36] proposed multilayer NMF (MLNMF), which iteratively decomposes the target observation matrix into multiple layers to achieve more levels of feature extraction. However, MLNMF is essentially only a series of sequential single-layer decompositions; it lacks cross-layer information feedback and fusion and fails to give full play to the hierarchical modeling advantages of deep structures. To enhance the modeling capability of data structures in HU, ref. [37] proposed a DNMF method incorporating $L_{1/2}$ sparsity constraint and TV regularization, referred to as SDNMF-TV. By leveraging a hierarchical pretraining and fine-tuning strategy, this method effectively captures the layered structural information of hyperspectral data, thereby improving the estimation accuracy of both endmembers and abundances. Building upon this, ref. [38] addressed the performance degradation of deep NMF methods under noise interference by introducing an $L_{1/2}$-norm constraint. The resulting robust DNMF model significantly enhances the model's resilience to outlier noise, confirming its superior unmixing performance. Moreover, the self-supervised robust deep matrix decomposition model (SSRDMF) jointly optimizes the endmember and abundance estimation processes through an encoder–decoder architecture in combination with sparse noise modeling and self-supervised constraints, significantly enhancing the robustness of the unmixing model against Gaussian noise and sparse noise and further improving its unmixing performance [39]. In summary, the DNMF method has significant advantages in terms of achieving improved HU accuracy and anti-noise performance because of its hierarchical modeling capabilities, adaptive optimization mechanism, and regularization strategy.

On the basis of the above analysis, considering the superior characteristics of DNMF in hierarchical feature learning tasks and drawing on the ideas of deep learning, an adversarial graph regularization and Gram sparsity-constrained DNMF method for spectral unmixing is proposed in this paper. Specifically, our method improves and enhances the existing model from three perspectives—spatial structure preservation, sparse representation, and robustness—by introducing graph structure priors and sparsity mechanisms (as shown in Figure 1). First, we utilize the manifold graph learning and adversarial learning concepts to design an adversarial graph regularization method composed of a local similarity graph and a heterogeneous repulsion graph, which effectively enhances the continuity within classes and the separability between classes during abundance estimation, thereby improving the ability of the model to identify complex spatial structures and class boundaries. Second, to effectively characterize the sparsity of the abundance components contained in multilayer structures, we introduce a novel sparsity constraint based on inner product penalties, which guides the task of generating unmixing results by penalizing the Gram matrix of abundances, enhancing the interpretability and discrimination of the model. Finally, the designed adversarial graph regularization scheme and Gram sparse constraint are uniformly integrated into the deep robust NMF model to further increase its robustness and accuracy. Additionally, during the iterative process of deep robust NMF, we innovatively introduce a truncated activation function, which truncates the low-amplitude terms in the hierarchical iterative matrix to further promote the generation of zero values in the abundance matrix, thereby enhancing the robustness of the model to noise and outliers while improving its computational efficiency and interpretability.

https://doi.org/10.3390/rs18010155

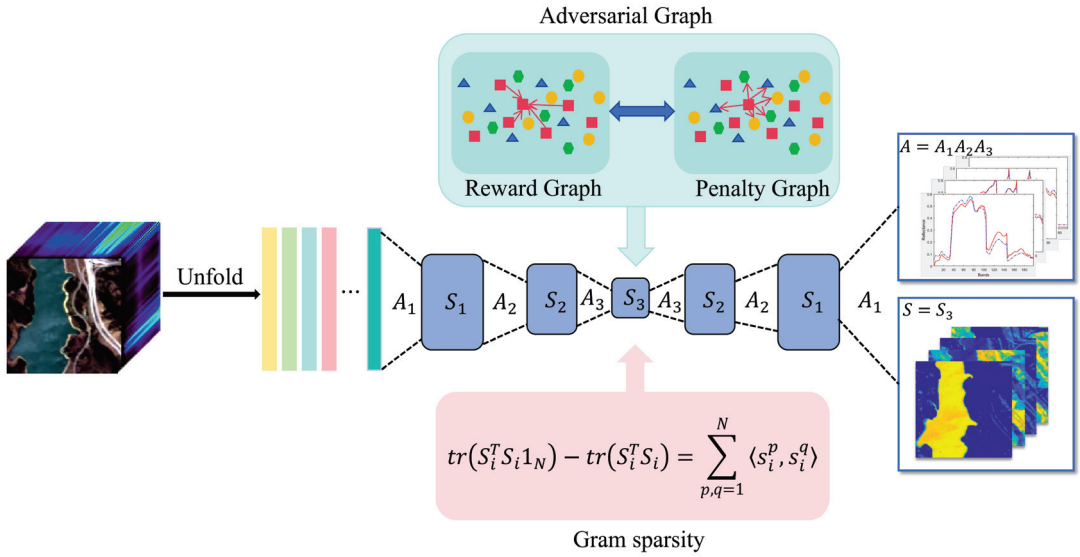


Figure 1. Schematic diagram of the proposed DNMF-AG model.

The main contributions of this study are summarized as follows.

(1) A robust DNMF framework is proposed, integrating adversarial graph regularization, a Gram sparsity constraint, and a reconstruction strategy that incorporates an improved ReLU function and the $L_{2,1}$-norm, thereby enhancing the expressive capacity, sparsity, nonlinearity modeling, and decomposition robustness of the multilayer unmixing model when dealing with complex hyperspectral data.

(2) An adversarial graph regularization method that considers a local similarity graph and a heterogeneous repulsion graph is designed; this approach dynamically balances the intraclass smoothness and interclass distinctiveness of the abundance matrix. Moreover, a sparse constraint based on the Gram inner product is introduced to guide the process of generating abundance components in each layer, further enhancing the sparse representation ability of the model and improving the interpretability of multilayer factor decomposition.

(3) The proposed model is solved using the alternating direction method of multipliers (ADMM), which is adopted as a standard numerical optimization technique to efficiently decompose and solve the resulting constrained optimization problem.

The remainder of this article is organized as follows. Section 2 introduces the relevant basic models. Section 3 describes the model of the proposed algorithm and its optimization algorithm. Section 4 describes and analyzes the results obtained on synthetic datasets and real datasets. Section 5 provides some conclusions and directions for future research.

2. Background

2.1. Linear Mixing Model (LMM)

In this work, all matrices are represented by capital bold letters, vectors are denoted by lowercase letters, and scalars are signified by capital letters.

The classic LMM assumes that the spectra of distinct materials do not interfere with each other. Therefore, a mixed pixel can be represented by a linear combination of the endmember spectra and the associated coefficient (representation abundance). The LMM is defined as follows:

$$\boldsymbol{X} = \boldsymbol{AS} + \boldsymbol{N} \tag{1}$$

where $\boldsymbol{X} \in \mathbb{R}^{B \times N}$ denotes an observation matrix with B bands and N pixels, $\boldsymbol{A} \in \mathbb{R}^{B \times P}$ denotes the endmember matrix, P represents the number of endmembers, $\boldsymbol{S} \in \mathbb{R}^{P \times N}$

denotes the abundance matrix, and $\boldsymbol{N} \in \mathbb{R}^{B\times N}$ denotes the noise matrix. In the LMM, the abundance matrix must satisfy the non-negativity constraint (ANC) and the sum-to-one constraint (ASC):

$$ANC : \boldsymbol{S}_{ij} \geq 0 \qquad (i = 1, 2, \ldots P, j = 1, 2, \ldots N)$$

$$ASC : \sum_{i=1}^{P} \boldsymbol{S}_{ij} = 1 \qquad (j = 1, 2, \ldots, N) \tag{2}$$

2.2. Non-Negative Matrix Factorization (NMF)

NMF is a widely used method in data analysis tasks, especially when the non-negative constraint is considered. The main purpose of NMF is to decompose a non-negative HSI matrix $\boldsymbol{X} \in \mathbb{R}^{B\times N}$ into the product of two non-negative low-rank submatrices. NMF can be written as follows:

$$\boldsymbol{X} \approx \boldsymbol{AS} \tag{3}$$

To measure the quality of the approximation results, the objective function usually minimizes the Euclidean distance between $\boldsymbol{X}$ and $\boldsymbol{AS}$; i.e.,

$$\min_{\boldsymbol{A}\geq 0, \boldsymbol{S}\geq 0} \mathcal{F}_{\mathrm{NMF}}(\boldsymbol{A}, \boldsymbol{S}) = \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{AS}\|_F^2 \tag{4}$$

where the operator $\|\cdot\|_F$ denotes the Frobenius norm and $\geq$ indicates a non-negativity constraint imposed on the values.

When exploring various optimization algorithms, the multiplicative update rule (MUR) is widely used because of its computational simplicity. Minimizing the loss function in (4) via the MUR yields the following update rule:

$$\boldsymbol{A} \leftarrow \boldsymbol{A} \cdot * \left(\boldsymbol{X}\boldsymbol{S}^T\right) \cdot / \left(\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^T\right) \tag{5}$$

$$\boldsymbol{S} \leftarrow \boldsymbol{S} \cdot * \left(\boldsymbol{A}^T\boldsymbol{X}\right) \cdot / \left(\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{S}\right) \tag{6}$$

where the operators $\cdot*$ and $\cdot/$ represent elementwise multiplication and division, respectively. $(\boldsymbol{.})^T$ represents the transpose of a matrix.

2.3. Deep Non-Negative Matrix Factorization (DNMF)

The traditional NMF method decomposes mixed pixels into a linear combination of an endmember spectral matrix and an abundance matrix by constructing a low-dimensional representation. However, in hyperspectral unmixing, the complexity of mixed pixels also arises from nonlinear mixing, endmember spectral variability (e.g., due to illumination or terrain effects), and noise interference. Shallow NMF struggles to capture these complex relationships, limiting the accuracy of endmember identification and abundance estimation. To overcome this limitation, ref. [38] proposed the DNMF framework, which decomposes the given data matrix $\boldsymbol{X}$ into $l + 1$ non-negative matrices through the following steps:

$$\begin{aligned} \boldsymbol{X} &\approx \boldsymbol{A}_1\boldsymbol{S}_1 \\ \boldsymbol{X} &\approx \boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{S}_2 \\ &\vdots \\ \boldsymbol{X} &\approx \boldsymbol{A}_1\boldsymbol{A}_2 \cdots \boldsymbol{A}_l\boldsymbol{S}_l \end{aligned} \tag{7}$$

where l denotes the total number of layers, $\boldsymbol{A}_i \in \mathbb{R}^{P_{i-1}\times P_i}$ represents the endmember matrix at the i-th layer, and $\boldsymbol{S}_l \in \mathbb{R}^{P_i\times N}$ denotes the corresponding abundance matrix. Moreover,

https://doi.org/10.3390/rs18010155

P_0 is the number of spectral bands B, and N is the number of pixels. The size of each layer is given by P_i, where $i = 1, 2, \cdots, l$. This hierarchical structure allows the dimensionality to gradually decrease across successive layers.

$$B = P_0 \geq P_1 \geq \cdots \geq P_{l-1} \geq P_l \tag{8}$$

where P denotes the number of endmembers. The optimization objective for the HU process based on DNMF is defined as follows:

$$\mathcal{F}_{Deep} = \min_{A_i, S_l} \|X - A_1 A_2 \cdots A_l S_l\|_F^2 \tag{9}$$

We perform NMF-based approximation for all layers.

$$A_l \leftarrow A_l \frac{X S_l^T}{A_l S_l S_l^T} \tag{10}$$

$$S_l \leftarrow S_l \frac{A_l^T X}{A_l^T A_l S_l} \tag{11}$$

Specifically, in each layer of the DNMF framework, the endmember matrix A_i and the corresponding abundance matrix S_l are explicitly considered, whereas the abundance matrix S_i is implicitly represented, where $i = 1, 2 \cdots l-1$. To facilitate a better understanding of the DNMF framework, Figure 2 illustrates a three-layer DNMF architecture.

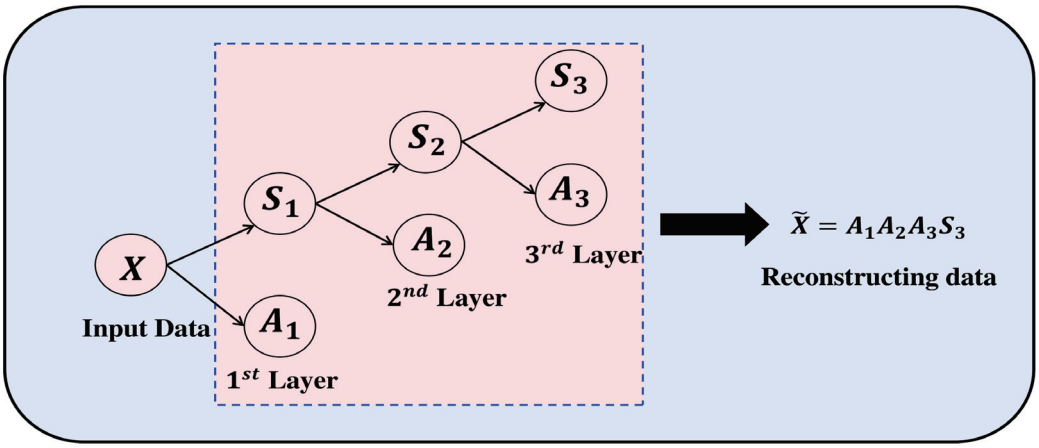


Figure 2. Demonstration of a three-layer DNMF framework. A hyperspectral data matrix $X \in R^{B \times N}$ is given, where B is the number of spectral bands and N is the number of pixels. In the first layer of DNMF, X is decomposed into $A_1 \in R^{B \times P_1}$ and $S_1 \in R^{P_1 \times N}$, where P_1 denotes the dimensionality of the first layer and $P_1 \leq B$. Next, the abundance matrix S_1 is further decomposed into $A_2 \in R^{P_1 \times P_2}$ and $S_2 \in R^{P_2 \times N}$, where P_2 is the dimensionality of the second layer and $P_2 \leq P_1$. This process continues to the third layer, where S_2 is decomposed into $A_3 \in R^{P_2 \times P_3}$ and $S_3 \in R^{P_3 \times N}$, with P_3 being the dimensionality of the third layer and $P_3 \leq P_2$. In summary, the three-layer DNMF process can be written as follows: $X \approx A_1 S_1$, $S_1 \approx A_2 S_2$, and $S_2 \approx A_3 S_3$. Finally, the reconstructed matrix is given by the following: $\widetilde{X} = A_1 A_2 A_3 S_3$.

3. Proposed Model

In this section, we first delve deeply into the mathematical modeling process of DNMF-AG. The optimization process and time complexity analysis of the algorithm are subsequently discussed in detail. Finally, the specific implementation details of the model are discussed.

https://doi.org/10.3390/rs18010155

3.1. Robust DNMF with Truncated Activation Functions

The traditional NMF method relies on linear decomposition and non-negativity constraints to perform matrix decomposition on data. However, in practical HU tasks, the abundance matrix is often disturbed by noise and outliers, causing some elements to deviate from their physically feasible domain, which in turn affects the unmixing accuracy and the stability of the constructed model. To address this issue, truncated activation functions (such as variants of the ReLU) are introduced in this paper; these functions map negative values or minimal values to zero through truncation, thereby achieving a nonlinear mapping that enhances the expressive power of the model while suppressing the interference of outliers during the optimization process. Additionally, this strategy can improve the stability and physical feasibility of abundance estimation while maintaining data sparsity. Its mathematical formalization is as follows:

$$\mathcal{F}_{\mathrm{NMF}} = \|\boldsymbol{X} - \boldsymbol{A}\sigma(\boldsymbol{S})\|_F^2 \tag{12}$$

where $\sigma(\cdot)$ denotes the element-wise truncated nonlinear activation function, which is employed to suppress unreasonable small values in the abundance matrix caused by noise or outliers. For each element s_{ij} in the abundance matrix $\boldsymbol{S}$, let t be the truncation threshold, the truncated activation function is then defined as:

$$\sigma(s_{ij}) = \begin{cases} 0, & s_{ij} \le t \\ s_{ij}, & s_{ij} > t \end{cases} \tag{13}$$

The traditional NMF method usually adopts a loss function that is based on the Frobenius norm. However, when addressing noise and outliers, the Frobenius norm exhibits strong instability. The fundamental reason for this finding is that the error term is usually expressed as the sum of the squared Euclidean norms of residuals (12). When extreme values exist, the squared terms significantly amplify the error, causing a few outliers to disproportionately impact the decomposition results, thereby reducing the stability and unmixing accuracy of the model. To overcome this problem, ref. [40] proposed a robust NMF method by introducing a $L_{2,1}$ norm to redefine the optimization objective, effectively weakening the interference caused by outliers.

$$\mathcal{F}_{\mathrm{NMF}} = \|\boldsymbol{X} - \boldsymbol{A}\,\sigma(\boldsymbol{S})\|_{2,1} \tag{14}$$

In terms of norm selection, the $L_{2,1}$ norm and the Frobenius norm have significant differences. For a given matrix $\boldsymbol{X} \in R^{B\times N}$, the $L_{2,1}$ norm is defined as $\|\boldsymbol{X}\|_{2,1} = \sum_{i=1}^{B} \|x_i\|_2$, whereas the Frobenius norm is defined as $\|\boldsymbol{X}\|_F^2 = \sum_{i=1}^{B} \|x_i\|_2^2$. Since the $L_{2,1}$ norm avoids the calculation of squared errors, it is more effective than the Frobenius norm when handling noisy data.

Inspired by the theoretical advantages of truncated activation functions and the $L_{2,1}$ norm in terms of improving the performance of NMF, we formulate the DNMF problem as follows:

$$\mathcal{F}_{\mathrm{Deep}} = \min_{\boldsymbol{A}_i,\boldsymbol{S}_l} \|\boldsymbol{X} - \boldsymbol{A}_1\sigma(\boldsymbol{A}_2\sigma(\cdots\sigma(\boldsymbol{A}_l\boldsymbol{S}_l)))\|_{2,1} \tag{15}$$

3.2. Adversarial Graph Regularization

Graph theory [41] and manifold learning theory [42] indicate that constructing a nearest-neighbor graph can effectively preserve the local geometric structure of data. However, the existing graph regularization methods (such as the Laplacian regularization

https://doi.org/10.3390/rs18010155

term) often overly rely on the local manifold assumption, focusing only on the similarity between samples and neglecting their differences. This limitation leads to overly strong continuity in the abundance matrix space, thereby damaging the detailed information contained within it. To address this issue, a dual-graph adversarial learning mechanism (Figure 1) is proposed in this paper; this mechanism effectively optimizes the intraclass compactness and interclass separability of the target abundance matrix by dynamically balancing the reward graph and the penalty graph, overcoming the shortcomings of the traditional graph regularization methods.

The reward graph is aimed at enhancing the consistency of the abundance distribution of the pixels within the local neighborhood. Specifically, given a data matrix $\boldsymbol{X} \in R^{B\times N}$, if a pixel x_i is one of the K-nearest neighbors of x_j, similarity connections are connected among the samples of the same type through the K-nearest neighbors (KNN) criterion:

$$\boldsymbol{W}_R(i,j) = \begin{cases} \exp\left(-\dfrac{\left\|x_i - x_j\right\|_2^2}{\tau}\right), & \text{if } x_i \in N_K(x_j) \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

where $N_K(x_j)$ represents the KNN of pixel x_i and τ is a temperature parameter that controls the rate of similarity decay. According to the Laplacian matrix $\boldsymbol{L}_R = \boldsymbol{D}_R - \boldsymbol{W}_R$, $\boldsymbol{D}_R$ is the degree matrix, and $\boldsymbol{D}_R(i,i) = \sum_j \boldsymbol{W}_R(i,j)$. The reward map restricts the abundance vectors of pixels belonging to the same class to be as close as possible in the manifold space, which suppresses the loose intraclass distribution caused by noise or gradual changes in the mixing ratio.

A penalty graph is used to enhance the differences between non-neighboring pixels, preventing the potential overlapping of heterogeneous samples in the abundance space. If pixel x_i is not among the KNNs of pixel x_j, a connecting edge is constructed and assigned a weight:

$$\boldsymbol{W}_P(i,j) = \begin{cases} \exp\left(-\dfrac{\left\|x_i - x_j\right\|_2^2}{\tau}\right), & \text{if } x_i \notin N_K(x_j) \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

Here, the corresponding Laplacian matrix $\boldsymbol{L}_P = \boldsymbol{D}_P - \boldsymbol{W}_P$ and its associated regularization term enhance interclass distinguishability by maximizing the abundance differences between non-neighboring samples.

It is worth noting that the parameter K is a crucial one in the construction of the reward graph and the penalty graph. It defines the local neighborhood range of each pixel, which directly affects the connectivity and discriminability of the graph structure. In the reward graph, K determines the connection strength between pixels of the same type. If it is too small, it will lead to insufficient intra-class information, affecting the consistency modeling of abundance distribution. In the penalty graph, the setting of K affects the degree of separation between different types of samples. If it is too large, it may introduce a large number of non-similar neighbors, causing the graph structure to be overly smoothed and weakening the discriminability between classes. Therefore, the reasonable setting of K requires a balance between intra-class compactness and inter-class separability.

By jointly optimizing the Laplacian constraints of the reward graph and the penalty graph, the objective function is defined as follows in each layer of the DNMF model:

$$G_{Deep} = \alpha tr\left(\boldsymbol{S}_i \boldsymbol{L}_R \boldsymbol{S}_i^T\right) - \beta tr\left(\boldsymbol{S}_i L_P \boldsymbol{S}_i^T\right) \tag{18}$$

https://doi.org/10.3390/rs18010155

where $i = 1, 2 \cdots l$ and α and β are weighting parameters that control the balance between intraclass compactness and interclass separability.

3.3. Gram Sparsity Constraint

From a statistical perspective, each pixel in a hyperspectral image is typically composed of only a few endmembers, so the column vectors of the abundance matrix naturally exhibit sparsity. Traditional element-wise sparsity constraints based on the L_0 or L_1 norm can directly control the sparsity of individual elements; however, solving the L_0 norm is an NP-hard problem, and the L_1 norm often provides limited sparsity in practical unmixing applications. Moreover, these methods generally overlook the redundancy and overlap between column vectors, making it difficult to effectively distinguish the endmember compositions of different pixels. To address this limitation, this study introduces an inner-product-based sparsity regularization method within the proposed DNMF framework, namely the Gram sparsity constraint. This constraint effectively suppresses high correlations between columns, thereby highlighting the primary endmember contributions of each pixel, indirectly enhancing sparsity while reducing the influence of noise on unmixing. Specifically, the constraint penalizes the off-diagonal elements of the Gram matrix, preventing the same endmember from exhibiting similar distributions across different pixel columns, thus reducing inter-column redundancy and avoiding potential mixing ambiguities during the unmixing process [43].

Specifically, considering the abundance matrix $\boldsymbol{S}_i = \left[s_i^1, s_i^2, \cdots, s_i^N\right]$ at the i-th layer, where $i = 1, 2, \cdots, l$, the corresponding Gram matrix $\boldsymbol{S}_i^T \boldsymbol{S}_i$ can be expressed as follows:

$$\boldsymbol{S}_i^T \boldsymbol{S}_i = \begin{pmatrix} \left\langle s_i^1, s_i^1 \right\rangle & \left\langle s_i^1, s_i^2 \right\rangle & \cdots & \left\langle s_i^1, s_i^N \right\rangle \\ \vdots & & \ddots & \vdots \\ \left\langle s_i^N, s_i^1 \right\rangle & \left\langle s_i^N, s_i^2 \right\rangle & \cdots & \left\langle s_i^N, s_i^N \right\rangle \end{pmatrix} \tag{19}$$

where $\left\langle s_i^p, s_i^q \right\rangle$ represents the inner product of the abundance vectors of columns p and q. The off-diagonal elements $\left\langle s_i^p, s_i^q \right\rangle (p \neq q)$ of the Gram matrix reflect the cooperative or competitive relationships among different endmembers during the mixing process. The expression $tr\left(\boldsymbol{S}_i^T \boldsymbol{S}_i 1_N\right) - tr\left(\boldsymbol{S}_i^T \boldsymbol{S}_i\right)$ corresponds to the sum of the off-diagonal elements in the Gram matrix $\boldsymbol{S}_i^T \boldsymbol{S}_i$. In simple terms,

$$tr\left(\boldsymbol{S}_i^T \boldsymbol{S}_i 1_N\right) - tr\left(\boldsymbol{S}_i^T \boldsymbol{S}_i\right) = \sum_{p,q=1}^{N} \left\langle s_i^p, s_i^q \right\rangle \tag{20}$$

where 1_N represents a matrix of all ones. The inner product $\left\langle s_i^p, s_i^q \right\rangle$ approaches zero if and only if s_i^p or s_i^q approaches zero. Therefore, this optimization process can induce sparsity in the abundance matrix along the column dimension, suppressing the global participation of redundant endmembers.

3.4. Proposed DNMF-AG Model

Considering the previous discussion, the structure of DNMF-AG consists of three main parts: (1) robust DNMF with a truncated activation function; (2) an exploration of

https://doi.org/10.3390/rs18010155

the problem of adversarial graphs; and (3) Gram sparsity regularization learning. On this basis, the optimization problem of the proposed model is formulated as follows:

$$\begin{aligned}\min_{\boldsymbol{A}_i,\boldsymbol{S}_i} &\|\boldsymbol{X}-\boldsymbol{A}_1\sigma(\boldsymbol{A}_2\sigma(\cdots\sigma(\boldsymbol{A}_l\boldsymbol{S}_l)))\|_{2,1}+\alpha\,\mathrm{tr}\Big(\boldsymbol{S}_i\boldsymbol{L}_R\boldsymbol{S}_i^T\Big)-\beta\,\mathrm{tr}\Big(\boldsymbol{S}_i\boldsymbol{L}_P\boldsymbol{S}_i^T\Big)\\ &+\gamma\left(\mathrm{tr}\Big(\boldsymbol{S}_i^T\boldsymbol{S}_i1_N\Big)-\mathrm{tr}\Big(\boldsymbol{S}_i^T\boldsymbol{S}_i\Big)\right)\end{aligned} \tag{21}$$

where the first term is a robust DNMF reconstruction with a truncated activation; the second and third are adversarial graph and Gram sparsity regularizations. Here, l is the number of layers, $\sigma(\boldsymbol{.})$ the improved nonlinear activation, and α, β, and γ are the parameters for reward graph, penalty graph, and Gram sparsity, respectively.

3.5. Optimization

To efficiently solve the optimization problem (21), this paper adopts the Lagrange multiplier method as a standard numerical optimization tool. Using this method, the constrained optimization problem (21) is transformed into an equivalent unconstrained formulation, which facilitates subsequent numerical solution.

$$\begin{aligned}\mathcal{F}\Big(\{\boldsymbol{A}_i\}_{i=1}^l,\{\boldsymbol{S}_i\}_{i=1}^l\Big)=&\Big\|\boldsymbol{X}-\boldsymbol{A}_1\,\sigma\Big(\boldsymbol{A}_2\,\sigma\big(\cdots\sigma(\boldsymbol{A}_l\boldsymbol{S}_l)\cdots\big)\Big)\Big\|_{2,1}+\alpha\sum_{i=1}^{l}\mathrm{tr}(\boldsymbol{S}_i\boldsymbol{L}_R\boldsymbol{S}_i^T)\\ &-\beta\sum_{i=1}^{l}\mathrm{tr}(\boldsymbol{S}_i\boldsymbol{L}_P\boldsymbol{S}_i^T)+\gamma\sum_{i=1}^{l}\Big(\mathrm{tr}(\boldsymbol{S}_i^T\boldsymbol{S}_i\boldsymbol{1}_N)-\mathrm{tr}(\boldsymbol{S}_i^T\boldsymbol{S}_i)\Big)\\ &+\sum_{i=1}^{l}\mathrm{tr}(\Theta_i\boldsymbol{A}_i^T)+\sum_{i=1}^{l}\mathrm{tr}(\Delta_i\boldsymbol{S}_i^T).\end{aligned} \tag{22}$$

where $i=1,2,\cdots,l$ and Θ_i and Δ_i are the Lagrange multipliers for $\boldsymbol{A}_i\in\mathbb{R}^{P_{l-1}\times P_l}$ and $\boldsymbol{S}_i\in\mathbb{R}^{P_l\times N}$, respectively.

(1) Calculate $\boldsymbol{A}_i$:

In subproblem $\boldsymbol{A}_i$, the following two situations may occur.

- Case 1: $i=1$. We assume that $\psi_2=\sigma(\boldsymbol{A}_2\sigma(\cdots\sigma(\boldsymbol{A}_l\boldsymbol{S}_l)))$.

$$\begin{aligned}\mathcal{F}(\boldsymbol{A}_1)&=\|\boldsymbol{X}-\boldsymbol{A}_1\psi_2\|_{2,1}+tr\Big(\Theta_1\boldsymbol{A}_1^T\Big)\\ &=-2tr\Big(\boldsymbol{X}^T\boldsymbol{A}_1\psi_2\Big)+tr\Big(\psi_2^T\boldsymbol{A}_1^T\boldsymbol{A}_1\psi_2\Big)+tr\Big(\Theta_1\boldsymbol{A}_1^T\Big)\end{aligned} \tag{23}$$

This means that

$$\frac{\partial\mathcal{F}}{\partial\boldsymbol{A}_1}=-2\boldsymbol{X}\psi_2^T+2\boldsymbol{A}_1\psi_2\psi_2^T+\Theta_1 \tag{24}$$

Based on the KKT condition $\Theta_1\boldsymbol{A}_1=0$ and assuming $\partial F/\partial\boldsymbol{A}_1=0$, the update rule for $\boldsymbol{A}_1$ is:

$$\boldsymbol{A}_1\leftarrow\boldsymbol{A}_1\sqrt{\frac{\boldsymbol{X}\psi_2^T}{\boldsymbol{A}_1\psi_2\psi_2^T}} \tag{25}$$

- Case 2: $i=2,3,\cdots,l$. We assume that $\varphi_{i-1}=\boldsymbol{A}_1\sigma(\boldsymbol{A}_2\sigma(\cdots\sigma(\boldsymbol{A}_{i-1})))$ and $\psi_{i+1}=\sigma(\boldsymbol{A}_{i+1}\sigma(\cdots\sigma(\boldsymbol{A}_l\boldsymbol{S}_l)))$ are ignored in the function $\mathcal{F}$, we obtain:

$$\begin{aligned}\mathcal{F}(\boldsymbol{A}_i)&=\|\boldsymbol{X}-\varphi_{i-1}\boldsymbol{A}_i\psi_{i+1}\|_{2,1}+\mathrm{tr}\Big(\Theta_i\boldsymbol{A}_i^T\Big)\\ &=-2\,\mathrm{tr}\Big(\boldsymbol{X}^T\varphi_{i-1}\boldsymbol{A}_i\psi_{i+1}\Big)+\mathrm{tr}\Big(\psi_{i+1}^T\boldsymbol{A}_i^T\varphi_{i-1}^T\varphi_{i-1}\boldsymbol{A}_i\psi_{i+1}\Big)\end{aligned} \tag{26}$$

Hence,

$$\frac{\partial\mathcal{F}}{\partial\boldsymbol{A}_i}=-2\varphi_{i-1}^T\boldsymbol{X}\psi_{i+1}^T+2\varphi_{i-1}^T\varphi_{i-1}\boldsymbol{A}_i\psi_{i+1}\psi_{i+1}^T \tag{27}$$

https://doi.org/10.3390/rs18010155

According to the KKT condition, the update rule for $\boldsymbol{A}_i$ is as follows:

$$\boldsymbol{A}_i \leftarrow \boldsymbol{A}_i \sqrt{\frac{\varphi_{i-1}^T \boldsymbol{X} \psi_{i+1}^T}{\varphi_{i-1}^T \varphi_{i-1} \boldsymbol{A}_i \psi_{i+1} \psi_{i+1}^T}} \tag{28}$$

(2) Calculate $\boldsymbol{S}_i$: In subproblem $\boldsymbol{S}_i$, the following two situations may occur:

- Case 1: For $i = l$, we assume that $\varphi_l = \boldsymbol{A}_1 \sigma(\boldsymbol{A}_2 \sigma(\cdots \sigma(\boldsymbol{A}_l)))$, and ignoring irrelevant terms in $\mathcal{F}$.

$$\begin{aligned} \mathcal{F}(S_l) &= \|\boldsymbol{X} - \varphi_l \boldsymbol{S}_l\|_{2,1} + \alpha tr\left(\boldsymbol{S}_l \boldsymbol{L}_R \boldsymbol{S}_l^T\right) - \beta tr\left(\boldsymbol{S}_l \boldsymbol{L}_P \boldsymbol{S}_l^T\right) + \gamma\left(tr\left(\boldsymbol{S}_l^T \boldsymbol{S}_l 1_N\right) - tr\left(\boldsymbol{S}_l^T \boldsymbol{S}_l\right)\right) + tr\left(\Delta_l \boldsymbol{S}_l^T\right) \\ &= -2tr\left(\boldsymbol{X}^T \varphi_l \boldsymbol{S}_l\right) + tr\left(\boldsymbol{S}_l^T \varphi_l^T \varphi_l \boldsymbol{S}_l\right) + \alpha tr\left(\boldsymbol{S}_l \boldsymbol{L}_R \boldsymbol{S}_l^T\right) - \beta tr\left(\boldsymbol{S}_l \boldsymbol{L}_P \boldsymbol{S}_l^T\right) + \gamma\left(tr\left(\boldsymbol{S}_l^T \boldsymbol{S}_l 1_N\right) - tr\left(\boldsymbol{S}_l^T \boldsymbol{S}_l\right)\right) + tr\left(\Delta_l \boldsymbol{S}_l^T\right) \end{aligned} \tag{29}$$

Hence,

$$\frac{\partial \mathcal{F}}{\partial \boldsymbol{S}_l} = -2\varphi_l^T \boldsymbol{X} + 2\varphi_l^T \varphi_l \boldsymbol{S}_l + 2\alpha \boldsymbol{S}_l \boldsymbol{L}_R - 2\beta \boldsymbol{S}_l \boldsymbol{L}_P + 2\gamma(\boldsymbol{S}_l 1_N - \boldsymbol{S}_l) + \Delta_l \tag{30}$$

According to the KKT condition, the update rule for $\boldsymbol{S}_l$ is as follows:

$$\boldsymbol{S}_l \leftarrow \boldsymbol{S}_l \sqrt{\frac{\varphi_l^T \boldsymbol{X} + \beta \boldsymbol{S}_l \boldsymbol{L}_P + \gamma \boldsymbol{S}_l}{\varphi_l^T \varphi_l \boldsymbol{S}_l + \alpha \boldsymbol{S}_l \boldsymbol{L}_R + \gamma \boldsymbol{S}_l 1_N}} \tag{31}$$

- Case 2: For $i = 1, 2, \cdots, l-1$, assuming $\varphi_i = \boldsymbol{A}_1 \sigma(\boldsymbol{A}_2 \sigma(\cdots \sigma(\boldsymbol{A}_l)))$ and ignoring irrelevant terms in $\mathcal{F}$, we have:

$$\begin{aligned} \mathcal{F}(S_i) &= \|\boldsymbol{X} - \varphi_i \boldsymbol{S}_i\|_{2,1} + \alpha tr\left(\boldsymbol{S}_i \boldsymbol{L}_R \boldsymbol{S}_i^T\right) - \beta tr\left(\boldsymbol{S}_i \boldsymbol{L}_P \boldsymbol{S}_i^T\right) + \gamma\left(tr\left(\boldsymbol{S}_i^T \boldsymbol{S}_i 1_N\right) - tr\left(\boldsymbol{S}_i^T \boldsymbol{S}_i\right)\right) + tr\left(\Delta_i \boldsymbol{S}_i^T\right) \\ &= -2tr\left(\boldsymbol{X}^T \varphi_i \boldsymbol{S}_i\right) + tr\left(\boldsymbol{S}_i^T {\varphi_i}^T \varphi_i \boldsymbol{S}_i\right) + \alpha tr\left(\boldsymbol{S}_i \boldsymbol{L}_R \boldsymbol{S}_i^T\right) - \beta tr\left(\boldsymbol{S}_i \boldsymbol{L}_P \boldsymbol{S}_i^T\right) + \gamma\left(tr\left(\boldsymbol{S}_i^T \boldsymbol{S}_i 1_N\right) - tr\left(\boldsymbol{S}_i^T \boldsymbol{S}_i\right)\right) + tr\left(\Delta_i \boldsymbol{S}_i^T\right) \end{aligned} \tag{32}$$

The following can then be obtained:

$$\frac{\partial \mathcal{F}}{\partial \boldsymbol{S}_i} = -2\varphi_i^T \boldsymbol{X} + 2\varphi_i^T \varphi_i \boldsymbol{S}_i + 2\alpha \boldsymbol{S}_i \boldsymbol{L}_R - 2\beta \boldsymbol{S}_i \boldsymbol{L}_P + 2\gamma(\boldsymbol{S}_i 1_N - \boldsymbol{S}_i) + \Delta_i \tag{33}$$

According to the KKT condition, the update rule for $\boldsymbol{S}_i$ is as follows:

$$\boldsymbol{S}_i \leftarrow \boldsymbol{S}_i \sqrt{\frac{\varphi_i^T \boldsymbol{X} + \beta \boldsymbol{S}_i \boldsymbol{L}_P + \gamma \boldsymbol{S}_i}{\varphi_i^T \varphi_i \boldsymbol{S}_i + \alpha \boldsymbol{S}_i \boldsymbol{L}_R + \gamma \boldsymbol{S}_i 1_N}} \tag{34}$$

The algorithm process can be seen in Algorithm 1.

3.6. Complexity Analysis

This section analyzes the computational complexity of the proposed method, which consists of hierarchical pretraining and global fine-tuning.

During the hierarchical pretraining stage, the classic NMF algorithm is adopted to initialize the decomposition layers at each level. Therefore, the computational complexity of this step can be expressed as $\mathrm{O}\left(It_{pre}\left(BNP_i + BP_i^2 + NP_i^2\right)\right)$, where P_i represents the size of each layer, and It_{pre} represents the number of iterations required by the NMF process. In the fine-tuning stage, each layer involves three main tasks: (1) constructing auxiliary matrices φ_{i-1} and ψ_{i+1}, (2) updating the matrix $\boldsymbol{A}_i$, and (3) updating the matrix $\boldsymbol{S}_i$. Constructing φ_{i-1} and ψ_{i+1} requires at most $\mathrm{O}\left(It_{fine}\left(P_i^3 + BP_i^2\right)\right)$ computational complexity. Moreover, the computational complexities for updating matrices $\boldsymbol{A}_i$ and $\boldsymbol{S}_i$

https://doi.org/10.3390/rs18010155

are $\mathrm{O}\left(It_{fine}\left(B^2P_{i-1}+BP_{i-1}^2\right)\right)$ and $\mathrm{O}\left(It_{fine}\left(N^2P_i+NP_i^2\right)\right)$, respectively, where It_{fine} is the maximum number of iterations involved in the fine-tuning stage of the proposed method. The comprehensive analysis indicates that when the size of each layer P_i is less than $\{B, N\}$ and $P_i \leq P_{i-1}$, the overall computational complexity of a single layer can be approximated as $\mathrm{O}\left(It_{fine}\left(N^2P_i+NP_i^2+B^2P_{i-1}+BP_{i-1}^2\right)\right)$. Therefore, by summing the costs of both the hierarchical pretraining and global fine-tuning stages across all layers, the total computational complexity of the proposed method is approximately $\mathrm{O}\big(l\cdot It_{\mathrm{pre}}(BNP+BP^2+NP^2)+l\cdot It_{\mathrm{fine}}(N^2P+NP^2+B^2P+BP^2+P^3)\big)$, where l is the number of layers.

Algorithm 1: Algorithm of the proposed method.

Input: A hyperspectral image $\boldsymbol{X}\in\mathbb{R}^{B\times N}$; the number of layers l; the size of each layer $P_i(i=1,2,\ldots,l)$; the parameters α,β,γ.
Pretraining all layers:
for $i=1,2,\ldots,l$ **do**
  $(\boldsymbol{A}_i,\boldsymbol{S}_i)\leftarrow \mathrm{NMF}(\boldsymbol{X},P_i)$.
  Set $\boldsymbol{X}=\boldsymbol{S}_i$.
end for
Fine-tuning all layers:
repeat
  for all layers **do**
    Define $\varphi_{i-1}=\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_{i-1}$, and set $\varphi_0=\boldsymbol{I}$.
    Define $\psi_{i+1}=\boldsymbol{A}_{i+1}\cdots\boldsymbol{A}_l$.
    Update $\boldsymbol{A}_i$ by using the rules in (25) and (28).
    Update $\boldsymbol{S}_i$ by using the rules in (31) and (34).
  end for
until convergence is achieved
Output: Set $\boldsymbol{A}=\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_l$ and $\boldsymbol{S}=\boldsymbol{S}_l$.

3.7. Implementation Details

To effectively implement the proposed algorithm, several issues need to be considered.

The first issue concerns the initialization of the endmember and abundance matrices. In our implementation, vertex component analysis (VCA) is used only to initialize the matrix $\boldsymbol{A}_1$, and the fully constrained least-squares (FCLS) algorithm is employed to initialize the matrix $\boldsymbol{S}_1$; the subsequent decomposition process follows the standard layer-wise structure of deep non-negative matrix factorization, without introducing any additional initialization strategies.

The second issue is how to ensure that the two LMM constraints are satisfied during the unmixing process. According to the MUR, we conclude that as long as the initial $\boldsymbol{A}_i$ and $\boldsymbol{S}_i$ are non-negative, the non-negative constraint can be achieved. In each iteration, we replace $\boldsymbol{X}$ and $\boldsymbol{A}$ with their enhanced forms:

$$\overline{\mathbf{X}}=\begin{pmatrix}\mathbf{X}\\ \delta\mathbf{1}_N^T\end{pmatrix},\quad \overline{\mathbf{A}}=\begin{pmatrix}\mathbf{A}\\ \delta\mathbf{1}_P^T\end{pmatrix} \tag{35}$$

where δ controls the weight of the ASC constraint in the objective function. Larger values improve ASC accuracy but slow convergence; therefore, we set $\delta=25$ to balance accuracy and convergence speed.

The third issue is choosing the number of layers in DNMF. Too few layers may fail to capture the complexity of the data, while too many can cause overfitting or higher computational cost. We conducted experiments on simulated datasets to evaluate different layer numbers and determined the optimal decomposition depth for subsequent experiments.

https://doi.org/10.3390/rs18010155

The fourth consideration is the use of the truncated activation function. This function enhances the sparsity of the abundance matrix and improves the interpretability of the unmixing results. However, different truncation thresholds can impact sparsity, accuracy, and computational efficiency. Based on ablation experiments on SC1, and taking SAD, RMSE, and runtime into account, a threshold of 1×10^{-5} was selected as the final value.

Finally, two stopping criteria are defined. The first is an error threshold: if the reconstruction error does not exceed this threshold for ten consecutive iterations, the iteration process is terminated. We set this threshold as 1×10^{-4}. The second criterion is the maximum number of iterations. In this model, we set the maximum number of iterations to 3000 to ensure full algorithmic convergence.

4. Experimental Results and Discussion

To comprehensively validate the effectiveness of the proposed algorithm, we conducted systematic experiments on two simulated datasets and four publicly available real hyperspectral datasets, comparing the results with several methods, including VCAFCLS (VCA [18] + FCLS [44]), L1/2NMF [27], GLNMF [45], MVNTFTV [33], MLNMF [36], SDNMFTV [37], HGNMFFS [46], BUDDIP [47], FaSUn [48], SGLRWRSU [49] and MSGACD [50]. All the experiments were conducted on a Windows™ 11 system with a 3.80-GHz AMD® Ryzen™ 7 7840H CPU and 16 GB of RAM using MATLAB® 2019a.

For quantitative evaluation purposes, the SAD and RMSE were used as metrics. They are defined as follows:

$$SAD = \sum_{i=1}^{P} \arccos\left(\frac{a_i^T \widetilde{a}_i}{\|a_i\| \cdot \|\widetilde{a}_i\|}\right) \tag{36}$$

$$RMSE = \left(\frac{1}{N}\sum_{i=1}^{n} \|s_i - \widetilde{s}_i\|_2^2\right)^{1/2} \tag{37}$$

where a_i and $\widetilde{a}_i$ denote the i-th estimated and true endmembers, and s_i and $\widetilde{s}_i$ denote the i-th estimated and true abundances.

4.1. Experiments Conducted on Synthetic Dataset

The experimental data generation involves two steps: endmember selection and abundance estimation. To evaluate the effectiveness of the proposed algorithm in HU, two simulated datasets, SC1 and SC2, are employed. Specifically, SC1 is constructed by randomly selecting six endmembers from the USGS spectral library Figure 3, generating abundances using the method in [51], and synthesizing hyperspectral data via the LMM, resulting in an image with a spatial resolution of 57 × 57 pixels and 188 spectral bands. Similarly, SC2 is generated by randomly selecting five endmembers from the USGS spectral library (see Figure 4), producing abundances with the method in [11], and applying the LMM to obtain a simulated dataset with 75 × 75 pixels and 188 bands.

To ensure a reliable and fair comparison, all competing NMF-based methods were uniformly initialized using the VCA and FCLS algorithms, while other baseline methods adopted their default initialization strategies as recommended in their original publications. To mitigate the influence of randomness, all experiments were randomly repeated twenty times, and the mean and standard deviation of each evaluation metric were reported.

Furthermore, all comparison algorithms were configured with their optimal parameter settings, see Table 1. VCAFCLS is used as a baseline method and does not involve any parameters. In L1/2NMF, λ controls the weight of the sparsity term. GLNMF introduces graph into sparse NMF, where λ, μ, and δ correspond to the weights of the sparsity, graph and the ASC constraint, respectively. MVNTFTV incorporates TV regularization, with λ regulating the strength of the TV term. MLNMF adopts a multilayer decomposition struc-

https://doi.org/10.3390/rs18010155

ture, where L denotes the number of decomposition layers and α_0 is the sparsity parameter. SDNMFTV combines deep NMF with TV(α) and sparsity regularization(λ), in which L represents the number of layers. HGNMFFS is based on high-order graph regularization, where λ and μ denote regularization weights, and β_1 and β_1 balance the first and second-order neighborhood constraints. BUDDIP employs a unified framework to address both linear and nonlinear blind unmixing problems, where $\alpha_1 - \alpha_4$ regulate the consistency between the network output and the initialization guidance, α_5 and α_6 control the data fitting term. In FaSUn, T_A and T_B denote the numbers of inner-loop iterations in the ADMM optimization for abundance and endmember estimation, respectively, while $\mu_1 - \mu_3$ are penalty parameters associated with the corresponding constraints. In SGLRWRSU, λ_s and λ_g control the sparsity and graph Laplacian regularization, respectively. In MSGACD, λ_1 and λ_2 are used to balance the reconstruction error and the abundance regularization term.

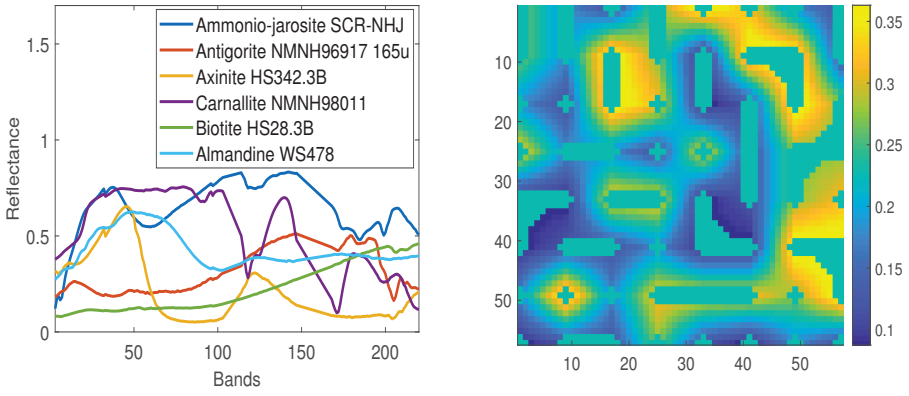


Figure 3. Synthetic data (SC1): (**a**) endmember spectra and (**b**) the simulated image.

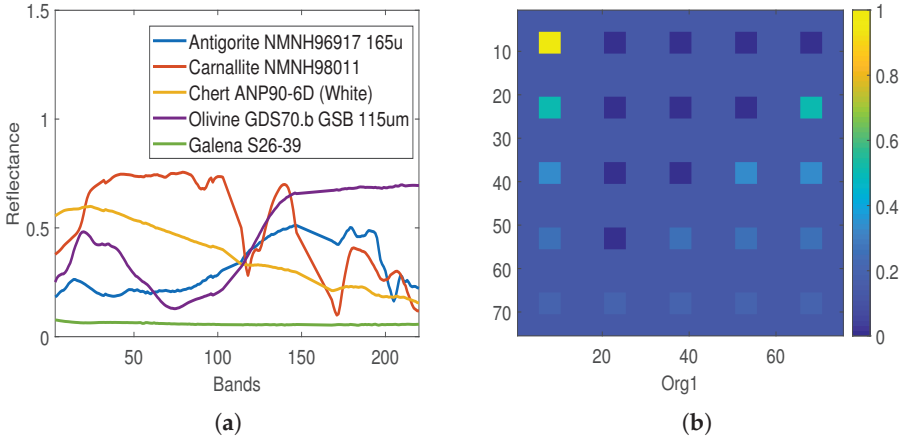


Figure 4. Synthetic data (SC2): (**a**) endmember spectra and (**b**) the simulated image.

Table 1. Parameter settings of the comparison algorithms.

Algorithm	VCAFCLS	L1/2NMF	GLNMF	MVNTFTV	MLNMF	SDNMFTV	HGNMFFS	BUDDIP	FaSUn	SGLRWRSU	MSGACD
Parameter	None	$\lambda = 0.1$	$\lambda = 0.1$ $\mu = 0.1$ $\delta = 15$	$\lambda = 0.1$ $\delta = 15$	$L = 10$ $\alpha_0 = 0.5$	$L = 3$ $\lambda = 0.02$ $\alpha = 0.005$ $\mu = 1000$	$\lambda = 0.01$ $\mu = 0.05$ $\beta_1 = 0.3$ $\beta_2 = 0.7$	$\alpha_{1,3,5} = 1.0$ $\alpha_2 = 0.001$ $\alpha_4 = 0.01$ $\alpha_6 = 0.1$	$T_A = T_B = 5$ $\mu_1 = 50$ $\mu_2 = 2$ $\mu_3 = 1$	$\lambda_s = 1$ $\lambda_g = 1e-3$	$\lambda_1 = 0.1$ $\lambda_2 = 0.002$

4.1.1. Experiments on SC1

Experiment 1(Parameter Selection):In this study, the parameters α, β, and γ respectively control the contribution weights of the reward graph, the penalty graph, and the Gram sparsity regularization term. To determine the optimal parameter combination, we conducted a systematic grid search on simulated data with an SNR of 30 dB under the DNMF framework. The candidate sets were α was $\{0, 5 \times 10^{-3}, 1 \times 10^{-2}, 5 \times 10^{-2}, 1 \times 10^{-1}, 2 \times 10^{-1}, 3 \times 10^{-1}\}$, that for β was $\{0, 5 \times 10^{-3}, 1 \times 10^{-2}, 2 \times 10^{-2}, 3 \times 10^{-2}, 5 \times 10^{-2}, 1 \times 10^{-1}\}$, and that for γ was $\{0, 1 \times 10^{-4}, 3 \times 10^{-4}, 1 \times 10^{-3}, 3 \times 10^{-3}, 1 \times 10^{-2}, 1 \times 10^{-1}\}$. As shown in Figure 5, the grid search results indicate that the combination $\alpha = 5 \times 10^{-2}$, $\beta = 2 \times 10^{-2}$, and $\gamma = 3 \times 10^{-3}$ achieves the best performance, and was thus selected as the global optimal parameter set.

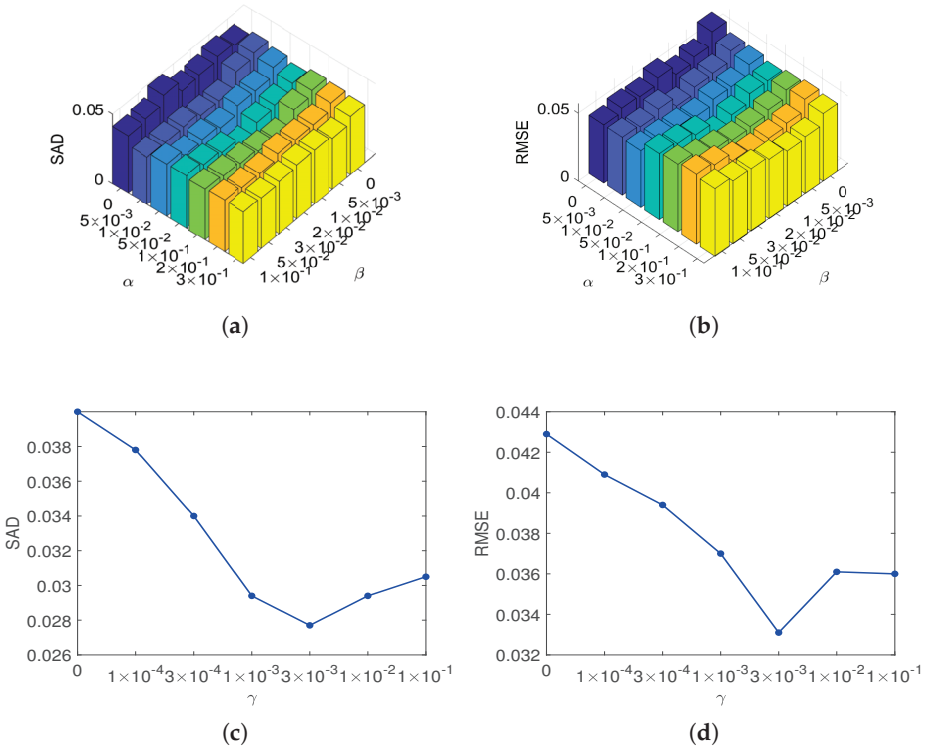


Figure 5. Parameter analysis: (**a**) SAD values corresponding to parameters α and β; (**b**) RMSE values corresponding to parameters α and β; (**c**) SAD values corresponding to parameter γ; (**d**) RMSE values corresponding to parameter γ.

Experiment 2(Sensitivity analysis of the parameter K): To systematically evaluate the impact of the parameter K on the effectiveness of adversarial graph regularization (as shown in Figure 6), all other parameters were fixed, and K was set to 3, 5, 10, 15, 20. The value of K controls the size of the local neighborhood in the graph structure, directly affecting its connectivity and discriminative capability. A smaller K results in an overly sparse graph that is sensitive to local noise and lacks sufficient intra-class information, while a larger K introduces too many inter-class neighbors, blurring class boundaries and weakening the discriminative power. In contrast, $K = 5$ strikes a good balance between intra-class consistency and inter-class separability, more accurately reflecting the actual distribution of mixed pixels and thus improving the accuracy and robustness of the unmixing model.

Experiment 3(Layer Analysis): The proposed model hierarchically represents hyperspectral data through a cascaded structure of multiple factor matrices, with the number of layers l controlling its complexity and representational capacity. Experiments on the SC1 dataset with an SNR of 30 dB (Figure 7) show that when $l = 1$ (NMF), SAD and RMSE reach

https://doi.org/10.3390/rs18010155

their maximum, while deep models ($l \geq 2$) clearly outperform the shallow structure. As l increases to 3, the model achieves the minimum SAD and RMSE; therefore, in this study, the training depth is set to $l = 3$ to balance performance and computational complexity.

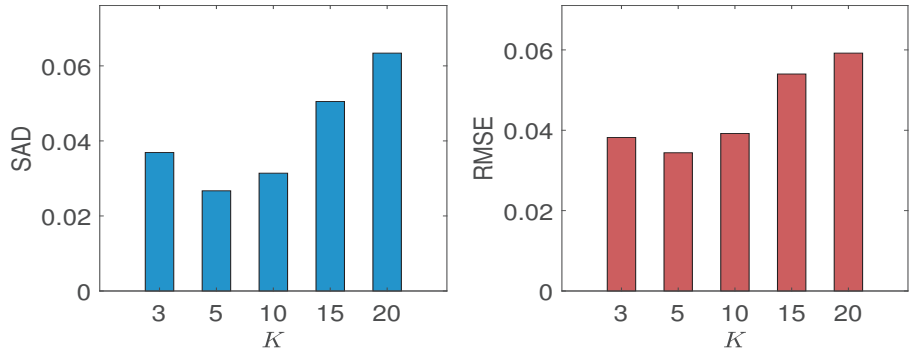
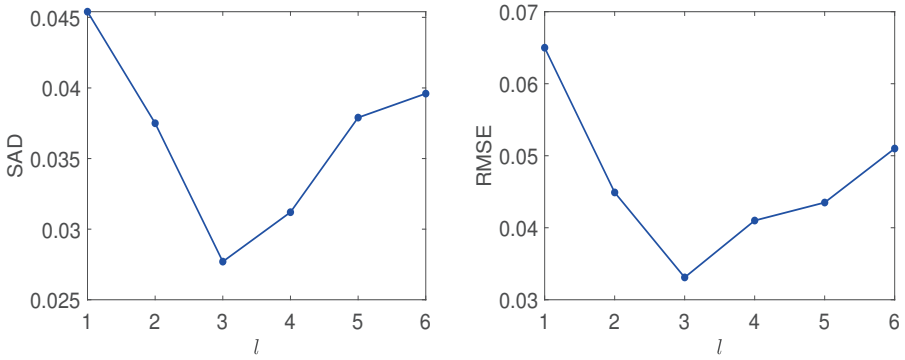


Figure 6. The impact of different K values on unmixing performance.



(**a**) (**b**)

Figure 7. Results of a performance analysis conducted with different numbers of layers: (**a**) SAD and (**b**) RMSE.

Experiment 4(Robustness Analysis): This experiment evaluated different noise scenarios by setting the SNR to 20, 30, and 40 dB. Table 2 shows the SAD and RMSE values of the proposed and comparison methods at these noise levels. The results indicate that the proposed method consistently achieves the lowest SAD and RMSE, significantly outperforming existing methods. Figure 8 presents the abundance maps for six endmembers at SNR = 30 dB, showing that our method accurately estimates ground object distributions. The last row of Table 2 reports the average running time on the SC1 dataset, indicating that VCAFCLS, being non-iterative, is the fastest, while the DNMF-based algorithm remains less time-efficient, with potential for future improvement.

Table 2. Results of SAD, RMSE, and Time under different noise levels on SC1 (best values in bold).

SNR	Metric		VCAFCLS [2005]	L1/2NMF [2011]	GLNMF [2013]	MVNTFTV [2019]	MLNMF [2015]	SDNMFTV [2018]	HGNMFFS [2023]	BUDDIP [2024]	FaSUn [2024]	SGLRWRSU [2024]	MSGACD [2025]	Proposed
SNR = 20	SAD	Mean	0.1349	0.0851	0.0714	0.1135	0.0725	0.1079	0.0927	0.0539	/	/	0.0805	**0.0397**
		Std	±2.04%	±2.78%	±2.61%	±2.42%	±1.47%	±2.81%	±3.02%	±1.73%	/	/	±3.05%	±2.83%
	RMSE	Mean	0.1187	0.0984	0.0899	0.1022	0.0691	0.0966	0.0976	0.0590	0.0996	0.0661	0.0752	**0.0446**
		Std	±2.64%	±2.22%	±2.93%	±2.52%	±1.09%	±2.34%	±3.77%	±1.78%	±1.03%	±1.72%	±2.48%	±2.52%
SNR = 30	SAD	Mean	0.1273	0.0737	0.0694	0.0801	0.0518	0.0766	0.0625	0.0358	/	/	0.0328	**0.0267**
		Std	±1.81%	±1.11%	±1.53%	±3.01%	±1.10%	±0.98%	±2.85%	±1.18%	/	/	±0.82%	±0.41%
	RMSE	Mean	0.1088	0.0910	0.0884	0.0962	0.0427	0.0946	0.0808	0.0557	0.0428	0.0495	0.0394	**0.0344**
		Std	±2.00%	±0.95%	±1.04%	±2.68%	±1.57%	±0.64%	±3.40%	±1.73%	±0.90%	±0.56%	±1.28%	±0.37%
SNR = 40	SAD	Mean	0.1058	0.0702	0.0404	0.0614	0.0581	0.0737	0.0248	0.0212	/	/	0.0332	**0.0166**
		Std	±1.08%	±0.91%	±1.81%	±3.30%	±0.83%	±1.20%	±1.13%	±0.96%	/	/	±0.31%	±0.47%
	RMSE	Mean	0.0915	0.0713	0.0487	0.0639	0.0410	0.0849	0.0493	0.0298	0.0275	0.0258	0.0377	**0.0204**
		Std	±1.04%	±0.94%	±2.25%	±1.42%	±0.63%	±0.61%	±0.90%	±1.04%	±0.76%	±1.06%	±0.32%	±0.46%
Times [s]			1.924	15.48	39.87	133.89	17.51	67.17	30.51	600.23	90.60	28.37	869.35	30.22

Note: FaSUn and SGLRWRSU are sparse regression methods, no endmember estimation needed.

https://doi.org/10.3390/rs18010155

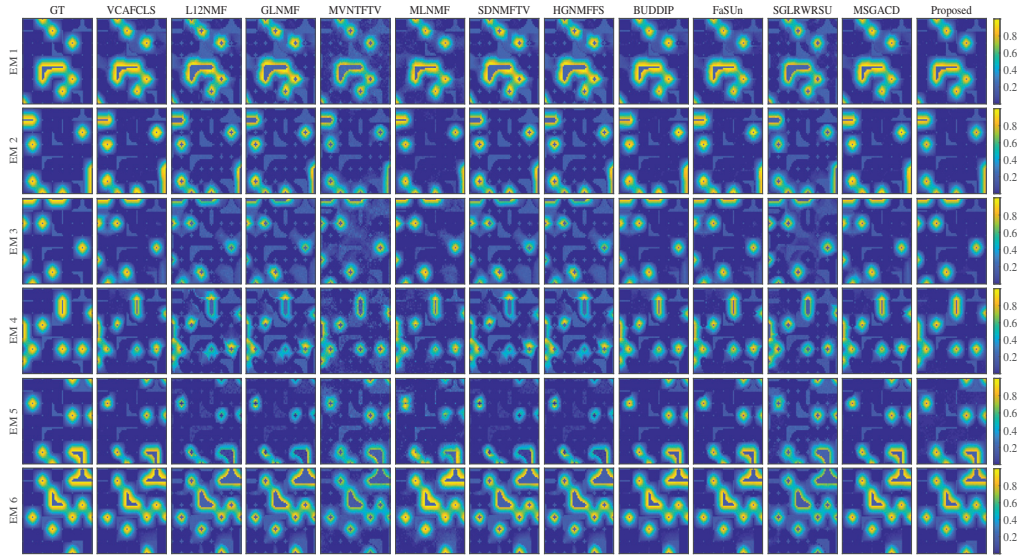


Figure 8. Abundance maps produced by all algorithms for SC1.

Experiment 5(Different Mixing Levels): This experiment systematically evaluated the impact of different mixing levels on HU algorithms. In SC1 with SNR = 30 dB, hyperspectral data with varying purities were generated by adjusting the mixing parameter θ (0.5–0.9), where higher θ indicates a lower level of data mixing. As shown in Figure 9, the proposed deep unmixing model consistently outperformed others across all mixing levels, maintaining the lowest errors even at $\theta = 0.9$. This demonstrates the algorithm's robustness to mixing-level variations and the multilevel analysis advantage of its deep architecture.

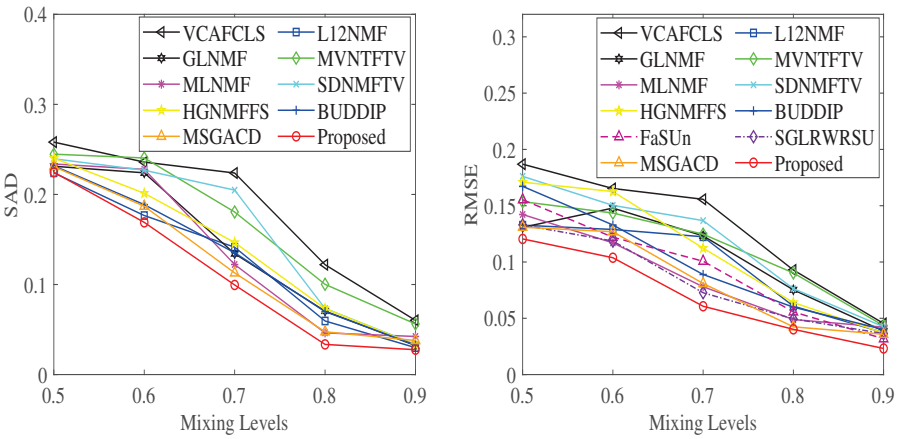


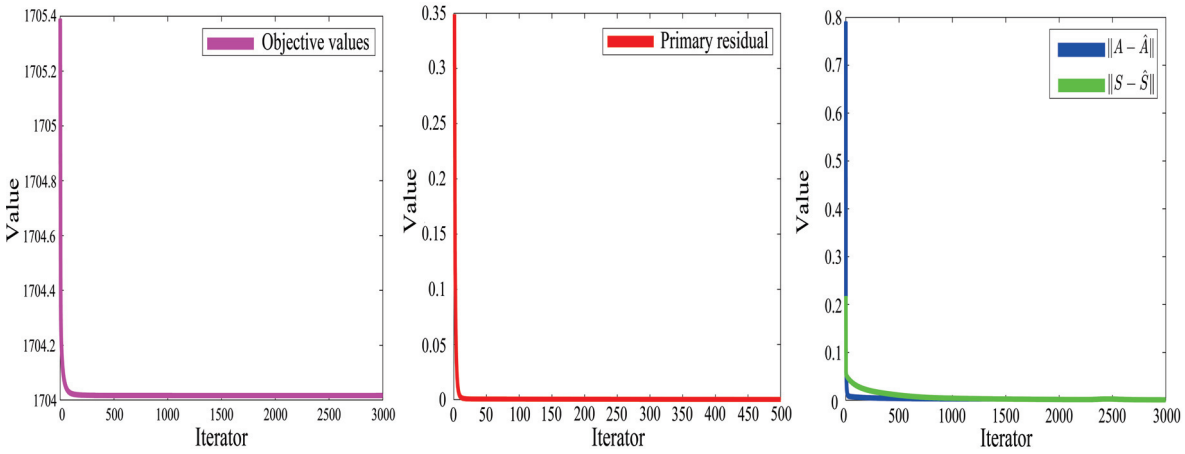
Figure 9. Unmixing results produced by different methods under different mixing levels: (**a**) SAD and (**b**) RMSE.

Experiment 6(Convergence Analysis): This experiment was conducted to analyze the convergence of the proposed algorithm. We set the number of endmembers to 6 and the SNR to 30 dB. Figure 10a,b shows the curves of the target function value and the main residual term changes, respectively. The target function value decreased rapidly after several iterations and finally converged to a stable value; the main residual term also tended to decrease gradually, which was consistent with the expected result. Furthermore, Figure 10c presents the iterative residual curves of the endmembers and abundance matrices. As the number of iterations increased, the residuals of these two matrices gradually approached zero, further validating the good convergence ability of the proposed algorithm.

https://doi.org/10.3390/rs18010155

$$\eta = \frac{1}{\sqrt{B}} \sum_{b} \frac{\sqrt{N} - \|x_b\|_1 / \|x_b\|_2}{\sqrt{N-1}} \tag{38}$$

where x_b represents the given hyperspectral data at band b.



(**a**) (**b**) (**c**)

Figure 10. Convergence curves: (**a**) objective function value, (**b**) residual value of reconstruction and (**c**) endmember and abundance residuals.

Experiment 7 (Truncated Activation Function Analysis): This experiment investigated the effect of the truncation threshold of the activation function (10^{-3} to 10^{-9}) on HU, analyzing the abundance matrix sparsity η, SAD, RMSE, and running time. The results (Table 3) show that decreasing the threshold reduces sparsity, especially between 10^{-3} and 10^{-5}, where the number of zero elements significantly decreases; below 10^{-6}, sparsity stabilizes. SAD and RMSE are optimal at 10^{-5}, indicating the highest unmixing accuracy; a too-large threshold truncates too many abundance values, increasing errors, while a smaller threshold increases running time. These results demonstrate that the truncation threshold significantly affects model performance and requires a tradeoff between accuracy and efficiency, with 10^{-5} being the optimal choice. According to [30], η is defined as follows:

Table 3. Analysis of truncated activation functions (bold font represents the best value).

t	η	SAD	RMSE	Time (s)
10^{-3}	**0.5128**	0.0329	0.0417	**28.94**
10^{-4}	0.6062	0.0276	0.0403	30.08
10^{-5}	0.6102	**0.0267**	**0.0344**	30.21
10^{-6}	0.6162	0.0294	0.0441	32.78
10^{-7}	0.6170	0.0355	0.0447	33.11
10^{-8}	0.6182	0.0368	0.0486	34.82
10^{-9}	0.6185	0.0395	0.0492	35.76

Experiment 8 (Ablation Analysis): To evaluate the effectiveness of truncated robust DNMF, adversarial graph regularization, and Gram sparsity regularization, we designed eight ablation experiments. First, within the DNMF framework, the traditional Frobenius norm reconstruction term was replaced with the $L_{2,1}$ norm to enhance robustness against noise and outliers (39) and (40). On this basis, a truncated activation function was intro-

duced to form the truncated robust DNMF model (41). The loss functions of the three models are as follows:

$$\min_{\boldsymbol{A}_i,\boldsymbol{S}_l} \|\boldsymbol{X} - \boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_l\boldsymbol{S}_l\|_F^2 \tag{39}$$

$$\min_{\boldsymbol{A}_i,\boldsymbol{S}_l} \|\boldsymbol{X} - \boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_l\boldsymbol{S}_l\|_{2,1} \tag{40}$$

$$\min_{\boldsymbol{A}_i,\boldsymbol{S}_l} \|\boldsymbol{X} - \boldsymbol{A}_1\sigma(\boldsymbol{A}_2\sigma(\cdots\sigma(\boldsymbol{A}_l\boldsymbol{S}_l)))\|_{2,1} \tag{41}$$

By establishing a dynamic mechanism for balancing the reward graph and the penalty graph, the proposed dual-graph adversarial regularization module (42) and (43) effectively suppressed the difficulty of distinguishing dissimilar abundances while maintaining the continuity of the space of similar samples. Furthermore, we conducted a systematic comparison between the Gram sparsity regularization proposed in this paper and the traditional $L_{1/2}$ regularization technique (44) and (45) and ultimately formed the model presented in this paper (46).

$$\min_{\boldsymbol{A}_i,\boldsymbol{S}_i} \|\boldsymbol{X} - \boldsymbol{A}_1\sigma(\boldsymbol{A}_2\sigma(\cdots\sigma(\boldsymbol{A}_l\boldsymbol{S}_l)))\|_{2,1} + \alpha\,\mathrm{tr}\left(\boldsymbol{S}_i\boldsymbol{L}_R{\boldsymbol{S}_i}^T\right) \tag{42}$$

$$\min_{\boldsymbol{A}_i,\boldsymbol{S}_i} \|\boldsymbol{X} - \boldsymbol{A}_1\sigma(\boldsymbol{A}_2\sigma(\cdots\sigma(\boldsymbol{A}_l\boldsymbol{S}_l)))\|_{2,1} + \alpha\,\mathrm{tr}\left(\boldsymbol{S}_i\boldsymbol{L}_R{\boldsymbol{S}_i}^T\right) - \beta\,\mathrm{tr}\left(\boldsymbol{S}_i\boldsymbol{L}_P{\boldsymbol{S}_i}^T\right) \tag{43}$$

$$\min_{\boldsymbol{A}_i,\boldsymbol{S}_i} \|\boldsymbol{X} - \boldsymbol{A}_1\sigma(\boldsymbol{A}_2\sigma(\cdots\sigma(\boldsymbol{A}_l\boldsymbol{S}_l)))\|_{2,1} + \lambda\,\|\boldsymbol{S}_i\|_{\frac{1}{2}} \tag{44}$$

$$\min_{\boldsymbol{A}_i,\boldsymbol{S}_i} \|\boldsymbol{X} - \boldsymbol{A}_1\sigma(\boldsymbol{A}_2\sigma(\cdots\sigma(\boldsymbol{A}_l\boldsymbol{S}_l)))\|_{2,1} + \gamma\left(\mathrm{tr}\left({\boldsymbol{S}_i}^T\boldsymbol{S}_i\boldsymbol{1}_N\right) - \mathrm{tr}\left({\boldsymbol{S}_i}^T\boldsymbol{S}_i\right)\right) \tag{45}$$

$$\min_{\boldsymbol{A}_i,\boldsymbol{S}_i} \|\boldsymbol{X} - \boldsymbol{A}_1\sigma(\boldsymbol{A}_2\sigma(\cdots\sigma(\boldsymbol{A}_l\boldsymbol{S}_l)))\|_{2,1} + \alpha\,\mathrm{tr}\left(\boldsymbol{S}_i\boldsymbol{L}_R{\boldsymbol{S}_i}^T\right) - \beta\,\mathrm{tr}\left(\boldsymbol{S}_i\boldsymbol{L}_P{\boldsymbol{S}_i}^T\right) + \gamma\left(\mathrm{tr}\left({\boldsymbol{S}_i}^T\boldsymbol{S}_i1_N\right) - \mathrm{tr}\left({\boldsymbol{S}_i}^T\boldsymbol{S}_i\right)\right) \tag{46}$$

To evaluate the performance of each model, experiments were conducted on synthetic data with six endmembers and an SNR of 30 dB (Figure 11). The results show that replacing the traditional Frobenius norm with the $L_{2,1}$ norm effectively suppresses interference from noise and outliers, significantly reducing SAD and RMSE. The robust DNMF model with the truncated activation function not only preserves the non-negativity of the abundance matrix but also truncates elements smaller than 10^{-5} to 0, enhancing sparsity and improving computational efficiency. On this basis, the dual-graph adversarial regularization module (43) outperforms the single-graph model (42) by simultaneously constructing a local similarity graph and a non-neighborhood repulsion graph, better capturing detailed features. Finally, the proposed Gram sparsity constraint achieves the best results compared to the traditional $L_{1/2}$ regularization, further improving unmixing performance.

Experiment 9(Outlier Robustness Analysis): To verify the algorithm's robustness against outliers, a two-dimensional experiment was designed. In the spectral dimension, abnormal pixels with full bands, 1/2 bands, and 1/3 bands (random negative values) were constructed (Figure 12a,b) to test adaptability to varying spectral anomalies. In the spatial dimension, with 1/3 abnormal bands fixed, the number of abnormal pixels (3, 5, and 10) was increased to evaluate spatial sensitivity (Figure 12c,d). Results show that in single-pixel tests, the algorithm maintains optimal endmember extraction (SAD) and abundance inversion (RMSE); in multi-pixel tests, thanks to the synergy of the activation function, $L_{2,1}$ norm constraint, and regularization term, the algorithm significantly outperforms comparison methods, demonstrating excellent robustness.

https://doi.org/10.3390/rs18010155

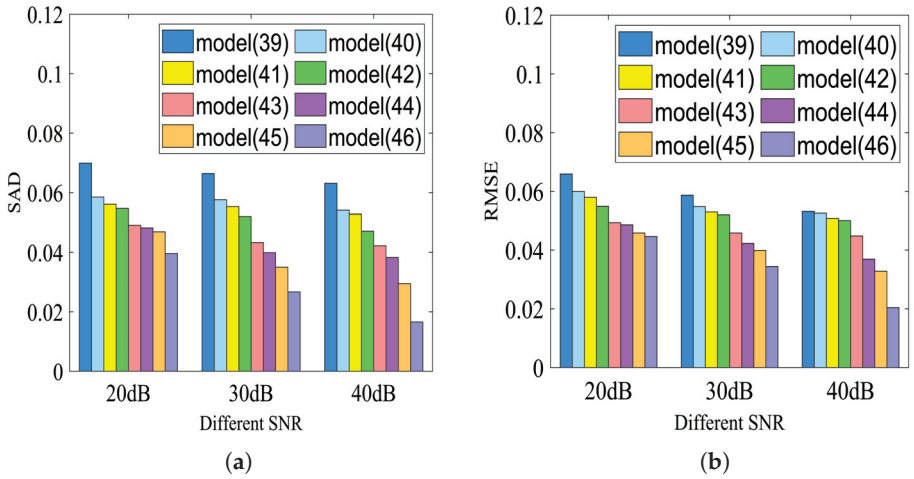


Figure 11. Results of ablation experiments conducted under different SNRs: (**a**) SAD and (**b**) RMSE.

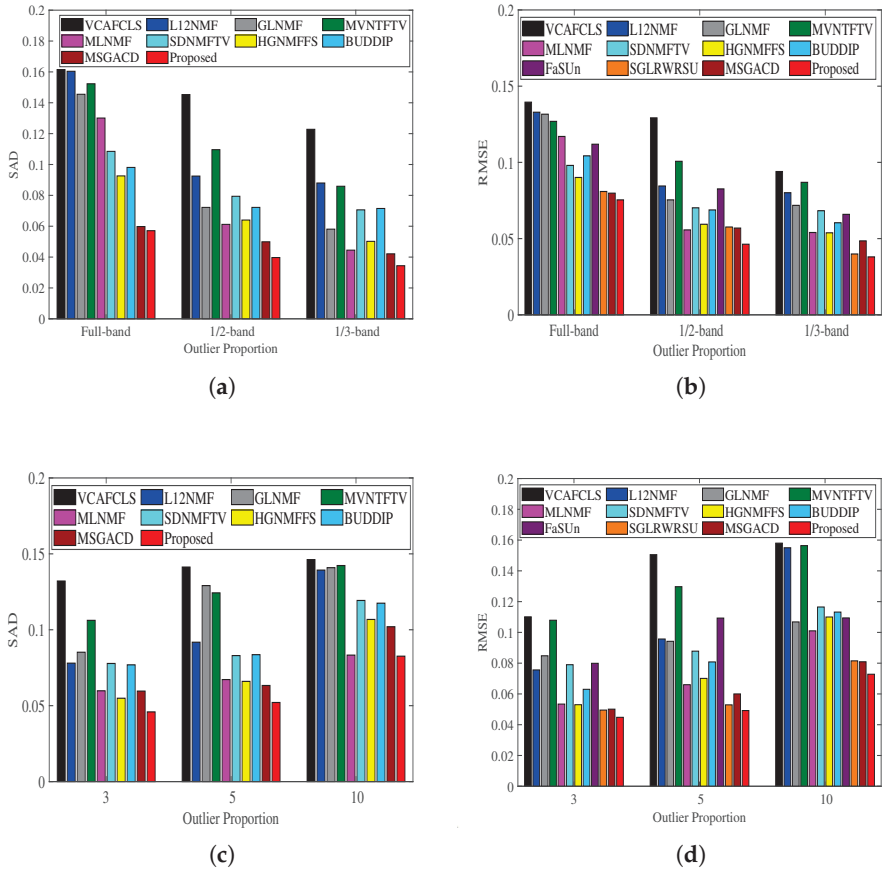


Figure 12. Outlier robustness analysis. (**a**) SAD under different spectral anomaly scenarios; (**b**) RMSE under different spectral anomaly scenarios; (**c**) SAD with varying numbers of anomalous pixels under the fixed 1/3 anomalous-band condition; (**d**) RMSE with varying numbers of anomalous pixels under the fixed 1/3 anomalous-band condition.

4.1.2. Experiments on SC2

This set of experiments evaluated the unmixing performance of the proposed algorithm on SC2 under SNRs of 20, 30, and 40 dB. As shown in Table 4, it achieves the best performance under most noise conditions. The abundance maps at 30 dB (Figure 13)

https://doi.org/10.3390/rs18010155

show that the algorithm preserves fine spatial structures in transition zones. These results indicate improved spectral separation and spatial resolution under varying noise levels.

Table 4. Results of SAD and RMSE under different noise levels on SC2 (best values in bold).

SNR	Metric		VCAFCLS [2005]	L1/2NMF [2011]	GLNMF [2013]	MVNTFTV [2019]	MLNMF [2015]	SDNMFTV [2018]	HGNMFFS [2023]	BUDDIP [2024]	FaSUn [2024]	SGLRWRSU [2024]	MSGACD [2025]	Proposed
SNR = 20	SAD	Mean	0.1041	0.0284	0.0659	0.0644	0.0523	0.0369	0.0253	0.0277	/	/	0.0347	**0.0189**
		Std	±2.95%	±1.97%	±2.06%	±3.72%	±2.45%	±1.03%	±1.39%	±1.35%	/	/	±2.56%	±1.01%
	RMSE	Mean	0.1098	0.0300	0.0512	0.0296	0.0388	0.0357	0.0289	0.1005	0.1090	**0.0244**	0.0286	0.0269
		Std	±2.70%	±1.98%	±1.94%	±2.44%	±2.63%	±1.20%	±0.81%	±1.45%	±1.96%	±1.07%	±1.95%	±0.57%
SNR = 30	SAD	Mean	0.0628	0.0272	0.0624	0.0367	0.0423	0.0353	0.0191	0.0058	/	/	0.0085	**0.0057**
		Std	±2.85%	±1.13%	±2.76%	±2.75%	±1.80%	±0.87%	±1.07%	±0.85%	/	/	±2.01%	±1.33%
	RMSE	Mean	0.1014	0.0217	0.0440	0.0231	0.0316	0.0218	0.0183	0.0807	0.1012	0.0183	0.0193	**0.0176**
		Std	±2.18%	±1.94%	±1.55%	±1.54%	±1.84%	±1.01%	±0.83%	±1.05%	±0.95%	±0.71%	±1.92%	±0.80%
SNR = 40	SAD	Mean	0.0620	0.0254	0.0619	0.0123	0.0349	0.0109	0.0058	0.0020	/	/	0.0023	**0.0018**
		Std	±3.23%	±2.86%	±0.91%	±2.22%	±1.59%	±0.77%	±0.93%	±0.83%	/	/	±2.01%	±0.81%
	RMSE	Mean	0.1006	0.0158	0.0436	0.0165	0.0111	0.0076	0.0094	0.0809	0.1005	0.0073	0.0078	**0.0065**
		Std	±2.93%	±2.90%	±0.87%	±1.44%	±1.35%	±0.97%	±0.70%	±0.81%	±0.79%	±0.90%	±0.65%	±0.42%

Note: FaSUn and SGLRWRSU are sparse regression methods, no endmember estimation needed.

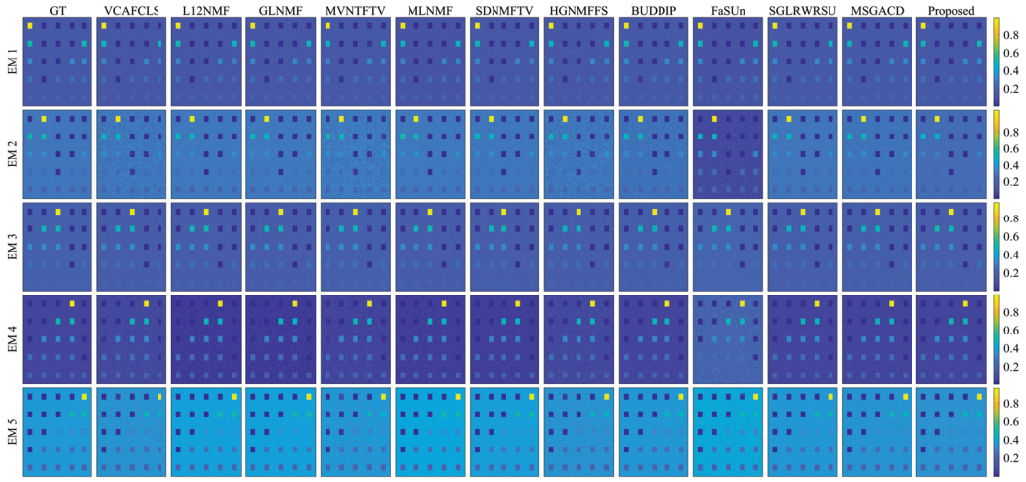


Figure 13. Abundance maps produced by all algorithms for SC2.

4.2. Experiments Conducted on Real Datasets

4.2.1. Samson Dataset

This experiment validated the proposed algorithm on the Samson dataset (95 × 95 pixels, 156 bands, containing trees, soil, and water). As shown in Figure 14, the extracted endmember spectra closely match the reference spectra. Figure 15 and Table 5 present the abundance maps and SAD values, indicating that the algorithm performs well on most materials and achieves the best overall performance.

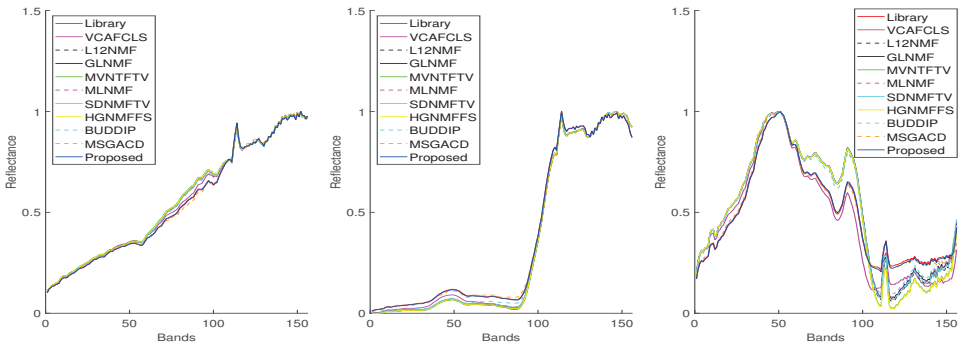


Figure 14. Endmember spectra from the Samson dataset (from left to right: soil, trees, and water).

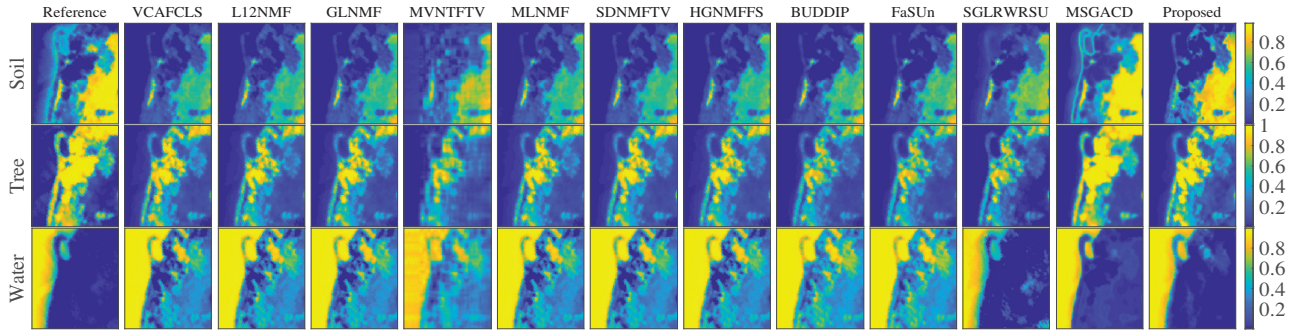


Figure 15. Abundance maps estimated by different algorithms for the Samson dataset.

Table 5. SAD values attained by different algorithms on the Samson dataset (best values in bold).

Methods		VCAFCLS [2005]	L1/2NMF [2011]	GLNMF [2013]	MVNTFTV [2019]	MLNMF [2015]	SDNMFTV [2018]	HGNMFFS [2023]	BUDDIP [2024]	MSGACD [2025]	Proposed
Tree	Mean	0.0407	0.0359	0.0316	0.0377	0.0301	0.0307	0.0357	0.0293	0.0396	**0.0077**
	Std	±5.94%	±11.21%	±0.11%	±0.65%	±0.08%	±7.98%	±8.27%	±14.88%	±7.13%	±11.95%
Soil	Mean	0.0895	0.0611	0.0559	0.0632	0.0556	0.0549	0.0608	0.0466	0.0552	**0.0394**
	Std	±0.13%	±4.81%	±0.21%	±0.18%	±0.01%	±3.54%	±3.13%	±6.09%	±1.09%	±5.28%
Water	Mean	0.1599	0.1503	0.1409	0.1787	0.1406	0.1474	0.1822	0.1235	**0.1126**	0.1224
	Std	±0.10%	±1.52%	±1.04%	±0.36%	±0.37%	±1.17%	±0.85%	±1.40%	±0.95%	±2.86%
Overall	Mean	0.0967	0.0824	0.0761	0.0932	0.0754	0.0777	0.0929	0.0665	0.0691	**0.0565**
	Std	±1.95%	±4.83%	±0.45%	±0.32%	±0.15%	±3.91%	±3.89%	±6.52%	±3.89%	±5.04%

Note: FaSUn and SGLRWRSU are sparse regression methods, no endmember estimation needed.

4.2.2. Jasper Ridge Dataset

This experiment was conducted on a 100 × 100-pixel subregion of the Jasper dataset (198 bands), containing four main land-cover types: tree, water, soil, and road. The results show that the endmember spectra estimated by the proposed algorithm closely match the USGS reference spectra (Figure 16), and the abundance maps (Figure 17) demonstrate strong spatial resolution. Table 6 indicates that the proposed method outperforms other algorithms in most land-cover categories and in average SAD, confirming its advantage in hyperspectral unmixing tasks.

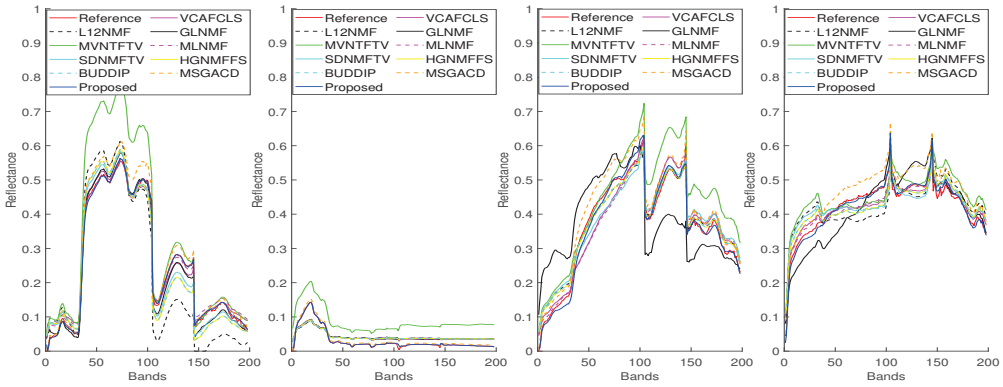


Figure 16. Endmember spectra from Jasper dataset (from left to right: soil, tree, water and road).

4.2.3. AVIRIS Cuprite Dataset

On the AVIRIS Cuprite dataset, a 250 × 191 × 188 subregion was selected for analysis, with the number of endmembers set to 12. (Figure 18) shows the reference endmembers and the endmember spectra estimated by each algorithm, indicating that the endmembers extracted by the proposed method are highly consistent with the references. (Figure 19) presents the abundance maps, where brighter pixels represent higher abundances. Table 7

https://doi.org/10.3390/rs18010155

lists the average SAD values for each algorithm, showing that the proposed method outperforms the comparison methods for most mineral categories and overall.

Table 6. SAD values attained by different algorithms on the Jasper dataset (best values in bold).

Methods		VCAFCLS [2005]	L1/2NMF [2011]	GLNMF [2013]	MVNTFTV [2019]	MLNMF [2015]	SDNMFTV [2018]	HGNMFFS [2023]	BUDDIP [2024]	MSGACD [2025]	Proposed
Tree	Mean	0.4889	0.4671	0.1073	0.1636	0.0849	0.4585	0.4642	0.1592	0.1016	**0.0774**
	Std	±1.28%	±2.68%	±2.62%	±2.82%	±4.99%	±0.76%	±2.41%	±3.03%	±0.97%	±1.62%
Water	Mean	0.1016	0.0832	0.4267	0.0764	0.0901	0.0650	0.0603	0.1227	0.0902	**0.0488**
	Std	±0.55%	±0.83%	±0.77%	±17.70%	±1.89%	±2.29%	±0.60%	±2.11%	±1.63%	±0.51%
Soil	Mean	0.1902	0.0715	0.0641	0.1537	**0.0560**	0.0700	0.0846	0.1147	0.1059	0.1158
	Std	±11.46%	±20.36%	±12.46%	±30.76%	±23.10%	±23.97%	±20.00%	±18.12%	±10.60%	±18.33%
Road	Mean	0.2492	0.0866	0.0676	0.3953	0.4473	0.1189	0.0829	0.2185	0.0961	**0.0800**
	Std	±6.70%	±2.10%	±1.95%	±7.68%	±4.45%	±10.61%	±20.18%	±3.96%	±2.35%	±2.36%
Overall	Mean	0.2575	0.1771	0.1664	0.1973	0.1696	0.1781	0.1730	0.1538	0.0985	**0.0805**
	Std	±4.46%	±5.28%	±3.22%	±6.79%	±6.03%	±5.44%	±5.53%	±5.07%	±4.61%	±4.67%

Note: FaSUn and SGLRWRSU are sparse regression methods, no endmember estimation needed.

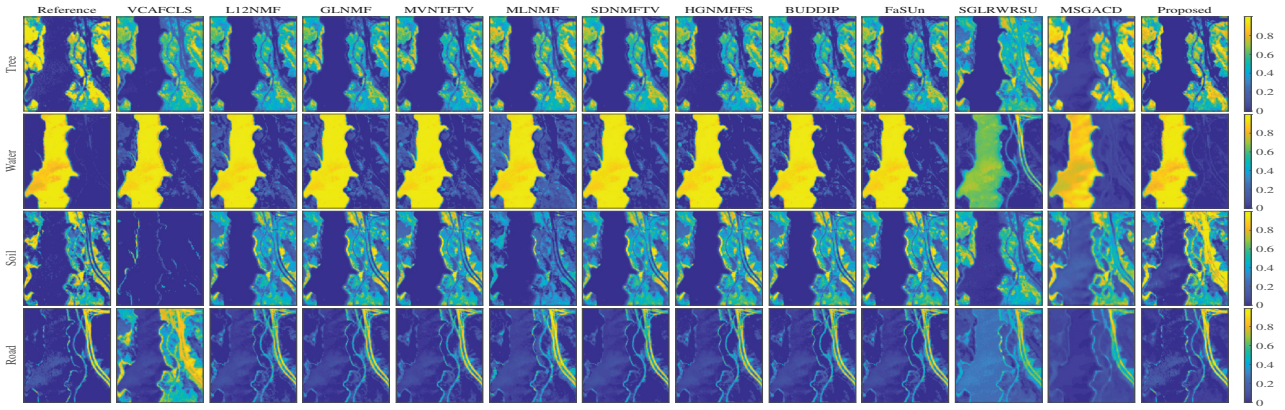


Figure 17. Abundance maps estimated by different algorithms for the Jasper dataset.

Table 7. SAD values attained by different algorithms on the Cuprite dataset (best values in bold).

Methods		VCAFCLS [2005]	L1/2NMF [2011]	GLNMF [2013]	MVNTFTV [2019]	MLNMF [2015]	SDNMFTV [2018]	HGNMFFS [2023]	BUDDIP [2024]	MSGACD [2025]	Proposed
Chalcedony	Mean	0.1429	**0.0464**	0.1198	0.0777	0.2797	0.0785	0.0831	0.3162	0.1193	0.1183
	Std	±1.59%	±1.34%	±2.17%	±2.31%	±4.01%	±2.45%	±2.48%	±1.40%	±5.08%	±1.67%
Andradite	Mean	**0.0590**	0.1694	0.0666	0.5293	0.0595	0.0690	0.0768	0.1106	0.0624	0.0729
	Std	±1.23%	±1.58%	±0.50%	±8.20%	±3.87%	±1.08%	±2.07%	±2.10%	±10.05%	±1.21%
Buddingtonite	Mean	0.1679	**0.0600**	0.1163	0.2709	0.0737	0.0692	0.0914	0.1969	0.0867	0.0856
	Std	±0.81%	±1.40%	±1.86%	±25.46%	±1.45%	±1.82%	±0.29%	±1.91%	±0.96%	±0.94%
Kaolin1	Mean	0.1398	0.0810	0.0976	0.0698	0.0958	0.0950	0.1098	0.0698	0.0738	**0.0684**
	Std	±0.98%	±1.55%	±1.26%	±18.20%	±2.10%	±1.34%	±1.16%	±1.49%	±1.14%	±0.90%
Kaolin2	Mean	0.2222	0.0812	0.2265	0.0402	0.0746	0.1186	0.0699	0.0984	0.1153	**0.0401**
	Std	±2.28%	±2.97%	±2.38%	±10.64%	±4.67%	±2.87%	±1.52%	±3.33%	±3.69%	±2.09%
Muscovite	Mean	0.1907	0.0957	0.1340	0.1250	0.1203	0.0912	0.1771	0.2057	0.1278	**0.0901**
	Std	±1.09%	±1.36%	±0.49%	±15.93%	±1.03%	±1.36%	±1.20%	±0.51%	±0.66%	±1.24%
Dumortierite	Mean	0.1926	0.0808	0.0755	**0.0515**	0.0743	0.1608	0.1098	0.0836	0.0791	0.1039
	Std	±1.50%	±0.85%	±0.58%	±2.49%	±3.55%	±2.86%	±3.01%	±0.77%	±2.29%	±0.88%
Montmorillonite	Mean	0.1839	0.0928	0.0598	0.0634	0.1200	**0.0481**	0.0638	0.1492	0.0914	0.0759
	Std	±1.39%	±1.02%	±1.84%	±27.13%	±1.19%	±0.75%	±1.17%	±3.27%	±1.05%	±0.69%
Nontronite	Mean	0.1415	0.0577	0.0587	0.0383	0.0981	0.1062	0.0923	0.2953	0.0975	**0.0346**
	Std	±1.37%	±0.16%	±0.34%	±35.93%	±1.75%	±0.37%	±0.21%	±0.93%	±1.03%	±0.32%
Pyrope	Mean	0.2050	0.1530	0.0985	0.0775	0.1444	0.0833	0.1589	0.1415	0.0878	**0.0625**
	Std	±1.02%	±1.51%	±2.12%	±24.84%	±6.53%	±1.70%	±2.12%	±2.09%	±2.69%	±1.94%
Sphene	Mean	0.1537	0.0832	0.0810	0.1119	0.1857	0.2606	0.0598	0.1374	0.0624	**0.0584**
	Std	±1.07%	±1.15%	±1.98%	±10.06%	±4.36%	±2.13%	±0.33%	±2.57%	±0.92%	±1.30%
Alunite	Mean	0.1956	0.2707	0.1116	**0.0453**	0.0530	0.0867	0.1110	0.0655	0.0913	0.0730
	Std	±2.13%	±1.69%	±2.05%	±4.12%	±3.83%	±2.15%	±0.65%	±3.69%	±6.09%	±1.92%
Overall	Mean	0.1662	0.1060	0.1039	0.1251	0.1149	0.1056	0.1003	0.1558	0.0912	**0.0736**
	Std	±0.23%	±0.36%	±0.33%	±3.87%	±0.83%	±0.37%	±0.20%	±0.71%	±0.79%	±0.30%

Note: FaSUn and SGLRWRSU are sparse regression methods, no endmember estimation needed.

https://doi.org/10.3390/rs18010155

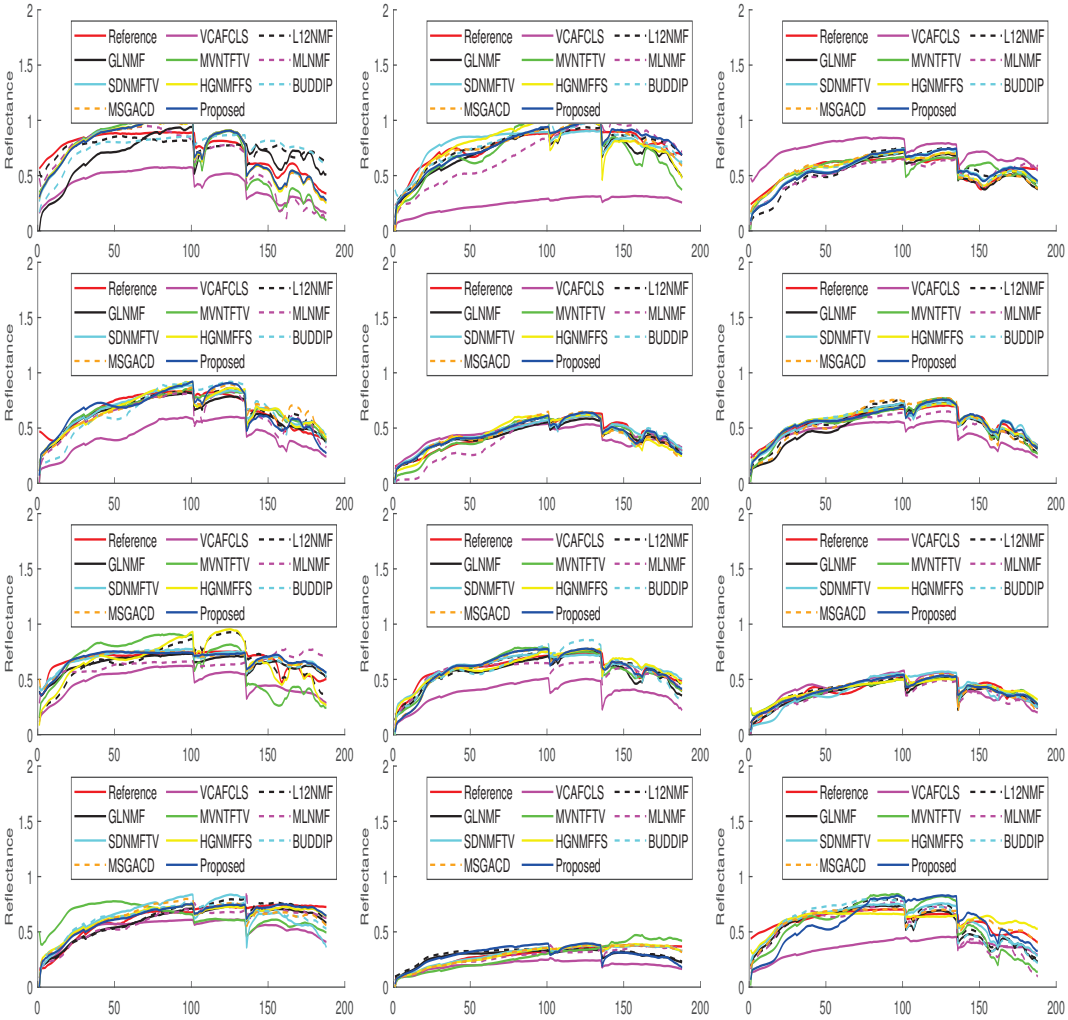


Figure 18. Endmember spectra from the AVIRIS Cuprite dataset.

4.2.4. Urban Dataset

The Urban dataset contains 307 × 307 pixels and 162 effective bands, covering six types of ground features: asphalt road, grass, roof1, tree, roof2, and concrete road. As shown in Figure 20, the endmember spectra estimated by the proposed algorithm closely match the reference endmembers. Figure 21 shows that the abundance maps are consistent with the ground-truth distribution, and Table 8 indicates that it achieves the lowest SAD values.

Table 8. SAD values attained by different algorithms on the Urban dataset (best values in bold).

Methods		VCAFCLS [2005]	L1/2NMF [2011]	GLNMF [2013]	MVNTFTV [2019]	MLNMF [2015]	SDNMFTV [2018]	HGNMFFS [2023]	BUDDIP [2024]	MSGACD [2025]	Proposed
Asphalt Road	Mean	0.1664	0.1652	0.1584	0.0360	0.1033	0.0820	0.1732	0.1356	0.1216	**0.0358**
	Std	±15.76%	±23.86%	±1.56%	3.11%	14.57%	±25.07%	±19.50%	±2.28%	±11.26%	±18.23%
Grass	Mean	0.0943	0.0827	0.0826	0.0367	0.0783	0.1587	0.0821	0.0823	0.0824	**0.0523**
	Std	±6.33%	±9.51%	±7.60%	±2.05%	±29.88%	±54.81%	19.77%	±5.55%	09.25%	±4.05%
Roof1	Mean	0.1646	0.1586	0.1604	**0.0433**	0.1115	0.0480	0.1630	0.1556	0.1509	0.0811
	Std	±9.07%	±10.08%	±0.17%	±3.07%	±16.75%	±20.81%	±10.78%	±7.21%	±8.15%	±7.89%
Tree	Mean	0.0820	0.0451	0.0445	0.2029	0.1357	0.1531	0.0466	0.0442	0.0441	**0.0437**
	Std	±9.10%	±21.96%	±24.37%	±35.36%	±9.66%	±50.52%	±16.63%	±4.50%	±13.72%	±5.73%
Roof2	Mean	0.0890	0.0740	0.0798	0.2220	0.0455	0.0453	0.0836	0.0799	0.0807	**0.0446**
	Std	±21.21%	±25.11%	±29.30%	±4.50%	±14.14%	±23.59%	±17.96%	±24.52%	±10.02%	±14.82%
Concrete Road	Mean	0.1124	0.0552	0.0514	0.0877	0.1066	0.0822	0.0536	0.0476	0.0459	**0.0471**
	Std	±3.11%	±28.54%	±0.30%	±11.35%	±33.00%	±36.24%	±21.34%	±8.12%	±17.39%	±0.95%
Overall	Mean	0.1181	0.0968	0.0962	0.1048	0.0968	0.0949	0.1004	0.0909	0.0876	**0.0508**
	Std	±3.20%	±7.97%	±3.89%	±3.94%	±6.28%	±8.98%	±4.61%	±3.81%	±3.95%	±3.33%

Note: FaSUn and SGLRWRSU are sparse regression methods, no endmember estimation needed.

https://doi.org/10.3390/rs18010155

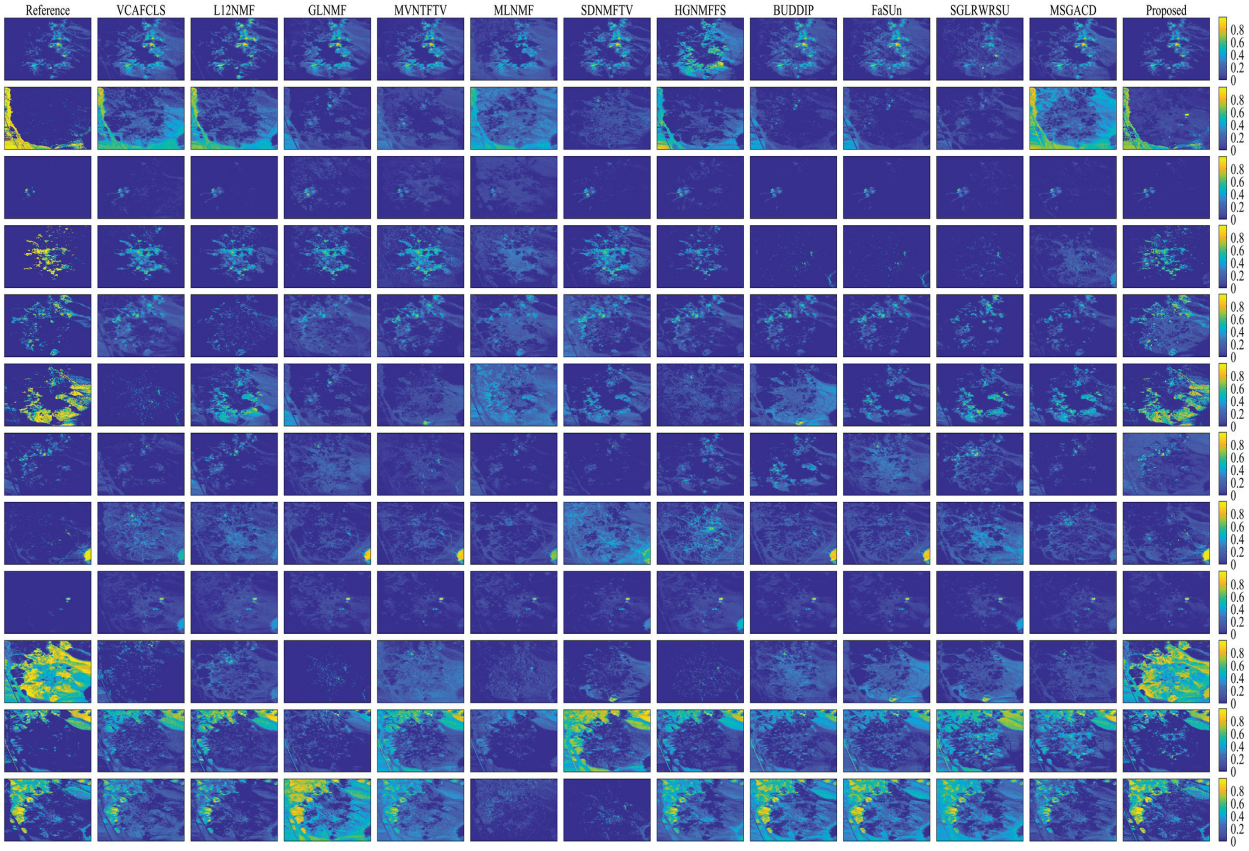


Figure 19. Abundancemaps estimated by different algorithms for the AVIRIS Cuprite dataset.

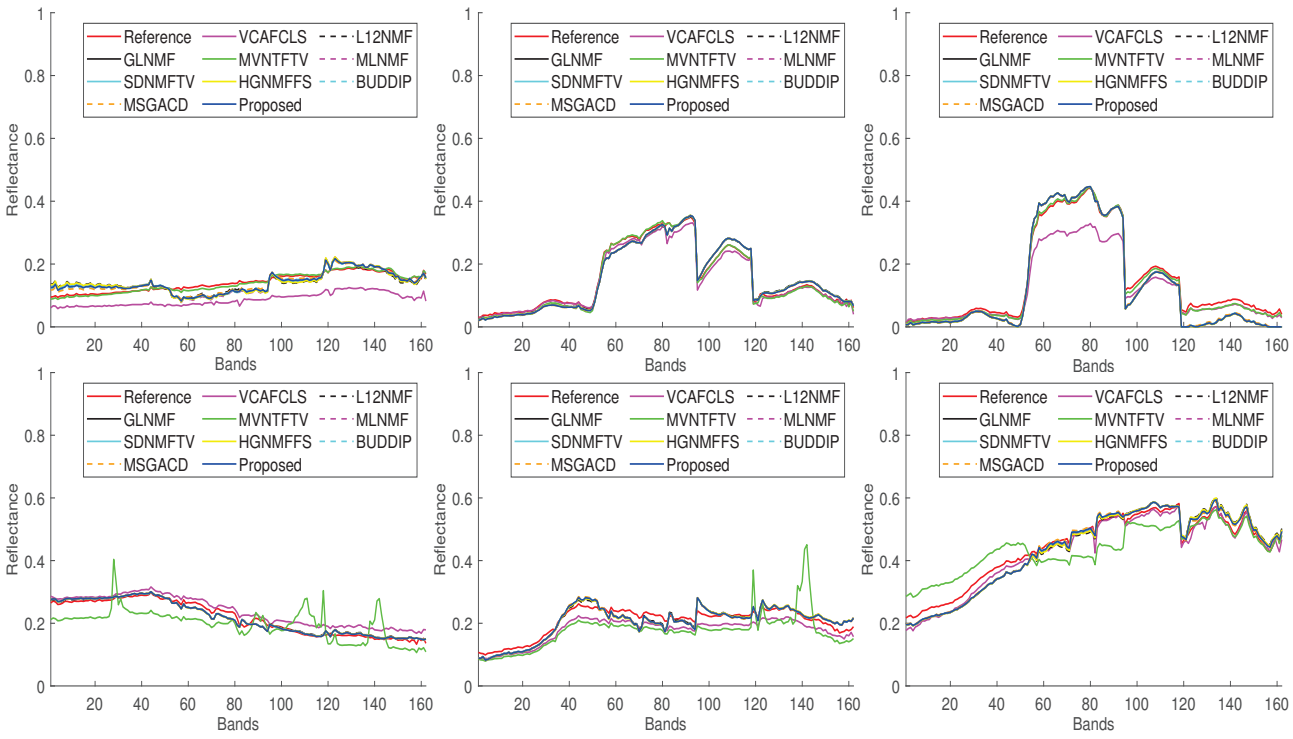


Figure 20. Comparison among the endmember spectra extracted by different algorithms from the Urban dataset.

https://doi.org/10.3390/rs18010155

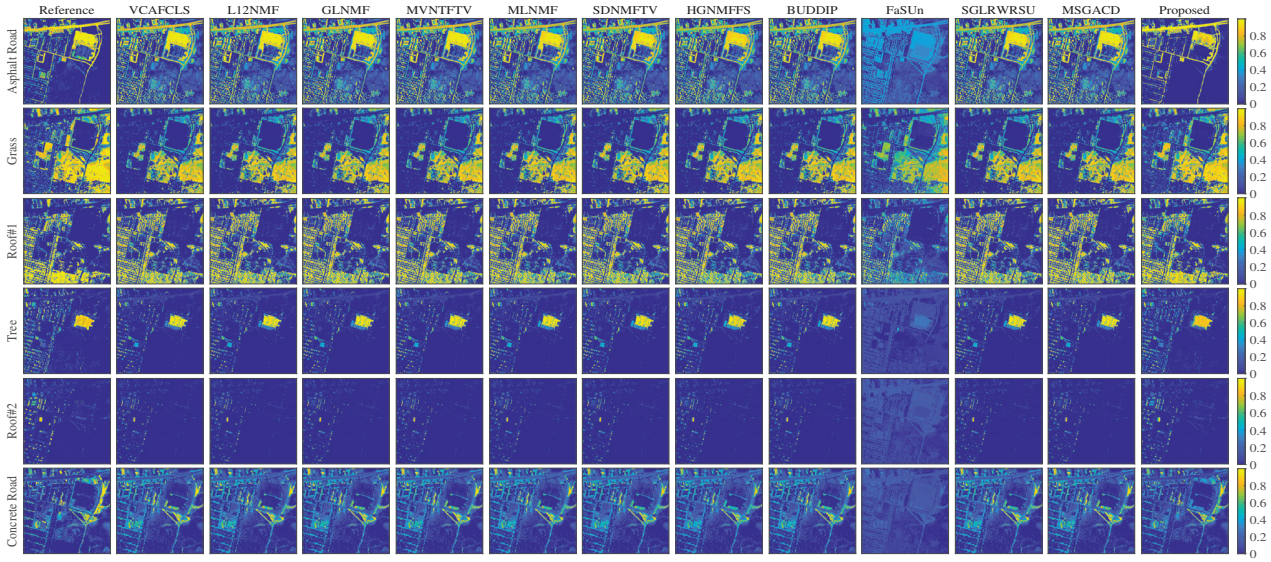


Figure 21. Abundance maps estimated by different algorithms for the Urban dataset.

5. Conclusions

This study proposed a deep non-negative matrix factorization method (DNMF-AG) for hyperspectral unmixing, incorporating adversarial graph regularization and Gram sparsity constraints. The adversarial graph integrates similarity and dissimilarity graphs to enhance intraclass continuity and interclass separability, while the Gram inner-product penalty constrains the abundance matrix to encourage structured sparsity and improve endmember distinguishability. These two regularizations are jointly embedded into the DNMF framework to enhance robustness and unmixing accuracy. A truncated activation function is introduced during iterative optimization to suppress low-amplitude noise, promote sparsity, and improve computational efficiency.

In future work, we will further explore the use of more spatial–spectral prior information to enhance the generalization ability of the proposed model in complex scenarios.

Author Contributions: Methodology, K.Q., X.L., and W.B.; Writing—original draft, X.L.; Writing—review and editing, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 62461001), Natural Science Project Fund for High-level Talents of North Minzu University (Grant No.: 2025BG239), and in part by the North Minzu University Postgraduate Innovation Project (Grant No. YCX24365).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bhargava, A.; Sachdeva, A.; Sharma, K.; Alsharif, M.H.; Uthansakul, P.; Uthansakul, M. Hyperspectral imaging and its applications: A review. *Heliyon* **2024**, *10*, e33208. [CrossRef]
2. Sun, L.; Wang, X.; Zheng, Y.; Wu, Z.; Fu, L. Multiscale 3-D–2-D Mixed CNN and Lightweight Attention-Free Transformer for Hyperspectral and LiDAR Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 2100116. [CrossRef]
3. Cheng, S.; Chan, R.; Du, A. CACFTNet: A Hybrid Cov-Attention and Cross-Layer Fusion Transformer Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–17. [CrossRef]
4. He, X.; Tang, C.; Liu, X.; Zhang, W.; Sun, K.; Xu, J. Object Detection in Hyperspectral Image via Unified Spectral–Spatial Feature Aggregation. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5521213. [CrossRef]

5. Rasti, B.; Zouaoui, A.; Mairal, J.; Chanussot, J. Image Processing and Machine Learning for Hyperspectral Unmixing: An Overview and the HySUPP Python Package. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5517631. [CrossRef]
6. Wei, J.; Wang, X. An Overview on Linear Unmixing of Hyperspectral Data. *Math. Probl. Eng.* **2020**, *2020*, 3735403. [CrossRef]
7. Li, H.C.; Feng, X.R.; Wang, R.; Gao, L.; Du, Q. Superpixel-Based Low-Rank Tensor Factorization for Blind Nonlinear Hyperspectral Unmixing. *IEEE Sens. J.* **2024**, *24*, 13055–13072. [CrossRef]
8. Su, H.; Jia, C.; Zheng, P.; Du, Q. Superpixel-Based Weighted Collaborative Sparse Regression and Reweighted Low-Rank Representation for Hyperspectral Image Unmixing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 393–408. [CrossRef]
9. Shen, X.; Chen, L.; Liu, H.; Su, X.; Wei, W.; Zhu, X.; Zhou, X. Efficient Hyperspectral Sparse Regression Unmixing With Multilayers. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5522614. [CrossRef]
10. Iordache, M.D.; Bioucas-Dias, J.M.; Plaza, A. Sparse Unmixing of Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2014–2039. [CrossRef]
11. Iordache, M.D.; Bioucas-Dias, J.M.; Plaza, A. Total Variation Spatial Regularization for Sparse Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4484–4502. [CrossRef]
12. Liang, Y.; Zheng, H.; Yang, G.; Du, Q.; Su, H. Superpixel-Based Weighted Sparse Regression and Spectral Similarity Constrained for Hyperspectral Unmixing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 6825–6842. [CrossRef]
13. Ayres, L.C.; Borsoi, R.A.; Bermudez, J.C.M.; de Almeida, S.J.M. A Generalized Multiscale Bundle-Based Hyperspectral Sparse Unmixing Algorithm. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 5502505. [CrossRef]
14. Xu, C. Spectral Weighted Sparse Unmixing Based on Adaptive Total Variation and Low-Rank Constraints. *Sci. Rep.* **2024**, *14*, 23705. [CrossRef]
15. Qu, K.; Luo, F.; Wang, H.; Bao, W. A New Fast Sparse Unmixing Algorithm Based on Adaptive Spectral Library Pruning and Nesterov Optimization. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 6134–6151. [CrossRef]
16. Boardman, J.W.; Kruscl, F.A.; Grccn, R.O. Mapping target signatures via partial unmixing of AVIRIS data. In Proceedings of the Fifth JPL Airborne Earth Science Workshop, Pasadena, CA, USA, 23–26 January 1995.
17. Li, H.C. An Algorithm for Fast Spectral Endmember Determination in Hyperspectral Data. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 2689–2692.
18. Nascimento, J.; Dias, J. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 898–910. [CrossRef]
19. Zhuang, L.; Lin, C.H.; Figueiredo, M.A.T.; Bioucas-Dias, J.M. Regularization Parameter Selection in Minimum Volume Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9858–9877. [CrossRef]
20. Feng, X.R.; Li, H.C.; Wang, R.; Du, Q.; Jia, X.; Plaza, A. Hyperspectral Unmixing Based on Nonnegative Matrix Factorization: A Comprehensive Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4414–4436. [CrossRef]
21. Li, X.; Zhang, X.; Yuan, Y.; Dong, Y. Adaptive Relationship Preserving Sparse NMF for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5504516. [CrossRef]
22. Dong, L.; Lu, X.; Liu, G.; Yuan, Y. A Novel NMF Guided for Hyperspectral Unmixing From Incomplete and Noisy Data. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5513515. [CrossRef]
23. Qu, K.; Li, Z. A Fast Sparse NMF Optimization Algorithm for Hyperspectral Unmixing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 1885–1902. [CrossRef]
24. Chetia, G.S. Hyperspectral Unmixing for Highly Correlated Endmembers using Scaled Endmembers and Abundance Sparsity Constraint NMF. In Proceedings of the 2024 6th International Conference on Energy, Power and Environment (ICEPE), Shillong, India, 20–22 June 2024; pp. 1–6.
25. Jia, S.; Qian, Y.-T.; Ji, Z. Nonnegative matrix factorization with piecewise smoothness constraint for hyperspectral unmixing. In Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition, Hong Kong, China, 30–31 August 2008; Volume 2, pp. 815–820.
26. Li, D.; Zhang, X.; Zhang, Y.; Liang, L.; Chen, X.; Jia, L. Superpixel-guided manifold sparse nonnegative matrix factorization for hyperspectral unmixing. *J. Appl. Remote Sens.* **2025**, *19*, 20. [CrossRef]
27. Qian, Y.; Jia, S.; Zhou, J.; Robles-Kelly, A. Hyperspectral Unmixing via $L_{1/2}$ Sparsity-Constrained Nonnegative Matrix Factorization. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4282–4297. [CrossRef]
28. Yuan, Q.; Yan, L.; Li, J.; Zhang, L. Remote sensing image super-resolution via regional spatially adaptive total variation model. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 3073–3076.
29. Qu, K.; Bao, W. Multiple-Priors Ensemble Constrained Nonnegative Matrix Factorization for Spectral Unmixing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 963–975. [CrossRef]
30. Liu, X.Y.; Gong, X.F.; Wang, L.; Feng, W.; Lin, Q.H. A Parametric Non-Negative Coupled Canonical Polyadic Decomposition Algorithm for Hyperspectral Super-Resolution. In Proceedings of the ICASSP 2025—2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 6–11 April 2025; pp. 1–5.

31. Jimena, G.M.; De Ketelaere, B.; Saeys, W. Shared subspace learning via partial Tucker decomposition for hyperspectral image classification. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2025**, *343*, 126584. [CrossRef] [PubMed]
32. Jiang, X.; Sun, L.; Lin, P. Local Sparsity Blocks and Tensor Low Rank Regularized Sparse Unmixing. In Proceedings of the 2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Athens, Greece, 31 October–2 November 2023; pp. 1–5.
33. Xiong, F.; Qian, Y.; Zhou, J.; Tang, Y.Y. Hyperspectral Unmixing via Total Variation Regularized Nonnegative Tensor Factorization. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2341–2357. [CrossRef]
34. Li, H.; Xiong, X.; Liu, C.; Ma, Y.; Zeng, S.; Li, Y. SFFNet: Staged Feature Fusion Network of Connecting Convolutional Neural Networks and Graph Convolutional Neural Networks for Hyperspectral Image Classification. *Appl. Sci.* **2024**, *14*, 2327. [CrossRef]
35. Dong, H.; Zhang, X.; Meng, H.; Jiao, L. A multistage graph-based autoencoder network with global-local features for hyperspectral unmixing. *Int. J. Remote Sens.* **2025**, *46*, 3709–3735. [CrossRef]
36. Rajabi, R.; Ghassemian, H. Spectral Unmixing of Hyperspectral Imagery Using Multilayer NMF. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 38–42. [CrossRef]
37. Feng, X.R.; Li, H.C.; Li, J.; Du, Q.; Plaza, A.; Emery, W.J. Hyperspectral Unmixing Using Sparsity-Constrained Deep Nonnegative Matrix Factorization With Total Variation. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6245–6257. [CrossRef]
38. Huang, R.; Jiao, H.; Li, X.; Chen, S.; Xia, C. Hyperspectral Unmixing Using Robust Deep Nonnegative Matrix Factorization. *Remote Sens.* **2023**, *15*, 2900. [CrossRef]
39. Li, H.C.; Feng, X.R.; Zhai, D.H.; Du, Q.; Plaza, A. Self-Supervised Robust Deep Matrix Factorization for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5513214. [CrossRef]
40. Kong, D.; Ding, C.; Huang, H. Robust nonnegative matrix factorization using L21-norm. In Proceedings of the CIKM '11: 20th ACM International Conference on Information and Knowledge Management, Glasgow, UK, 24–28 October 2011; pp. 673–682.
41. Prokhorenkova, L.; Shekhovtsov, A. Graph-based nearest neighbor search: From practice to theory. In Proceedings of the ICML'20: 37th International Conference on Machine Learning, Virtual, 13–18 July 2020.
42. Xu, X.; Huang, Z.; Zuo, L.; He, H. Manifold-Based Reinforcement Learning via Locally Linear Reconstruction. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 934–947. [CrossRef]
43. Han, J.; Sun, Z.; Hao, H. Selecting feature subset with sparsity and low redundancy for unsupervised learning. *Knowl.-Based Syst.* **2015**, *86*, 210–223. [CrossRef]
44. Heinz, D.; Chein-I-Chang. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 529–545. [CrossRef]
45. Lu, X.; Wu, H.; Yuan, Y.; Yan, P.; Li, X. Manifold Regularized Sparse NMF for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2815–2826. [CrossRef]
46. Qu, K.; Li, Z.; Wang, C.; Luo, F.; Bao, W. Hyperspectral Unmixing Using Higher-Order Graph Regularized NMF With Adaptive Feature Selection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5511815. [CrossRef]
47. Zhou, C.; Rodrigues, M.R.D. Hyperspectral Blind Unmixing Using a Double Deep Image Prior. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 16478–16492. [CrossRef]
48. Rasti, B.; Zouaoui, A.; Mairal, J.; Chanussot, J. Fast Semisupervised Unmixing Using Nonconvex Optimization. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5526713. [CrossRef]
49. Zou, X.; Xu, M.; Liu, S.; Sheng, H. Superpixel-Based Graph Laplacian Regularized and Weighted Robust Sparse Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5526415. [CrossRef]
50. Cao, F.; Situ, Y.; Ye, H. A Joint Multiscale Graph Attention and Classify-Driven Autoencoder Framework for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5504514. [CrossRef]
51. Miao, L.; Qi, H. Endmember Extraction From Highly Mixed Data Using Minimum Volume Constrained Nonnegative Matrix Factorization. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 765–777. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

remote sensing

MDPI

Article

Weighted Total Variation for Hyperspectral Image Denoising Based on Hyper-Laplacian Scale Mixture Distribution

Xiaoyu Yu [1], Jianli Zhao [1,*], Sheng Fang [1], Tianheng Zhang [1], Liang Li [1] and Xinyue Huang [2]

[1] College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China; 202384060026@sdust.edu.cn (X.Y.); fangsheng@tsinghua.org.cn (S.F.); 202281060002@sdust.edu.cn (T.Z.); liangli@sdust.edu.cn (L.L.)

[2] College of Engineering Computing and Cybernetics, Australian National University, Canberra 0200, Australia; u7643012@anu.edu.au

* Correspondence: jlzhao@sdust.edu.cn

Highlights

What are the main findings?

- Proposes a Hyper-Laplacian Scale Mixture prior for HSI gradient modeling.
- Introduces a region-adaptive weighted total variation model for HSI denoising.

What are the implications of the main findings?

- Improves denoising performance in both smooth and textured regions with better texture preservation.
- Achieves state-of-the-art performance and can enhance other TV-based methods when integrated.

Abstract

Conventional total variation (TV) regularization methods based on Laplacian or fixed-scale Hyper-Laplacian priors impose uniform sparsity penalties on gradients. These uniform penalties fail to capture the heterogeneous sparsity characteristics across different regions and directions, often leading to the over-smoothing of edges and loss of fine details. To address this limitation, we propose a novel regularization Hyper-Laplacian Adaptive Weighted Total Variation (HLAWTV). The proposed regularization employs a proportional mixture of Hyper-Laplacian distributions to dynamically adapt the sparsity decay rate based on image structure. Simultaneously, the adaptive weights can be adjusted based on local gradient statistics and exhibit strong robustness in texture preservation when facing different datasets and noise. Then, we propose an hyperspectral image (HSI) denoising method based on the HLAWTV regularizer. Extensive experiments on both synthetic and real hyperspectral datasets demonstrate that our denoising method consistently outperforms state-of-the-art methods in terms of quantitative metrics and visual quality. Moreover, incorporating our adaptive weighting mechanism into existing TV-based models yields significant performance gains, confirming the generality and robustness of the proposed approach.

Keywords: hyperspectral image (HSI) denoising; Hyper-Laplacian Scale Mixture (HLSM); total variation (TV); tensor completion

1. Introduction

Hyperspectral images acquire data across hundreds of contiguous spectral bands, enabling detailed characterization of real-world scenes through rich spectral and spatial information. This makes hyperspectral images indispensable in a variety of applications, including anomaly detection [1], environmental monitoring [2], military surveillance [3], precision agriculture [4], and object detection [5]. However, during the acquisition process, HSI is often corrupted by various noise types such as Gaussian noise, impulse noise, stripe noise, and deadlines. These mixed noises significantly deteriorate the image quality and adversely affect subsequent processing tasks [6–9]. Therefore, robust and efficient denoising techniques are essential to enhance the reliability of hyperspectral images in practical scenarios.

An intuitive approach is to treat each band of a hyperspectral image as an individual grayscale image and directly apply traditional grayscale denoising algorithms, such as non-local means (NLM) [10], block-matching 3D filtering (BM3D) [11], and weighted nuclear norm minimization (WNNM) [12]. However, these methods ignore the spectral correlations of HSI. To solve this problem, some algorithms process the entire hyperspectral data as a 3D cube, such as block-matching 4D filtering (BM4D) [13] and video block-matching 4D filtering (V-BM4D) [14]. However, without prior constraints on the inherent characteristics of HSI, it is difficult to achieve optimal denoising performance. Consequently, researchers have begun developing denoising methods that exploit spectral correlations and spatial structures. These methods are generally categorized into two classes as follows: (1) Deep learning-based methods directly learn mapping functions between noisy and clean HSI pairs from training datasets to avoid explicit modeling of prior knowledge. (2) Prior model-based methods formulate HSI denoising as a constrained optimization problem by incorporating hyperspectral priors such as low-rankness in the spectral dimension, local smoothness in the spatial domain, and self-similarity of pixel patches.

In recent years, deep learning–based hyperspectral image denoising has advanced from simple spatial–spectral convolutional neural networks (CNNs) to more sophisticated architectures that effectively balance local and global context. HSID-CNN [15] pioneered the use of deep residual convolution with multiscale and multilevel feature fusion to suppress noise while preserving spectral–spatial consistency. QRNN3D [16] introduced alternating-direction quasi-recurrent pooling alongside 3D convolutions to model both spatio-spectral correlations and global spectral dependencies simultaneously. More recently, SSRT-UNet [17] integrated non-local spatial self-similarity and global spectral correlation within a single recurrent transformer U-Net block, enabling unified exploitation of hyperspectral priors across shift-window layers. TRQ3DNet [18] combined 3D quasi-recurrent blocks with U-Transformer modules connected via a bidirectional integration bridge, capturing both local textures and long-range dependencies and delivering state-of-the-art denoising performance. ILRNet [19] integrates the strengths of model-driven and data-driven approaches by embedding a rank minimization module within a U-Net architecture. However, these deep learning models still face significant limitations. Their performance heavily relies on paired clean–noisy datasets and assumed noise priors, which hinders their adaptability in real-world scenarios where such datasets are not available, or the noise characteristics may differ significantly. This reliance on prior knowledge and training on specific datasets restricts the generalization capability of these models when exposed to complex or unseen noise types, highlighting a key challenge in applying them to diverse and unstructured real-world noise conditions.

Prior model-based approaches for hyperspectral image denoising have attracted significant attention due to their flexibility and strong interpretability. Rather than relying on large-scale annotated datasets, these methods explicitly incorporate various types of

https://doi.org/10.3390/rs18010135

prior knowledge about hyperspectral images, such as low-rankness and local smoothness. By leveraging these priors, model-based methods can adapt more effectively to different noise types and complex real world scenarios.

Among these priors, low-rankness has proven especially useful for modeling the strong spectral correlation inherent in hyperspectral data. Building upon this idea, Zhang et al. proposed the low-rank matrix recovery (LRMR) framework [20], which exploits spectral low-rank priors to suppress noise. However, methods based solely on low-rank tend to overlook spatial structure information, resulting in suboptimal performance when dealing with complex noise such as stripes or structured artifacts.

To address the above problem, spatial smoothness priors have been introduced. For example, He et al. incorporated TV regularization [21] in their low-rank matrix denoising (LRTV) method [22], applying both low-rankness and local smoothness to better preserve spatial details and enhance denoising performance. Nevertheless, this method requires unfolding hyperspectral data into matrices, which can destroy the intrinsic spatial structure. To overcome this limitation, recent methods have modeled hyperspectral images as 3D tensors. For example, LRTDTV [23] combines Tucker decomposition with spatial–spectral TV (SSTV) [24] to effectively remove complex noise. However, balancing the parameters between low-rank and smoothness regularizations remains challenging. Recognizing that low-rank regularization is often coupled with smoothness priors, Wang et al. [25] further proposed directly imposing low-rank constraints on gradient tensors to encode both low-rankness and local smoothing simultaneously within the tensor framework. This formulation improves the adaptivity and representation power of total variation regularization across different structural patterns in hyperspectral images. To improve the texture preserving capability of TV regularization, some methods introduce adaptive weighting strategies for sparsity constraints across spatial and spectral directions. For example, Anisotropic Spectral Spatial TV (ASSTV) [26] uses anisotropic full variability in three directions, which explores HSI spectral continuity and spatial smoothing while taking into account the directionality of the texture. Graph Spatio-Spectral TV (GSSTV) [27] weights the spatial difference operator of SSTV based on a graph reflecting the spatial structure of the HSI image, which can better respond to the spatial structure of the image. However, most of these weighting schemes apply the same sparsity constraint to every pixel, causing textured areas to be over-smoothed since they ignore local gradient variations. To solve this problem, Chen et al. proposed a TPTV regularizer [28] with gradient map weighting l_1-parameter. The gradient map is utilized to adaptively adjust the weights so as to reduce the sparsity constraints on pixels with large variations.

In recent years, Hyper-Laplacian TV (HLTV) methods have garnered significant attention because they can more accurately model the peaky nature and heavy-tailed characteristics of spatial–spectral gradients in hyperspectral images compared to traditional Laplacian TV (LTV) [29,30]. However, although both LTV and HLTV approaches strive to equation the regularization with the true sparse distribution of gradients, current TV-based denoising methods still have the following problems in hyperspectral image denoising:

(1) Existing HLTV regularizers use a fixed l_p-norm constraint for each gradient direction, implicitly assuming that gradients originating from smooth areas or richly textured regions, follow the same heavy-tailed distribution. In reality, hyperspectral images exhibit markedly different gradient statistics across spatial versus spectral dimensions and between smooth versus textured regions as shown in Figure 1. By ignoring this heterogeneity, HLTV regularization fails to capture the true prior distribution of the data. As a result, it often over-smooths details in textured areas, leading to suboptimal noise removal.

(2) Although LTV models introduce heuristic weight functions to protect edges and textures, these weights are typically fixed mappings derived from simplified priors. They

https://doi.org/10.3390/rs18010135

lack the flexibility to adapt dynamically to local gradient characteristics. Consequently, such models struggle to discriminate effectively between flat regions and highly textured ones as follows: over-smoothing may be introduced in flat regions, while insufficient protection is provided in highly textured regions, resulting in the underutilization of the sparsity feature of HSI gradients. In addition, such heuristics are difficult to adaptively set reasonable weight sizes based on the statistical properties of HSI, thus limiting the overall regularization performance.

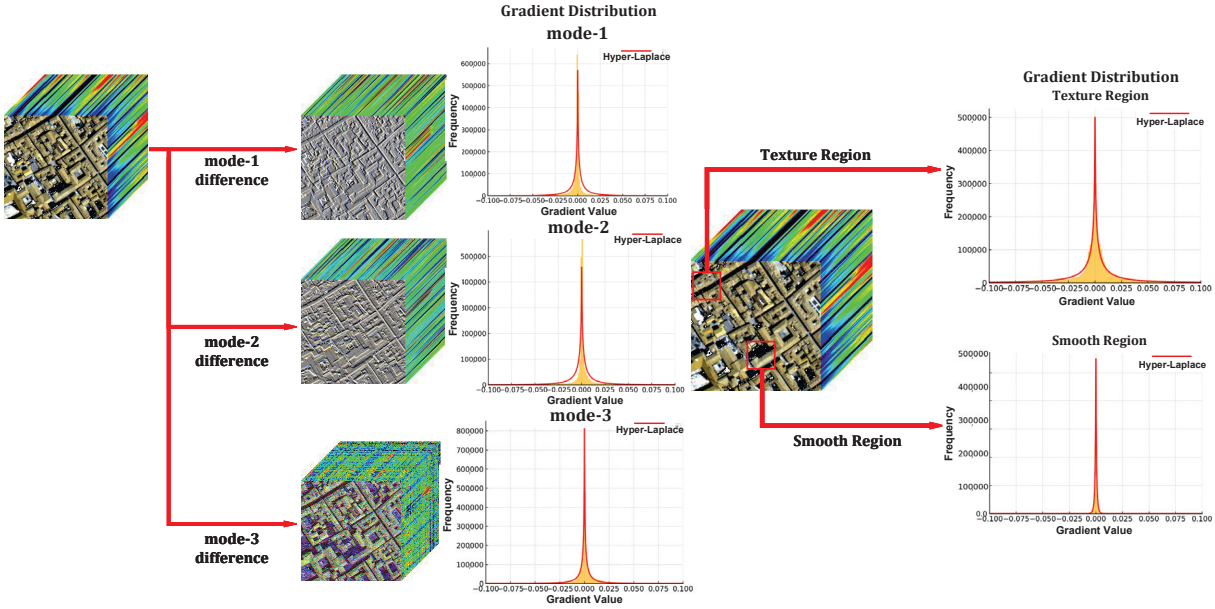


Figure 1. Hyper-Laplacian modeling of spatial–spectral gradients and region-wise (smooth vs. textured) gradient distributions.

Based on the above observations, we propose a novel HLAWTV regularization method, which is built upon a Hyper-Laplacian Scale Mixture (HLSM) distribution model. Our method assumes that gradients in hyperspectral images follow Hyper-Laplacian distributions with different scales in smooth and textured regions. We embed this assumption into a maximum a posteriori (MAP) estimation framework, which yields an adaptive weighted regularization term. Specifically, the regularization fits variable-scale Hyper-Laplacian distributions based on both the direction-specific gradient statistics and the magnitude of local gradient values, enabling the model to control the sparsity decay rate dynamically. This model improves the adaptivity of total variation regularization across different structural patterns in hyperspectral images.

The main contributions of this article can be summarized as follows.

(1) We propose a sparse prior modeling method based on the Hyper-Laplacian scale mixture distribution to more accurately characterize the statistical properties of hyperspectral image gradients. Specifically, considering that the HSI gradients usually exhibit distribution patterns with varying degrees of spiking and heavy tailing in different directions, we model them using a variable scale super-Laplacian distribution. To further introduce regional adaptation, we derive a probabilistic scale mixture model by exploiting the prior conjugacy of the Hyper-Laplacian distribution with the gamma distribution.

(2) Based on the HLSM distribution, we further propose a new weighted total variation regularization model HLAWTV. To efficiently solve the resulting optimization problem, we develop a tailored algorithm based on the Alternating Direction Method of Multipliers (ADMM), which achieves effective and scalable denoising performance.

https://doi.org/10.3390/rs18010135

(3) Experimental results demonstrate that the proposed denoising method performs better than other advanced methods. Furthermore, migration experiments show that the denoising performance can be significantly improved by embedding our weighting scheme into existing TV regularizers.

The rest of this article is organized as follows. The second part gives signs and preliminary knowledge. The third part introduces the HLSM distribution model, HLAWTV regularizer, and denoising model. The fourth part reports the experimental results. The fifth part is the conclusion.

2. Preliminaries

2.1. Notations

We summarize notations used in this article in Table 1. Next, we introduce the definitions of mode-n unfolding, the mode-k tensor-matrix product, and the mode-k gradient map.

Table 1. Notation declarations.

Notations	Interpretations
x, $\mathbf{x}$, X, $\mathcal{X}$	scalar, vector, matrix, tensor
$\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$	Nth-order tensor
$x_{i_1,i_2,\ldots,i_N}$	element at position $(i_1, i_2, \ldots, i_N)$ of $\mathcal{X}$
$\|\mathcal{X}\|_1$	$\sum_{i_1,i_2,\ldots,i_N} \lvert x_{i_1,i_2,\ldots,i_N} \rvert$
$\|\mathcal{X}\|_q$	$\sqrt[q]{\sum_{i_1,i_2,\ldots,i_N} \lvert x_{i_1,i_2,\ldots,i_N} \rvert^q}$

The n-mode unfolding of N-order tensors can be divided into horizontal unfolding $\mathrm{X}_{(n)} \in \mathbb{R}^{I_1 \ldots I_{n-1} I_{n+1} \ldots I_N}$ and vertical unfolding $\mathrm{X}^{(n)} \in \mathbb{R}^{I_1 \ldots I_{n-1} I_{n+1} \ldots I_N}$.

The mode-k tensor-matrix product of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$ and a matrix $\mathrm{A} \in \mathbb{R}^{J_1 \times I_k}$ is defined as follows:

$$\mathcal{Y} = \mathcal{X} \times_k \mathrm{A}. k = 1, 2, 3 \tag{1}$$

The gradient map of $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ along mode-k can be denoted as $\mathcal{G}_k(\mathcal{X})$, which is defined as follows:

$$\mathcal{G}_k(\mathcal{X}) = \nabla_k(\mathcal{X}) = D_k \mathcal{X}, k = 1, 2, 3, \tag{2}$$

where D_k represent the first-order finite difference operators.

2.2. The Laplacian Scale Mixture Distribution

A random variable s_i is said to follow a Laplacian Scale Mixture (LSM) distribution if it can be expressed as follows:

$$s_i = \lambda_i^{-1} u_i, \tag{3}$$

where u_i is a Laplacian random variable with the following unit scale:

$$p(u_i) = \frac{1}{2} e^{-|u_i|}, \tag{4}$$

and λ_i is a positive random variable with density $p(\geq_i)$. We assume u_i and λ_i are independent. Conditioned on λ_i, the distribution of s_i becomes a Laplacian with inverse scale λ_i as follows:

$$p(s_i|\lambda_i) = \frac{\lambda_i}{2} e^{-\lambda_i |s_i|}. \tag{5}$$

https://doi.org/10.3390/rs18010135

Therefore, the marginal distribution of s_i is an infinite mixture of Laplacians with different scales as follows:

$$p(s_i) = \int_0^\infty p(s_i \mid \lambda_i)p(\lambda_i)d\lambda_i = \int_0^\infty \frac{\lambda_i}{2}e^{-\lambda_i|s_i|}p(\lambda_i)d\lambda_i. \tag{6}$$

2.3. TV Regularizatiers

TV regularizatiers methods have emerged in hyperspectral image restoration, such as the early proposed 3-D TV model [23], the subsequent improved Enhanced 3-D TV (E3DTV) [31], and the correlated TV model [32]. To facilitate the subsequent analysis, we abstract and generalize their structures and use a unified representation framework to represent the above different TV regularizations as follows:

$$\|\mathcal{X}\|_{3DTV} = \|D_1\mathcal{X}\|_1 + \|D_2\mathcal{X}\|_1 + \|D_3\mathcal{X}\|_1, \tag{7}$$

where $\mathcal{X}$ is the mathematical representation of HSIs in different TV regularization models.

2.4. HSI Degradation Model

Since HSIs are often contaminated with various types of noise, the degradation model is typically represented as follows:

$$\mathcal{Y} = \mathcal{X} + \mathcal{E}, \tag{8}$$

where the tensor $\mathcal{Y}$ represents the observed hyperspectral image, the tensor $\mathcal{X}$ represents the clean HSI, and the tensor $\mathcal{E}$ is the noise term. The goal of HSI denoising is to estimate the clean image $\mathcal{X}$ using the corrupted image $\mathcal{Y}$. The denoising model for HSIs can be represented as

$$\arg\min_{\mathcal{X},\mathcal{E}} \frac{1}{2}\|\mathcal{Y} - \mathcal{X} - \mathcal{E}\|_F^2 + \eta J_1(\mathcal{X}) + \lambda J_2(\mathcal{E}), \tag{9}$$

where the first term is the data fidelity term, the second term is used to describe the prior information of the clean HSI [33–35], and the third term is used to describe the sparsity of the mixed noise.

3. Methodology

In this section, we propose the HLAWTV regularizer and give its technical details, then propose a denoising model based on the HLAWTV regularizer, and finally utilize the ADMM algorithm to solve the denoising model.

3.1. HLSM Model

Different TV regularizers allow flexibility in balancing denoising and texture preservation capabilities by selecting different forms of $\mathcal{X}$. In traditional regularization frameworks, the l_1-norm is widely used to constrain the sparsity of the gradient map, implicitly assuming that the gradient map follows a fixed-scale LSM distribution [36,37] with probability density function (PDF) $p(x;\lambda) \propto e^{-\lambda|x|}$. However, some recent research has found that both spatial and spectral gradients in natural images exhibit significantly more heavy-tailed distributions. Specifically, their heavy-tailed distributions are closer to the Hyper-Laplacian distribution with probability density function $p(x;\lambda) \propto e^{-\lambda|x|^{\alpha_k}}$, where α_k controls the strength of the sparsity in that direction. A smaller α_k represents a stronger sparse prior.

Based on the above theory, we propose a Hyper-Laplacian scale mixture distribution model. Current TV regularization based on the Hyper-Laplacian distribution assumes that the gradient values in the same difference direction independently and identically obey the Hyper-Laplacian distribution of the same scale. However, different regions of HSIs often

have completely different structural characteristics. To better capture the nonstationary and sparse nature of gradient variations, we assume that the gradient values at each pixel point follow an independent distribution of the HLSM model. As shown in Figure 2, the methods based on HLSM distribution model closely matches the gradient distribution of $\mathcal{X}$. Specifically, for each pixel-wise gradient value $\mathcal{G}_{i,j,b}$, we assume the following model:

$$p(\mathcal{G}_{i,j,b}) = \prod_{i,j,b} p(\mathcal{G}_{i,j,b}). \tag{10}$$

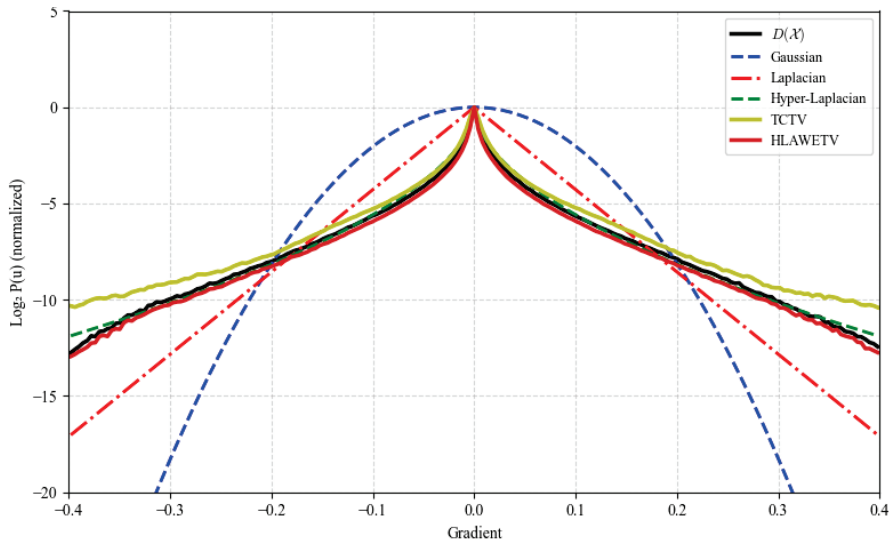


Figure 2. Empirically fitted curves for the gradient of the PaviaC dataset.

Instead of assuming a LSM distribution, we define the following:

$$\begin{aligned} p(\mathcal{G}_{i,j,b}) &= \int p(\mathcal{G}_{i,j,b} \mid \Theta_{i,j}) p(\Theta_{i,j}) d\Theta_{i,j} \\ p(\mathcal{G}_{i,j,b} \mid \Theta_{i,j}) &= \frac{\alpha_k \Theta_{i,j}}{2} |\mathcal{G}_{i,j,b}|^{\alpha_k - 1} \exp\left(-\Theta_{i,j} |\mathcal{G}_{i,j,b}|^{\alpha_k}\right), \end{aligned} \tag{11}$$

where $0 < \alpha_k < 1$ is a sparsity-controlling exponent, allowing stronger sparsity than the Laplacian distribution (which corresponds to $\alpha = 1$).

Since the integral form of $p(\mathcal{G}_{i,j,b})$ does not yield a closed-form solution, we adopt an Expectation-Maximization (EM) strategy to optimize the marginal log-likelihood $\ln(p(\mathcal{G}))$.

E step: We compute the evidence lower bound (ELBO) of the log-marginal probability, given the current estimate $\mathcal{G}_{(t)}$. The ELBO is denoted as follows:

$$Q(\mathcal{G}, \mathcal{G}^{(t)}) = \int \prod_{i,j} p(\Theta_{i,j} \mid \mathcal{G}^{(t)}) \ln \prod_{i,j,b} p(\mathcal{G}_{i,j,b} \mid \Theta_{i,j}) d\Theta_{i,j}. \tag{12}$$

Using the form of the Hyper-Laplacian conditional distribution, the optimization problem can be reformulated as follows:

$$Q(\mathcal{G}, \mathcal{G}^{(t)}) = -\sum_{i,j,b} \langle \Theta_{i,j} \rangle \cdot |\mathcal{G}_{i,j,b}|^{\alpha_k} + const, \tag{13}$$

where $\langle \Theta_{i,j} \rangle = W(i,j)$ is the expectation of $\Theta_{i,j}$ under the current posterior estimate. $W(i,j)$ is a function of $\mathcal{G}^{(t)}$. $\alpha_k = f(\mu_k, \sigma_k)$. Both $W(i,j)$ and $f(\mu_k, \sigma_k)$ will be discussed later. In practice, we assume that the posterior $p(\Theta_{i,j} \mid \mathcal{G}^{(t)})$ obeys a Gamma distribution, or use local statistics to approximate it.

https://doi.org/10.3390/rs18010135

3.2. HLAWTV Regularizer with HLSM

M step: We maximize the ELBO w.r.t. $\mathcal{G}$, resulting in the following optimization:

$$\mathcal{G}^{(t+1)} = \arg\min_{\mathcal{G}} \sum_{i,j,b} W(i,j) \cdot |\mathcal{G}_{i,j,b}|^{\alpha_k}. \tag{14}$$

Consequently, following the optimization structure induced by the EM algorithm, we define the HLAWTV regularization term as follows:

$$\begin{aligned} \|\mathcal{Z}\|_{\text{HLAWTV}} &= \sum_{k=1}^{3} \|D_k\mathcal{X}\|_{\alpha_k,W} \\ &= \sum_{k=1}^{3} \sum_{m,n,b} W_k(i,j) \cdot |\mathcal{G}_k(i,j,b)|^{\mathbf{ff}_k}. \end{aligned} \tag{15}$$

In this way, incorporating HLAWTV as the regularization term is equivalent to maximizing the joint log-likelihood of the gradient priors along all directions, formulated as follows: $-\sum_{i=1}^{3} \ln(p(\mathcal{G}_i^{\mathcal{X}}))$. This demonstrates that the adoption of HLAWTV serves not only as a regularization strategy but also as a probabilistic modeling of gradients, grounded in a prior-driven MAP estimation framework.

In our denoising framework, the adaptive weight map W plays a crucial role by acting as the scale parameter in the HLSM. Rather than assuming a fixed scale, we adopt a spatially adaptive formulation where W reflects the local structural complexity of the HSI. Because this scale parameter is a latent variable in HLSM, it is inherently unknown and its value is influenced by local spatial characteristics. To make W more data-adaptive, we explicitly design $W(i,j)$ to approximate this latent scale, effectively linking the statistical sparsity prior with image-driven structural information. In particular, large weights correspond to stronger contractions to ensure sparsity in the smoothing region, while small weights preserve larger gradient magnitudes in texture or edge regions.

To robustly estimate $W(i,j)$, we use low-rank filtering to obtain denoising estimates for hyperspectral images, which reduces the interference of noise in weight calculation. HSI exhibits consistent texture patterns across all bands, whereas the noise in each band is random. In order to extract the texture region of HSI more accurately, we use mean filtering along the spectral dimension to obtain a gradient map with better texture detail quality. The calculation formula is as follows:

$$G_k = \frac{1}{B} \sum_{b=1}^{B} (\mathcal{G}_k)_{(:,:,b)}, k = 1,2,3 \tag{16}$$

where $G_k(i,j)$ is a matrix that provides a more reliable characterization of texture structures along the three modes of HSIs. To enhance the adaptivity of the regularization strength across different image regions, we use the Gamma distribution as a hyperprior over the scale parameter $\Theta_{i,j}$ in the HLSM as follows:

$$p(\Theta_i) = (b^a/\Gamma(a))\Theta_i^{a-1} e^{-b\Theta_i}. \tag{17}$$

This choice is justified on multiple grounds as follows: (1) The gamma distribution is conjugate to the exponential family, enabling tractable marginalization. (2) By adjusting the shape parameter a and inverse scale parameter b, the gamma distribution flexibly models diverse behavioral characteristics ranging from strong sparsity to weak sparsity. In our method, the weighting scheme originates from Bayesian inference under the gamma-

https://doi.org/10.3390/rs18010135

exponential conjugate model. Specifically, when the prior distribution of $\Theta_{i,j}$ is gamma and the conditional probability of $\mathcal{G}_{i,j,b}$ follows a Hyper-Laplacian distribution with inverse scale $\Theta_{i,j}$, the posterior probability of $\Theta_{i,j}$ given $\mathcal{G}_{i,j,b}$ is also gamma. This conjugacy allows us to directly update the posterior distribution of latent variables via the E-step of the EM algorithm as shown in Equation (13). Therefore, the posterior distribution of $\Theta_{i,j}$ is computed based on the current gradient $\mathcal{G}_{i,j,b}$ and remains a Gamma distribution as follows:

$$\Theta_k | G_k \sim Gamma(a+1, b+|G_k|). \tag{18}$$

The expectation of this posterior yields the following adaptive weight function:

$$W_k(i,j) = \mathbb{E}[\Theta_{i,j} | G_k(i,j)] = \frac{a+1}{b+|G_k(i,j)|}. \tag{19}$$

The proposed weighting function is monotonically decreasing in $G_k(i,j)$, which directly supports a sparse-modeling prior, as follows: when $|G_k(i,j)| \to 0$ in smooth regions, the weight becomes large to encourage smoothing; when $G_k(i,j)$ is large at edges or textures, the weight decreases to preserve fine details. Additionally, this weighting scheme improves robustness against noise, mitigating instability caused by gradient fluctuations.

3.3. Proposed Model

To better characterize the spatial–spectral structures of hyperspectral images, we propose a novel denoising model that integrates spectral subspace representation with the HLAWTV regularizer. By projecting the HSI gradient tensor into the spectral subspace, we are able to utilize the inherent spectral redundancy to suppress noise and maintain structural consistency across bands. This subspace representation improves denoising robustness while maintaining computational efficiency and is a strong complement to our proposed HLAWTV. Therefore, the spectral gradient of HSIs can be expressed as follows:

$$\mathcal{D}_k(\mathcal{X}) = \mathcal{U}_k \times_3 \mathrm{V}_k{}^T, \tag{20}$$

where $\mathcal{U}_k \in \mathbb{R}^{M \times N \times r}$ contains the corresponding spatial coefficients, and $\mathrm{V}_k \in \mathbb{R}^{B \times r}$ is the spectral basis. This transformation allows the spatial variation across bands to be compactly encoded in $\mathcal{U}_i$. We use the HLAWTV regularizer to constrain $\mathcal{U}_i$ to enhance both sparsity and adaptivity across different spatial regions and gradient directions. Based on the above discussion, our proposed HLAWTV denoising model can be expressed as follows:

$$\begin{aligned} \underset{\mathcal{X},\mathcal{E},\mathcal{U}_k,\mathrm{V}_k}{\arg\min} & \frac{1}{2}\|\mathcal{Y}-\mathcal{X}-\mathcal{E}\|_F^2 + \lambda \sum_{k=1}^{3} \|\mathcal{U}_k\|_{\mathrm{HLAWTV}} + \|\mathcal{E}\|_1 \\ & \mathcal{D}_k(\mathcal{X}) = \mathcal{U}_k \times_3 \mathrm{V}_k{}^T \\ & \mathrm{V}_k^T \mathrm{V}_k = I. \quad k = 1,2,3 \end{aligned} \tag{21}$$

https://doi.org/10.3390/rs18010135

3.4. Algorithm

The proposed HSI denoising model is solved using ADMM. According to the ADMM method, the optimization problem described in Equation (21) can be solved by minimizing the following augmented Lagrangian function:

$$\begin{aligned}\mathcal{L}(\mathcal{X},\mathcal{E},\mathcal{U}_k,\mathcal{V}_k,\mathcal{M}_k) =&\lambda\sum_{k=1}^{3}\|\mathcal{U}_k\|_{\alpha_k,W}+\|\mathcal{E}\|_1\\&+\sum_{k=1}^{3}\frac{\mu}{2}\|\mathcal{D}_k(\mathcal{X})-\mathcal{U}_k\times_3\mathrm{V}_k^T+\frac{\mathcal{M}_k}{\mu}\|_F^2\\&+\frac{\mu}{2}\|\mathcal{Y}-\mathcal{X}-\mathcal{E}+\frac{\mathcal{M}_4}{\mu}\|_F^2,\end{aligned}\tag{22}$$

where $\mathcal{M}_k$ (k = 1, 2, 3) and $\mathcal{M}_4$ are Lagrange multipliers. μ is the penalty parameter. Each variable can be updated alternatively, while the other variables remain constant.

(1) *Update* $\mathcal{U}_k$ *and* V_k: The subproblem with respect to $\mathcal{U}_i$ can be expressed as

$$\min_{\mathcal{U}_k}\lambda\|\mathcal{U}_i\|_{\alpha_k,W}+\frac{\mu}{2}\left\|\mathcal{D}_k(\mathcal{X})+\frac{\mathcal{M}_k}{\mu}-\mathcal{U}_k\times_3\mathrm{V}_k\right\|_F^2.\tag{23}$$

For each directional gradient tensor $\mathcal{D}_k$, we first unfold the 3D tensor into a 2D matrix to facilitate subspace modeling as follows:

$$\mathrm{D}_k=\mathrm{Unfold}(\mathcal{D}_k(\mathcal{X}))\in\mathbb{R}^{MN\times B}.\tag{24}$$

Then, the update of the subspace coefficient matrix U_k is obtained by solving the following:

$$\min_{\mathrm{U}_k}\frac{\mu}{2}\left\|\mathrm{D}_k+\frac{\mathrm{M}_k}{\mu}-\mathrm{U}_k\mathrm{V}_k^\top\right\|_F^2+\lambda\|\mathrm{U}_k\|_{\alpha_k,\mathrm{W}},\tag{25}$$

where D_k, U_k and M_k represent the mode-3 unfolding matrices of $\mathcal{D}_k$,$\mathcal{U}_k$ and $\mathcal{M}_k$. This problem is solved using the following two-step procedure:

In the first step, we update the matrices $\mathrm{U}_k^{(temp)}$ and $\mathrm{V}_k^{(t+1)}$ with Singular Value Decomposition (SVD) as follows:

$$\{\mathrm{U}_k^{(temp)},\mathrm{V}_k^{(t+1)}\}=\mathrm{SVD}\left(\mathrm{D}_k+\frac{1}{\mu}\mathrm{M}_k\right).\tag{26}$$

In the second step, we perform weighted nonconvex shrinkage on $\mathrm{U}_k^{(temp)}$ to update $\mathrm{U}_k^{(t+1)}$ as follows:

$$\mathrm{U}_k^{(t+1)}=\mathrm{Shrink}^q_{\frac{\lambda W_k}{\mu}}(\mathrm{U}_k^{(temp)}),\tag{27}$$

where $\mathrm{Shrink}^q(\cdot)$ is a generalized shrankage/thresholding operator [38]. To set parameters that better alight with the sparsity characteristics of gradients in both spatial and spectral dimensions, we set the value of q as α_k, shown in Equation (35).

(2) *Update* $\mathcal{E}$: The subproblem with respect to $\mathcal{E}$ can be expressed as

$$\min_{\mathcal{E}}\left\|\mathcal{Y}-\mathcal{X}+\frac{\mathcal{M}_4}{\mu}-\mathcal{E}\right\|_F^2+\frac{1}{\mu}\|\mathcal{E}\|_1.\tag{28}$$

The solution of $\mathcal{E}$ can be computed using the following soft-thresholding operator:

$$\mathcal{E}^{(t+1)}=\mathcal{E}_{\frac{1}{\mu}}(\mathcal{Y}-\mathcal{X}+\frac{\mathcal{M}_4}{\mu}).\tag{29}$$

https://doi.org/10.3390/rs18010135

(3) *Update* $\mathcal{X}$: The subproblem with $\mathcal{X}$ respect to can be expressed as

$$\begin{aligned}\min_{\mathcal{X}} &\left\| \mathcal{Y} - \mathcal{E} + \frac{\mathcal{M}_4}{\mu} - \mathcal{X} \right\|_F^2 \\ &+ \sum_{k=1}^{3} \left\| \mathcal{U}_k \times_3 V_k^T - \frac{\mathcal{M}_k}{\mu} - \mathcal{D}_k(\mathcal{X}) \right\|_F^2.\end{aligned} \tag{30}$$

Optimizing the above problem can be considered as solving the following linear system:

$$\begin{aligned}\left(\mu I + \mu \sum_{k=1}^{3} \mathcal{D}_k^T \mathcal{D}_k \right) \mathcal{X} &= \mu(\mathcal{Y} - \mathcal{E}) + \mathcal{M}_4 \\ &+ \mu \sum_{i=1}^{3} \mathcal{D}_k^T \left(\mathcal{U}_k \times_3 V_k^T - \frac{\mathcal{M}_k}{\mu} \right)\end{aligned} \tag{31}$$

where D_k^T indicates the transpose of D_k. This quadratic problem has an analytic solution on the Fourier domain as follows:

$$\begin{cases} \mathcal{H} = \mathcal{Y} - \mathcal{E} + \frac{\mathcal{M}_4}{\mu} + \sum_{k=1}^{3} \mathcal{D}_k^T \left(\mathcal{U}_k \times_3 V_k - \frac{\mathcal{M}_k}{\mu} \right) \\ \mathcal{T} = |\mathcal{F}(\mathcal{D}_1)|^2 + |\mathcal{F}(\mathcal{D}_2)|^2 + |\mathcal{F}(\mathcal{D}_3)|^2 \\ \mathcal{X}^{t+1} = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(\mathcal{H})}{\mathbf{1} + \mathcal{T}} \right) \end{cases} \tag{32}$$

where **1** represents the tensor with all elements as 1. $\mathcal{F}(\cdot)$ denotes the Fourier transform operator, and $|\cdot|^2$ represents the squared magnitude of the element.

(4) *Update* W_k: To solve this subproblem, we calculate the parameter α_k for each direction based on the mean and standard deviation of the current gradient. Specifically, for each direction, the calculation of α_k is broken down into several steps as follows:

First, we normalize the rank of the current low-rank approximation matrix as follows:

$$R = \frac{r_0 - r_{\min}}{r_{\max} - r_{\min}}. \tag{33}$$

As the iteration proceeds, R represents the rank of the current low-rank approximation matrix. It is normalized based on the current rank r_0, and the minimum and maximum ranks r_{min} and r_{max}. This normalization ensures that the adjustment of the sparsity penalty is based on the current rank, dynamically adjusting it during iterations. Next, we compute the penalty term based on the standard deviation σ_k and mean μ_k of the gradient to fit the gradient distribution in different directions as follows:

$$P_k = \exp\left(-\frac{\sigma_k}{\mu_k + \epsilon} \right). \tag{34}$$

Combining the ratio R and the penalty term P_k, we calculate α_k using the following formula:

$$\alpha_k = f(\mu_{i,j}, \sigma_{i,j}) = (1 - RP_k) \cdot \alpha_{\max} + RP_k \cdot \alpha_{\min}. \tag{35}$$

Using the calculated values α_k as the shape parameter a, we update the weight W_k in each pixel (i, j) for each direction. The weight update is given by the following:

$$W_k(i,j) = \frac{\alpha_k + 1}{b + |G_k(i,j)|}. \tag{36}$$

https://doi.org/10.3390/rs18010135

To stabilize the weights, we normalize them as follows:

$$W_k(i,j) = \frac{W_k(i,j)}{\max(W_k(i,j))}. \tag{37}$$

The multipliers are updated as follows:

$$\begin{cases} \mathcal{M}_k^{(t+1)} = \mathcal{M}_k^{(t)} + \mu\left(\mathcal{D}_k(\mathcal{X}) - \mathcal{U}_k \times_3 V_k^T\right), \quad k = 1,2,3 \\ \mathcal{M}_4^{(t+1)} = \mathcal{M}_4^{(t)} + \mu(\mathcal{Y} - \mathcal{X} - \mathcal{E}). \end{cases} \tag{38}$$

The specific process of the ADMM solver for HLAWTV is shown in Algorithm 1.

Algorithm 1 ADMM solver for HLAWTV

Require: Noisy HSI $\mathcal{Y}$; regularization λ; rank bounds $r_{\min}$, $r_{\max}$; shape-parameter bounds $\alpha_{\min}$, $\alpha_{\max}$
Ensure: Recovered HSI $\mathcal{X}$
1: $\mathcal{X} \leftarrow \mathcal{Y}$, $\mathcal{E} \leftarrow 0$, $\mathcal{M}_k \leftarrow 0$
2: $t \leftarrow 1$
3: **while** not converged **do**
4: Update $\mathcal{U}_i$, $\mathcal{V}_i$ by Equations (23)–(27)
5: Update $\mathcal{E}$ by Equation (29)
6: Update $\mathcal{X}$ by Equation (32)
7: Update W_k by Equation (37)
8: Update $\mathcal{M}_k$ by Equation (38)
9: $t \leftarrow t + 1$
10: **end while**

4. Results

To validate the superiority of the proposed method, we conduct extensive comparative experiments on both synthetic and real HSI datasets. Our method is compared with the following thirteen classical or state-of-the-art methods: the QRNN3D [16] using the recurrent network to explore the spatial–spectral and the global correlation along the spectrum, the ILRNet [19] embedding a rank minimization module within a U-Net architecture to integrate the strengths of model-driven and data-driven approaches, the LRTV [22] applying the traditional 2-D total variation regularizer to each band independently, the LRTDTV [23] integrating the SSTV regularizer into a low-rank tensor decomposition framework, TCTV [25] considering the spectral correlation and sparsity prior on gradient tensor, E3DTV [31] adopting edge-preserving strategies to enhance spatial detail retention, RCTV [32] imposing TV regularization on representative low-rank coefficients, and ETPTV [28] introducing a texture-preserving weighted TV regularizer. In addition to TV-based models, we also consider the following low-rank- and non-local-based methods: TNN [35] employing a tensor nuclear norm to exploit global low-rankness, LRTF-DFR [39] introducing a dual-factor low-rank tensor decomposition with spatial regularization, HNN [40] using a two-dimensional frontal Haar wavelet transform to disentangle low- and high-frequencies, FastHyDe [41] combining spectral subspace projection with non-local denoising for high efficiency, and NGmeet [42] jointly leveraging global low-rank spectral subspace and non-local self-similarity through block matching and iterative refinement. The relationships and differences between the proposed method and the compared methods are presented in Table 2. All deep learning models are implemented in PyTorch 2.5.1 and trained on a workstation equipped with an Intel Xeon Platinum 8480+ CPU and Nvidia H800 GPU, using a batch size of 4. All traditional methods are implemented in MATLAB R2018b and executed on a workstation equipped with an Intel Core i5-12600KF processor (2.50 GHz) and 16 GB of RAM.

https://doi.org/10.3390/rs18010135

4.1. Synthetic Data Experiments

To ensure a fair and comprehensive comparison, we performed simulated experiments on two hyperspectral datasets with distinct spatial and spectral characteristics. The Pavia City Centre (PaviaC) dataset1 with a size of 200 × 200 × 80 and the Washington DC Mall (WDC) dataset2 whose size is 256 × 256 × 144. These two datasets can be regarded as ground-truth hyperspectral data, as they exhibit no perceptible noise. The pixel intensities of each spectral band are normalized to the range [0, 1]. To simulate realistic acquisition scenarios, we synthetically generate six types of noise and add them to the clean dataset.

Case 1: Each pixel was corrupted by additive *i.i.d.* Gaussian noise with a standard deviation of $\sigma = 0.1$.

Case 2: *non-i.i.d.* Gaussian noise was simulated by adding *i.i.d.* Gaussian noise with different variances to each spectral band, where the SNR for each band was randomly selected from the range of [1, 15] dB.

Case 3: A mixture of Gaussian noise and impulse noise is added to each band. The Gaussian noise is added as *Case 1*. The percentage of impulse noise is set to 0.2.

Case 4: The impulse noise is added as in *Case 3*. Furthermore, 20% (PaviaC) or 30% (WDC) of the bands were randomly selected to add random stripes, with the number of corrupted columns ranging from 10 to 30. A total of 10% of the bands were randomly selected to simulate deadline noise, in which 5 to 25 rows were randomly chosen and set to zero to emulate dead sensor lines.

Case 5: The stripe noise and deadlines are added as *Case 4*. The percentage of impulse noise is set to 0.5 (PaviaC) and 0.3 (WDC).

Case 6: A mixture of Gaussian noise, impulse noise, stripe noise, and deadlines are added to each band. The Gaussian noise is added as *Case 2*. The stripe noise and deadlines are added as *Case 4*. The percentage of impulse noise is set to 0.2.

To ensure the statistical reliability of the experimental results, each denoising experiment was repeated 20 times for each noise case. The denoising performance was quantitatively evaluated using three metrics: mean peak signal-to-noise ratio (MPSNR) [43], mean structural similarity index (MSSIM) [44], and spectral angle mapper (SAM), which together reflect spatial fidelity, structural preservation, and spectral accuracy. Furthermore, the average runtime of each method was recorded to assess computational efficiency.

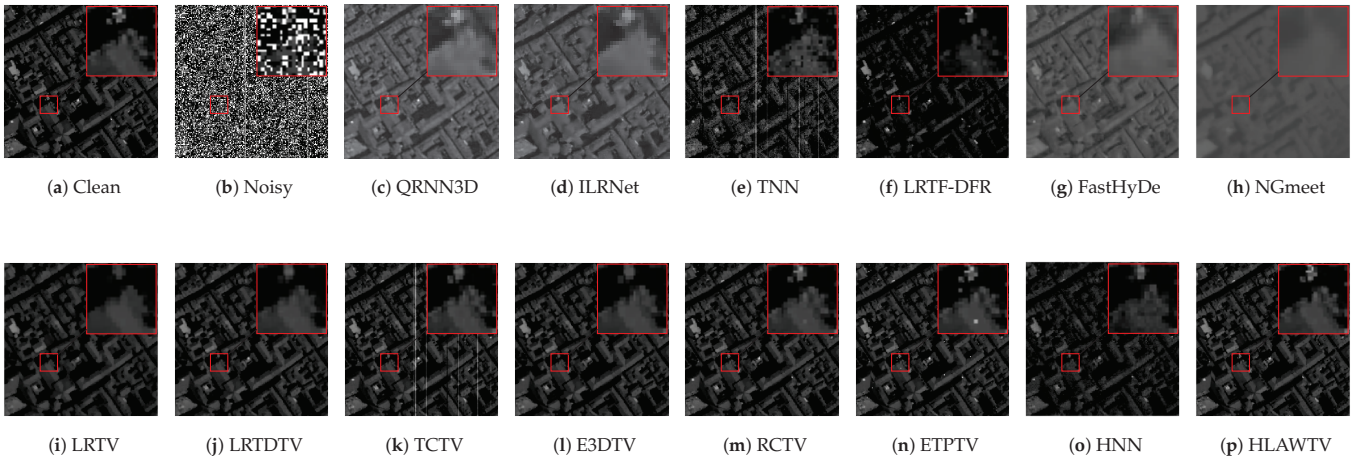
Quantitative Comparison: We present the MPSNR, MSSIM, MSAM, and the mean time values of 20 repetitive experiments obtained by all the compared methods in Tables 3 and 4. It can be seen that deep-learning denoising methods perform less effectively than traditional machine-learning approaches in both the PaviaC and WDC datasets, mainly because the training dataset ICVL is not similar to these datasets. In Case 2, FastHyDe achieves excellent performance in terms of MPSNR and MSSIM. This is because FastHyDe transforms non-i.i.d. Gaussian noise into i.i.d. noise through whitening, which significantly simplifies the noise model. However, it underperforms the proposed method in other cases. As observed, the proposed HLAWTV achieves overall superior results over the compared methods under nearly all cases. In both PaviaC and WDC datasets, our algorithm still achieves optimal results with i.i.d. Gaussian noise and non-i.i.d Gaussian noise compared to the algorithm using Gaussian noise constraint terms. In the case of impulse noise and structural noise (stripe noise and deadlines), our algorithm has less performance degradation with noise enhancement compared to other TV regularization based methods. This is due to the ability of our HLAWTV regularizer to adaptively adjust the sparse penalty strength based on structural information in different regions and directions.

Visual Comparison: Figures 3 and 4 illustrate the denoising performance of all competing algorithms on the 25th band of the PaviaC and WDC datasets. The noisy observations (b) are severely degraded by impulse noise, vertical stripes, and deadlines, causing

most scene details to be overwhelmed. Deep-learning-based methods can remove noise effectively, but they suffer from heavy over-smoothing. Traditional low-rank models such as TNN, LRTF-DFR, and FastHyDe suppress part of the impulse noise yet leave conspicuous stripe residuals. NGmeet and LRTV further attenuate the stripes but suffer from heavy over-smoothing. The texture in PaviaC and the architectural details in WDC become blurred. LRTDTV, TCTV, E3DTV, and HNN successfully remove structured noise but introduce blocky artifacts. RCTV and ETPTV preserve geometric structures, though isolated impulse dots remain visible. The proposed HLAWTV almost completely eliminates impulse and stripe/deadlines corruption while retaining sharp boundaries. These observations substantiate the effectiveness and robustness of the HLSM prior combined with our adaptive weighting.

Table 2. Comparison of HSI denoising methods in terms of modeling properties and noise assumptions.

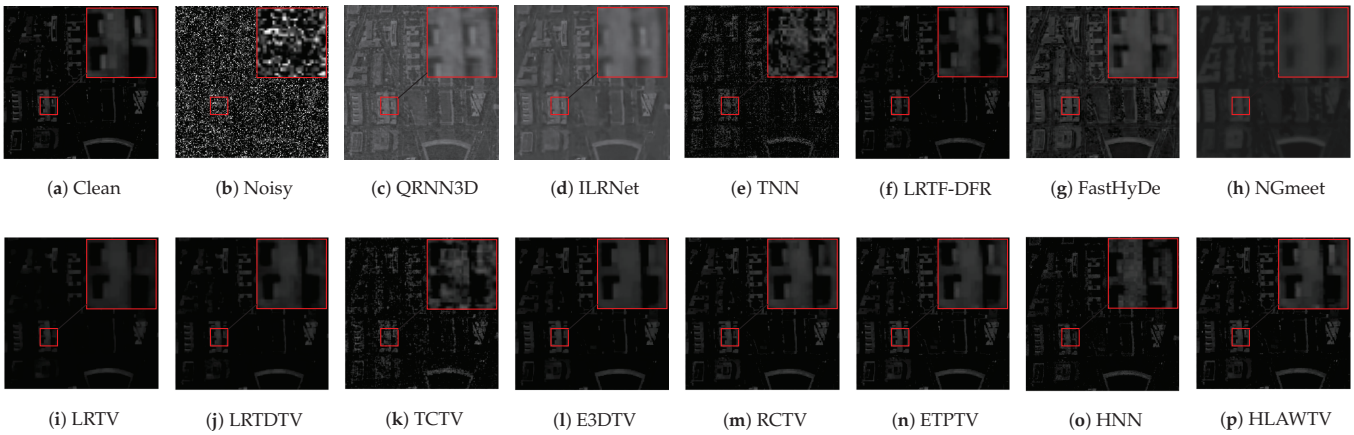
Method	Deep Learning Method	Low-Rank Modeling			Local Continuity		Non-Local Similarity	Gradient Distribution
		Spatial	Spectral	Both	Spatial	Spectral		
QRNN3D	✓							-
ILRNet	✓							-
TNN				✓				-
LRTF-DFR				✓	✓	✓		Laplacian
FastHyDe			✓				✓	-
NGmeet			✓				✓	-
LRTV			✓		✓			Laplacian
LRTDTV				✓	✓	✓		Laplacian
TCTV			✓	✓	✓	✓		Laplacian
E3DTV					✓	✓		Laplacian
RCTV			✓		✓			Laplacian
ETPTV				✓	✓	✓		Laplacian
HNN				✓				-
HLAWTV (Ours)				✓	✓	✓		Hyper-Laplacian



(**a**) Clean (**b**) Noisy (**c**) QRNN3D (**d**) ILRNet (**e**) TNN (**f**) LRTF-DFR (**g**) FastHyDe (**h**) NGmeet

(**i**) LRTV (**j**) LRTDTV (**k**) TCTV (**l**) E3DTV (**m**) RCTV (**n**) ETPTV (**o**) HNN (**p**) HLAWTV

Figure 3. Denoising results of different methods on Band 25 of the PaviaC dataset under Case 6.



(**a**) Clean (**b**) Noisy (**c**) QRNN3D (**d**) ILRNet (**e**) TNN (**f**) LRTF-DFR (**g**) FastHyDe (**h**) NGmeet

(**i**) LRTV (**j**) LRTDTV (**k**) TCTV (**l**) E3DTV (**m**) RCTV (**n**) ETPTV (**o**) HNN (**p**) HLAWTV

Figure 4. Denoising results of different methods on Band 25 of the WDC dataset under Case 6.

4.2. Real Data Experiments

To validate the proposed method under practical conditions, we executed experiments on two real HSI datasets. One of the real dataset selected is the Urban dataset whose size is 307 × 307 × 210, the other is the Indian Pines dataset with a size of 145 × 145 × 220. Both datasets are severely polluted, thereby furnishing a stringent benchmark for performance evaluation.

Figures 5 and 6 show two representative bands of Urban. After denoising, methods such as QNNN3D, TNN, FastHyDe, LRTDTV, TCTV, RCTV, and HNN still exhibit obvious stripe noise. ILRNet, NGmeet, LRTV, and E3DTV produced noticeable over-smoothing. In contrast, HLAWTV simultaneously removed noise while preserving building edges and road textures.

Figures 7 and 8 show two bands from the Indian Pines dataset. The low-rankness based algorithm such as TNN, TCTV, and HNN mitigates Gaussian noise but is less effective against impulse noise. The conventional TV-based model performs better in removing impulse noise but blurs the farmland boundaries. HLAWTV effectively removes noise while maintaining clear texture.

Figures 9 and 10 plot the spectral features of the Urban pixel at (185, 132) and the Indian Pines pixel at (125,20). After ILRNet, TNN, NGmeet, LRTDTV, TCTV, RCTV, and HNN denoising, the strong spikes in the original spectral are partially suppressed. However, there are still noticeable fluctuations in several noisy bands, which implies that there is still residual noise interference. In contrast, the curves produced by the other methods are much smoother after denoising, indicating that the noise is well suppressed on all the bands.

To evaluate the denoising performance of various methods on real-noise datasets, we employ the Blind/Unreferenced Image Space Quality Evaluator (BRISQUE) [45] as the unreferenced metric to assess the results of different methods on the Urban and Indian Pines datasets. BRISQUE is a perceptual quality assessment metric that evaluates image quality based on the statistical characteristics of natural scenes, requiring no real reference images for evaluation. A lower BRISQUE score indicates higher image quality. The results are presented in Table 5.

https://doi.org/10.3390/rs18010135



(**a**) Noisy (**b**) QRNN3D (**c**) ILRNet (**d**) TNN (**e**) LRTF-DFR

(**f**) FastHyDe (**g**) NGmeet (**h**) LRTV (**i**) LRTDTV (**j**) TCTV

(**k**) E3DTV (**l**) RCTV (**m**) ETPTV (**n**) HNN (**o**) HLAWTV

Figure 5. Denoising comparison on the real scenario at band 207 of Urban.

https://doi.org/10.3390/rs18010135

Table 3. Quantitative indices of the PaviaC dataset with different methods under different cases.

Index	Noisy	QRNN3D	ILRNet	TNN	LRTF-DFR	FastHyDe	NGmeet	LRTV	LRTDTV	TCTV	E3DTV	RCTV	ETPTV	HNN	HLAWTV
Case 1: i.i.d. Gaussian Noise (σ = 0.1)															
MPSNR	20.592	31.085	34.474	25.381	34.345	34.532	24.513	30.383	32.881	30.046	33.700	34.161	34.411	32.322	34.892
MSSIM	0.416	0.913	0.928	0.628	0.926	0.925	0.489	0.814	0.884	0.818	0.909	0.918	0.924	0.932	0.937
MSAM	0.478	0.115	0.083	0.313	0.079	0.098	0.125	0.101	0.099	0.197	0.091	0.108	0.098	0.108	0.077
Time	0.000	0.155	1.494	50.793	13.338	0.197	23.126	8.913	26.564	1345.642	9.541	2.850	46.544	28.024	40.991
Case 2: non-i.i.d. Gaussian Noise (SNR $\in$ [1,15] dB)															
MPSNR	27.494	33.246	39.197	31.966	38.270	42.050	25.067	34.101	36.090	36.089	36.176	38.644	39.787	36.932	39.997
MSSIM	0.712	0.948	0.940	0.866	0.967	0.984	0.533	0.915	0.945	0.942	0.951	0.963	0.973	0.977	0.974
MSAM	0.319	0.107	0.065	0.180	0.066	0.052	0.119	0.095	0.081	0.116	0.073	0.089	0.066	0.071	0.069
Time	0.000	0.271	1.893	50.020	13.118	0.188	22.637	8.675	26.502	1435.799	14.595	4.676	236.784	29.841	160.112
Case 3: Mixture of Gaussian and Impulse Noise															
MPSNR	11.286	29.526	29.582	21.966	30.781	21.786	19.847	28.092	30.620	27.330	30.796	30.845	32.249	30.759	33.531
MSSIM	0.093	0.891	0.852	0.446	0.856	0.700	0.300	0.715	0.817	0.720	0.846	0.821	0.887	0.805	0.890
MSAM	0.788	0.113	0.130	0.406	0.106	0.151	0.163	0.133	0.120	0.239	0.107	0.177	0.113	0.117	0.103
Time	0.000	0.393	1.971	52.623	14.164	0.197	23.290	25.309	70.988	1508.890	13.481	4.532	228.653	30.250	153.020
Case 4: Mixed Impulse Noise + Stripes + Deadlines (Impulse Ratio 20%)															
MPSNR	11.492	29.397	29.392	38.970	33.944	23.166	20.798	35.587	36.822	41.094	41.600	43.634	49.014	42.903	49.884
MSSIM	0.113	0.887	0.846	0.926	0.948	0.751	0.360	0.951	0.958	0.937	0.988	0.985	0.996	0.996	0.996
MSAM	0.799	0.117	0.134	0.206	0.078	0.144	0.156	0.092	0.083	0.200	0.045	0.051	0.034	0.043	0.033
Time	0.000	0.317	1.53	185.985	23.641	0.390	60.302	27.270	158.261	2348.431	9.611	2.811	46.316	27.931	42.589
Case 5: Mixed Impulse Noise + Stripes + Deadlines (Impulse Ratio 50%)															
MPSNR	7.564	23.417	24.203	24.490	25.422	15.763	15.170	29.918	33.150	36.340	34.378	35.530	37.937	36.500	43.255
MSSIM	0.027	0.709	0.593	0.603	0.811	0.401	0.122	0.854	0.921	0.933	0.951	0.932	0.976	0.985	0.986
MSAM	0.859	0.158	0.170	0.342	0.113	0.203	0.204	0.147	0.110	0.172	0.074	0.109	0.054	0.065	0.054
Time	0.000	0.378	1.569	52.503	13.587	0.185	23.167	8.553	26.311	1146.588	9.176	2.690	45.512	40.109	45.334
Case 6: Mixed Gaussian + Impulse + Stripes + Deadlines															
MPSNR	11.245	28.420	28.727	21.386	30.205	21.846	19.796	27.659	30.278	25.822	30.614	30.453	32.081	31.149	32.389
MSSIM	0.091	0.871	0.825	0.422	0.847	0.692	0.282	0.700	0.810	0.669	0.841	0.810	0.884	0.879	0.888
MSAM	0.789	0.122	0.142	0.431	0.119	0.158	0.168	0.157	0.133	0.292	0.111	0.203	0.117	0.129	0.109
Time	0.000	0.281	1.559	49.781	13.102	0.190	23.073	8.782	25.923	1178.679	9.476	2.652	46.392	94.718	40.634

https://doi.org/10.3390/rs18010135

Table 4. Quantitative indices of the WDC dataset with different methods under different cases.

Index	Noisy	QRNN3D	ILRNet	TNN	LRTF-DFR	FastHyDe	NGmeet	LRTV	LRTDTV	TCTV	E3DTV	RCTV	ETPTV	HNN	HLAWETV
Case 1: i.i.d. Gaussian Noise (σ = 0.1)															
MPSNR	20.713	30.260	34.145	25.296	35.137	35.708	35.681	30.235	32.508	29.570	32.947	35.228	35.587	33.536	35.874
MSSIM	0.407	0.877	0.933	0.615	0.953	0.956	0.955	0.837	0.904	0.807	0.922	0.950	0.956	0.933	0.957
MSAM	0.448	0.152	0.097	0.289	0.084	0.081	0.078	0.093	0.085	0.176	0.097	0.075	0.079	0.089	0.070
Time	0.000	0.727	3.27	592.092	78.912	0.701	105.312	117.485	1318.765	1303.256	53.931	13.624	1204.346	192.072	762.212
Case 2: non-i.i.d. Gaussian Noise (SNR $\in$ [1,15] dB)															
MPSNR	28.366	31.838	38.018	32.462	37.408	42.157	34.787	33.327	34.507	35.615	35.817	38.326	39.323	37.980	39.977
MSSIM	0.739	0.907	0.968	0.888	0.973	0.988	0.946	0.915	0.940	0.949	0.958	0.976	0.982	0.979	0.984
MSAM	0.305	0.131	0.073	0.140	0.062	0.047	0.085	0.081	0.070	0.093	0.075	0.057	0.048	0.057	0.050
Time	0.000	0.837	3.741	207.009	50.741	0.367	43.734	54.412	254.309	1246.466	58.075	14.880	1267.650	214.052	829.846
Case 3: Mixture of Gaussian and Impulse Noise															
MPSNR	11.425	28.677	30.002	23.137	33.716	22.258	19.693	28.954	31.447	27.914	31.331	33.607	34.015	32.005	34.230
MSSIM	0.113	0.860	0.856	0.508	0.937	0.779	0.462	0.796	0.880	0.746	0.891	0.928	0.939	0.906	0.940
MSAM	0.792	0.175	0.132	0.350	0.099	0.220	0.273	0.109	0.098	0.207	0.116	0.097	0.093	0.110	0.087
Time	0.000	0.926	3.84	387.397	54.884	0.371	39.023	47.733	100.479	645.359	58.416	12.720	365.138	91.146	258.205
Case 4: Mixed Impulse Noise + Stripes + Deadlines (Impulse Ratio 20%)															
MPSNR	11.385	28.521	29.795	34.978	35.197	22.935	20.019	34.553	34.717	36.545	38.742	42.170	43.950	43.840	44.913
MSSIM	0.122	0.857	0.852	0.923	0.969	0.795	0.468	0.949	0.946	0.930	0.984	0.992	0.995	0.995	0.996
MSAM	0.804	0.179	0.136	0.175	0.064	0.209	0.270	0.064	0.065	0.170	0.042	0.028	0.027	0.019	0.018
Time	0.000	0.836	3.485	319.254	92.515	0.749	131.510	136.032	1168.649	951.130	65.150	16.020	870.936	85.302	274.423
Case 5: Mixed Impulse Noise + Stripes + Deadlines (Impulse Ratio 30%)															
MPSNR	9.659	27.126	28.538	33.323	32.859	19.720	17.881	32.936	34.343	35.744	37.420	40.843	43.460	43.710	44.770
MSSIM	0.075	0.827	0.813	0.914	0.957	0.702	0.418	0.928	0.941	0.929	0.979	0.989	0.994	0.994	0.995
MSAM	0.851	0.193	0.151	0.176	0.077	0.262	0.302	0.084	0.069	0.167	0.050	0.030	0.028	0.020	0.020
Time	0.000	0.787	3.543	765.559	94.188	0.383	41.364	52.584	116.259	847.853	65.440	12.865	681.155	83.519	308.118
Case 6: Mixed Gaussian + Impulse + Stripes + Deadlines															
MPSNR	11.214	28.171	29.370	22.136	32.377	21.828	21.425	28.492	30.995	26.562	30.574	32.578	33.488	29.224	33.809
MSSIM	0.104	0.851	0.842	0.464	0.921	0.761	0.650	0.781	0.870	0.704	0.871	0.911	0.935	0.837	0.937
MSAM	0.800	0.184	0.145	0.389	0.119	0.231	0.254	0.134	0.110	0.249	0.128	0.115	0.102	0.162	0.097
Time	0.000	0.844	3.548	636.327	168.988	0.367	39.662	48.769	102.941	933.142	55.822	14.374	533.271	89.439	238.935

https://doi.org/10.3390/rs18010135

Table 5. Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) comparison on the real data.

Index	Noisy	QRNN3D	ILRNet	TNN	LRTF-DFR	FastHyDe	NGmeet	LRTV	LRTDTV	TCTV	E3DTV	RCTV	ETPTV	HNN	HLAWTV
Urban	72.50	39.42	36.95	47.97	31.79	34.98	38.43	35.81	55.24	48.63	36.71	36.85	36.97	50.46	28.51
Indian Pines	93.57	31.72	25.12	27.75	34.74	45.96	61.85	27.71	28.96	64.02	24.76	35.88	35.39	57.97	23.49

https://doi.org/10.3390/rs18010135

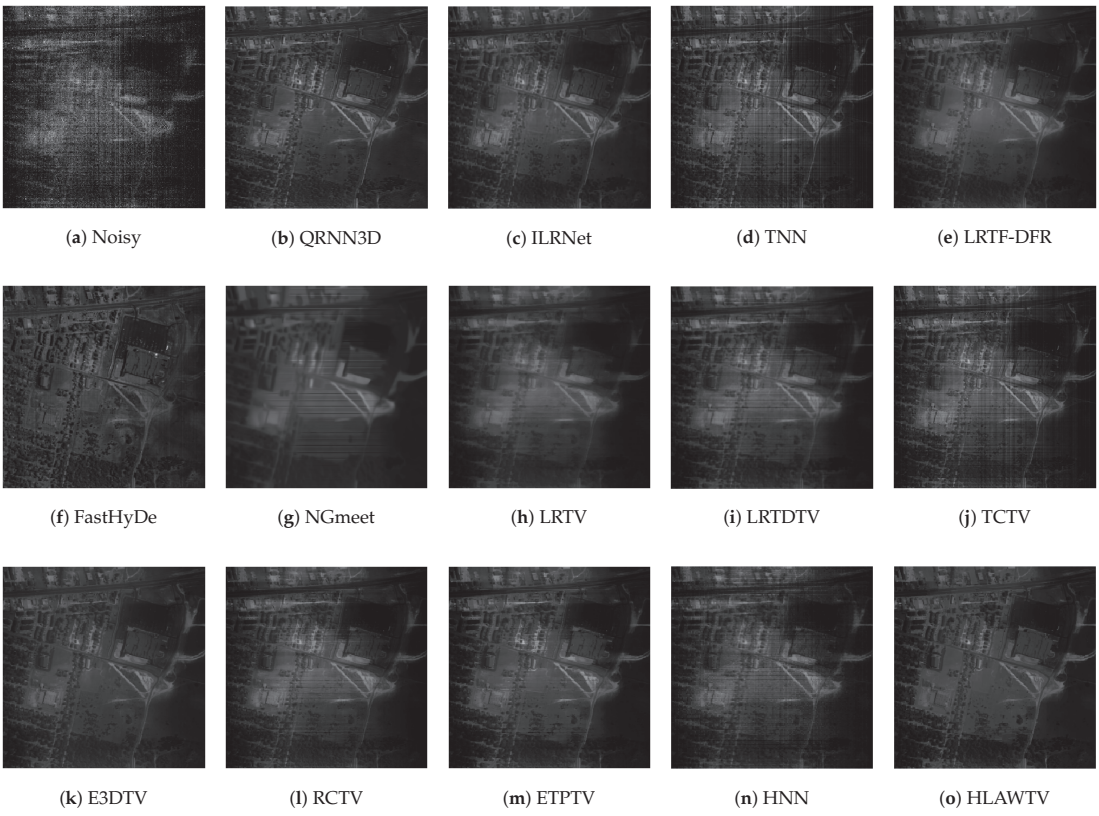


Figure 6. Denoising comparison on the real scenario at band 108 of Urban.

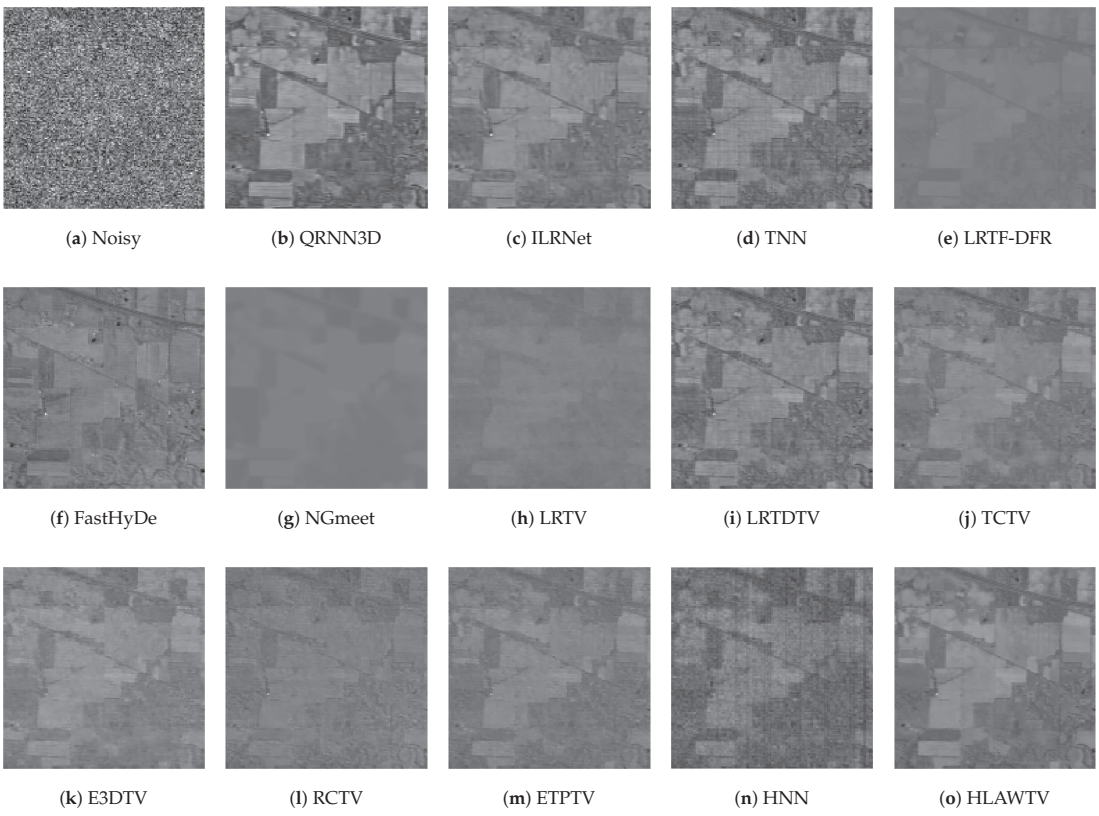
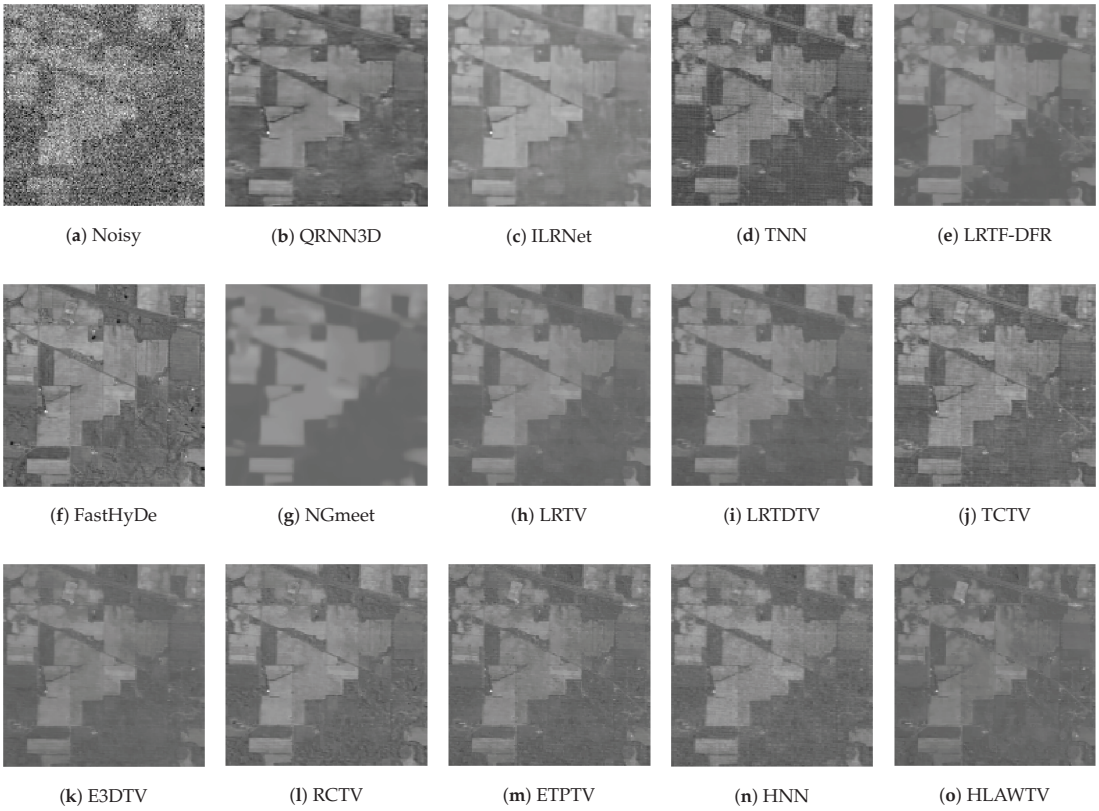


Figure 7. Denoising comparison on the real scenario at band 106 of Indian Pines.

https://doi.org/10.3390/rs18010135

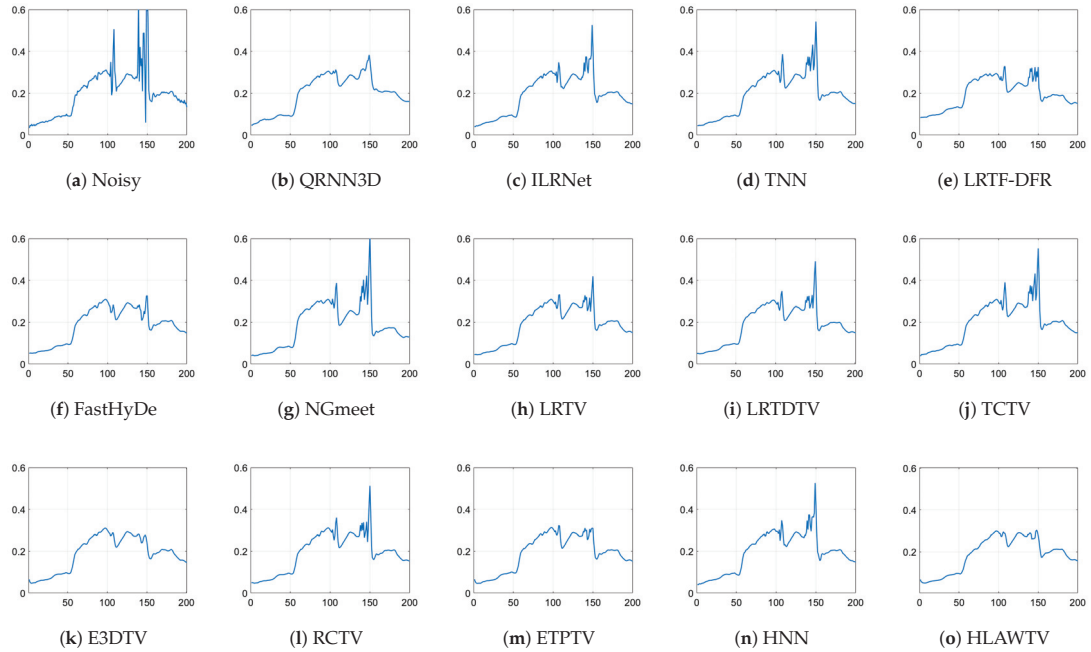


(**a**) Noisy (**b**) QRNN3D (**c**) ILRNet (**d**) TNN (**e**) LRTF-DFR

(**f**) FastHyDe (**g**) NGmeet (**h**) LRTV (**i**) LRTDTV (**j**) TCTV

(**k**) E3DTV (**l**) RCTV (**m**) ETPTV (**n**) HNN (**o**) HLAWTV

Figure 8. Denoising comparison on the real scenario at band 220 of Indian Pines.

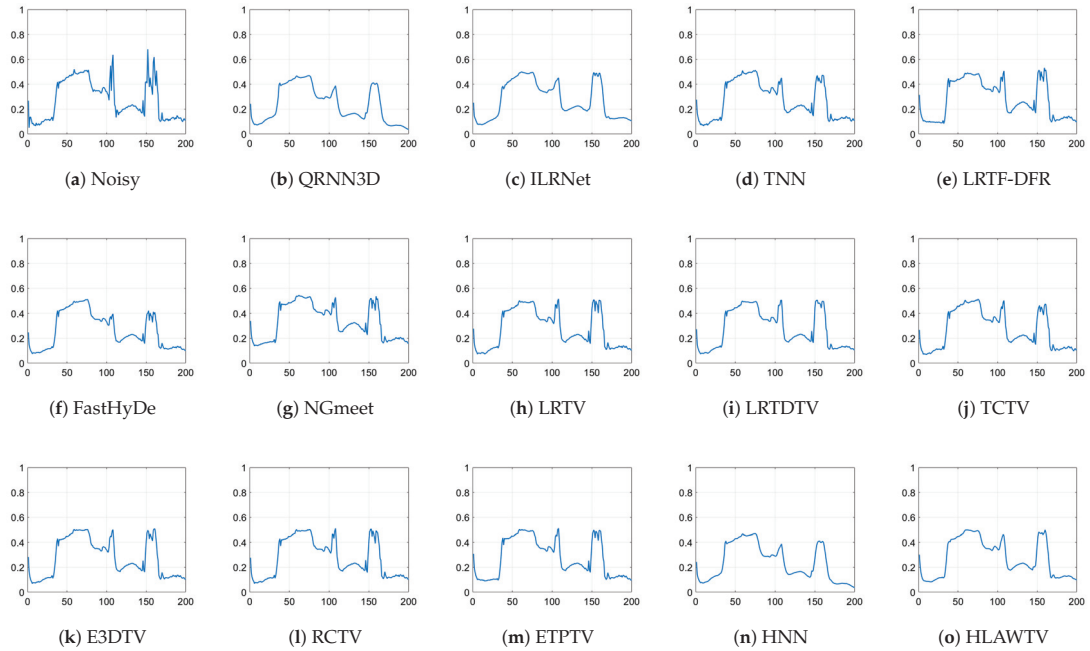


(**a**) Noisy (**b**) QRNN3D (**c**) ILRNet (**d**) TNN (**e**) LRTF-DFR

(**f**) FastHyDe (**g**) NGmeet (**h**) LRTV (**i**) LRTDTV (**j**) TCTV

(**k**) E3DTV (**l**) RCTV (**m**) ETPTV (**n**) HNN (**o**) HLAWTV

Figure 9. Spectral signatures at pixel (185, 132) in the Urban dataset before and after denoising by different methods.



(**a**) Noisy (**b**) QRNN3D (**c**) ILRNet (**d**) TNN (**e**) LRTF-DFR

(**f**) FastHyDe (**g**) NGmeet (**h**) LRTV (**i**) LRTDTV (**j**) TCTV

(**k**) E3DTV (**l**) RCTV (**m**) ETPTV (**n**) HNN (**o**) HLAWTV

Figure 10. Spectral signatures at pixel (125,20) in the Indian Pines dataset before and after denoising by different methods.

5. Discussion

5.1. Transferring to Other TV Regularizers

In this section, we investigate the portability of the proposed HLAWTV framework by integrating it into several TV-based regularizers. Therefore, three representative baselines with distinct TV formulations are selected as follows: LRTV applied 2-D TV to each band individually. LRTDTV considered pixel-variation distributions along the three tensor modes and assigned mode-specific weights to the corresponding gradient maps. RCTV adopted 2-D TV on representative coefficients of HSIs. For each baseline, we replace its original TV term with the HLAWTV regularization to obtain HLAW-LRTV, HLAW-LRTDTV, and HLAW-RCTV. The solver only needs to make local modifications to the TV minimization subproblem, while the overall optimization framework remains unchanged.

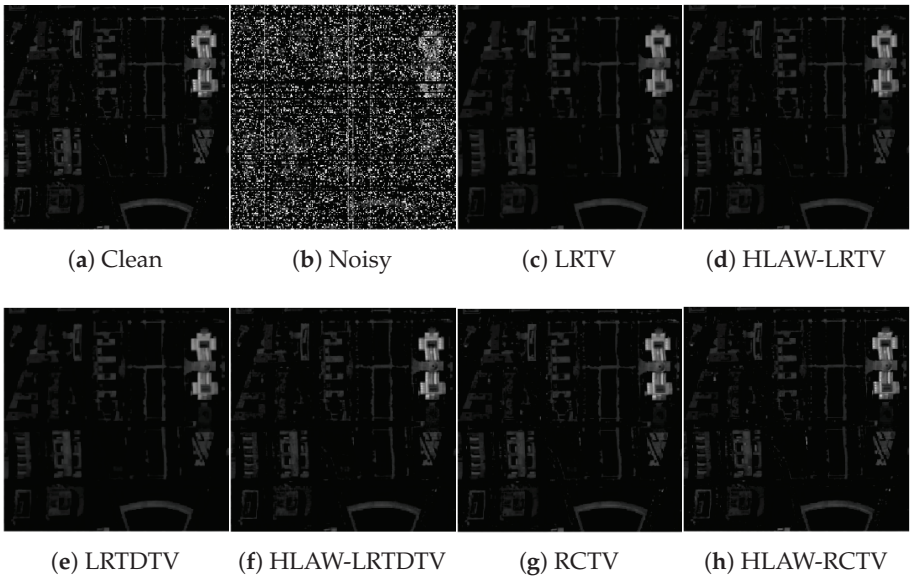
We conducted comparison experiments on the WDC dataset with the same noise settings as the synthetic data experiments, and the quantitative and visual results are shown in Table 6 and Figure 11. It can be seen that the denoising performance is greatly improved after using our HLAWTV framework, which indicates that the weighting scheme can be finely ported to other TV regularizers and significantly improve the performance of the original model.

Table 6. Performance comparison of all competing methods on the WDC dataset.

Case	Index	Noisy HSI	LRTV	HLAW-LRTV	LRTDTV	HLAW-LRTDTV	RCTV	HLAW-RCTV
Case 1	MPSNR	20.713	30.235	32.419	32.508	33.984	35.228	35.982
	MSSIM	0.407	0.837	0.894	0.904	0.924	0.950	0.954
	SAM	0.448	0.093	0.080	0.085	0.070	0.070	0.068
	Time (s)	0.000	117.485	53.616	1318.765	120.483	13.624	11.541
Case 2	MPSNR	28.366	33.327	35.967	34.507	35.514	38.326	39.297
	MSSIM	0.739	0.915	0.952	0.940	0.949	0.976	0.977
	SAM	0.305	0.081	0.071	0.070	0.064	0.057	0.052
	Time (s)	0.000	54.412	199.817	254.309	127.898	14.880	19.052

Table 6. *Cont.*

Case	Index	Noisy HSI	LRTV	HLAW-LRTV	LRTDTV	HLAW-LRTDTV	RCTV	HLAW-RCTV
Case 3	MPSNR	11.425	28.954	30.981	31.447	33.040	33.607	34.043
	MSSIM	0.113	0.796	0.863	0.880	0.908	0.928	0.929
	SAM	0.792	0.109	0.097	0.098	0.085	0.087	0.086
	Time (s)	0.000	47.733	199.955	100.479	837.385	12.720	11.162
Case 4	MPSNR	11.385	34.553	38.156	34.717	35.495	42.170	44.246
	MSSIM	0.122	0.949	0.976	0.946	0.950	0.992	0.995
	SAM	0.804	0.064	0.057	0.065	0.060	0.028	0.025
	Time (s)	0.000	136.032	54.355	1168.649	144.298	16.020	13.809
Case 5	MPSNR	9.659	32.936	36.625	34.343	35.396	40.843	43.245
	MSSIM	0.075	0.928	0.966	0.941	0.949	0.989	0.993
	SAM	0.851	0.084	0.066	0.069	0.060	0.030	0.026
	Time (s)	0.000	52.584	136.405	116.259	677.877	12.865	20.191
Case 6	MPSNR	11.214	28.492	30.395	30.995	32.177	32.578	33.050
	MSSIM	0.104	0.781	0.851	0.870	0.889	0.911	0.928
	SAM	0.800	0.134	0.125	0.110	0.109	0.115	0.107
	Time (s)	0.000	48.769	57.872	102.941	136.427	14.374	10.380



(**a**) Clean (**b**) Noisy (**c**) LRTV (**d**) HLAW-LRTV

(**e**) LRTDTV (**f**) HLAW-LRTDTV (**g**) RCTV (**h**) HLAW-RCTV

Figure 11. Comparison between original models and HLAW versions on Band 27 of the WDC dataset under case 5.

5.2. Ablation Study

To evaluate the effectiveness of each component in our proposed HLAWTV model, we perform a comprehensive ablation study on the WDC dataset under Case 5. Starting from the baseline LTV model, we progressively incorporate the following two key modules: adaptive weighting (AW) and HLSM. The results are shown in Table 7.

The LTV+AW model enables adaptive adjustment of the sparsity constraint between smooth and textured regions, thereby better preserving textures and edges in the image. However, due to the absence of global gradient sparsity modeling, its performance in terms of structural similarity and spectral accuracy remains suboptimal. In contrast, the LTV+HLSM model enhances global gradient sparsity by adjusting a heavy-tailed distribution on the gradients. Yet, it applies uniform penalties across the entire image without distinguishing between smooth and textured regions, leading to a lack of local adaptivity.

Although both AW and HLSM individually lead to slight decreases in SSIM and SAM compared to the baseline, their combination in the HLAWTV model results in the

best overall performance across all three metrics (MPSNR, MSSIM, and MSAM). This clearly demonstrates the strong coupling between the following two components: adaptive weights enhance local sensitivity, while the Hyper-Laplacian prior ensures global denoising robustness.

Moreover, applying our proposed adaptive weighting design to the baseline LTV significantly reduces computation time. This confirms the effectiveness of leveraging the conjugacy between the Hyper-Laplacian and Gamma distributions to derive efficient, data-driven weight estimation. The resulting balance between computational efficiency and denoising quality highlights the practical value of our method in real-world applications.

Table 7. Ablation study results.

Method	MPSNR	MSSIM	MSAM	Time (s)
LTV	43.460	0.994	0.028	681
LTV+AW	43.901	0.983	0.032	232
LTV+HLSM	43.923	0.985	0.026	319
HLAWTV	44.770	0.995	0.020	308

5.3. Parameter Analysis

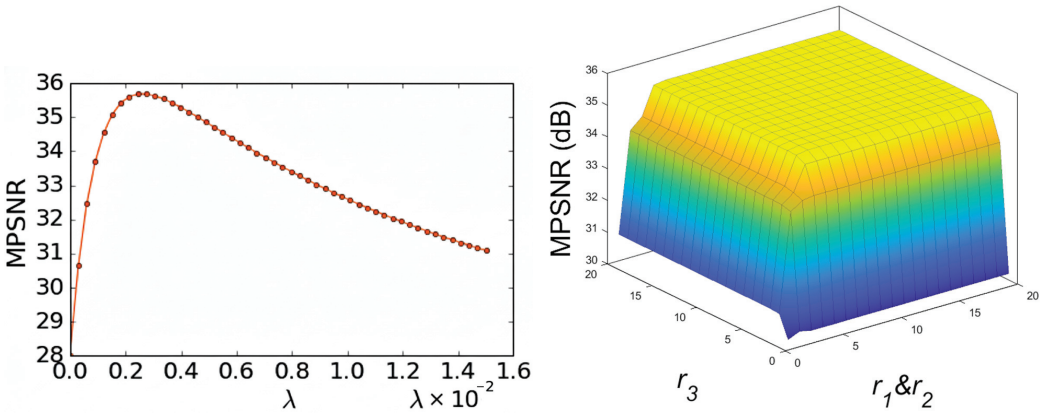
The robustness of the proposed HLAWTV model to its hyper-parameters λ and rank (r_1, r_2, r_3) is evaluated on the simulated data of Case 4. In this case, the denoising quality is monitored with MPSNR and MSSIM while a single parameter is varied and all others are fixed. Each experiment is repeated ten times to obtain statistically reliable curves.

First, we investigated the effect of the regularization parameter λ on denoising performance. As shown in Figure 12a, when λ is small, the HLAWTV regularization plays a minor role in the model, resulting in inadequate noise suppression and thus low MPSNR values with visible residual artifacts remaining in the denoised images. As λ gradually increases, more noise is effectively eliminated, leading to a rapid improvement in MPSNR and optimal preservation of image structures at the peak value. However, further increasing λ causes the regularization to become excessively strong, which in turn leads to over-smoothing and a loss of fine spatial details, causing the MPSNR to gradually decrease. Nevertheless, it is noteworthy that within a certain range of λ, the MPSNR remains consistently high, indicating that the model exhibits acceptable sensitivity to this parameter in practical applications. In summary, λ controls the strength of HLAWTV regularization, and its optimal value can be chosen based on the smoothness of the image as follows: smoother images generally require a larger λ, while images with more texture or detail benefit from a smaller λ.

The factor rank in HLAWTV is specified by a triplet (r_1, r_2, r_3), which controls the dimensionality of the low-rank subspace along the two spatial modes and the spectral mode. As shown in Figure 12b, varying these three ranks exhibits a clear pattern as follows: MPSNR increases rapidly as r_1, r_2, and r_3 grow from small values, reflecting the enhanced preservation of spatial and spectral structures. As $r1 \geq 5$ and $r2 \geq 5$ rise, the MPSNR plateaus and remains stable over a broad range of higher rank values, indicating that the model is robust to the choice of these parameters within this interval and does not easily overfit.

Overall, these parameter studies demonstrated the robustness of the proposed method and provided clear practical guidelines for selecting optimal parameter settings under various noise.

https://doi.org/10.3390/rs18010135



(**a**) (**b**)

Figure 12. Sensitivity analysis to parameters λ and r for the PaviaC dataset in case 5. (**a**) Sensitivity of MPSNR to parameter λ. (**b**) Sensitivity of MPSNR to parameters (r_1, r_2, r_3).

6. Conclusions

In this article, we proposed the Hyper-Laplacian Adaptive Weighted Total Variation model for hyperspectral image denoising. By integrating a Hyper-Laplacian Scale Mixture prior and leveraging the conjugacy between Gamma and Hyper-Laplacian distributions, our method efficiently computes adaptive weights that reflect the local sparsity and heavy-tailed characteristics of spatial-spectral gradients. The adaptive weighting, shared across all spectral bands at each pixel but varying spatially, enables precise noise suppression and detail preservation. Extensive experiments on both synthetic and real datasets demonstrated that HLAWTV consistently outperforms existing state-of-the-art methods. Furthermore, embedding our adaptive weighting mechanism into conventional total variation frameworks significantly enhances their effectiveness, confirming the broad applicability and robustness of our approach.

Author Contributions: Conceptualization, X.Y. and J.Z.; Methodology, J.Z.; Software, X.Y.; Validation, J.Z., S.F., and T.Z.; Formal analysis, T.Z.; Investigation, L.L.; Resources, J.Z.; Data curation, X.Y.; Writing—original draft preparation, X.Y.; Writing—review and editing, J.Z., S.F., T.Z., L.L., and X.H.; Visualization, T.Z.; Supervision, J.Z.; Project administration, J.Z.; Funding acquisition, X.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China under Grant 62072288; the Natural Science Foundation of Shandong Province under Grant ZR2021MF104 and ZR2021MF113; and the Open Project of the National Key Laboratory of Large Scale Personalized Customization System and Technology under Grant H&C-MPC-2023-02-04.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Huo, Y.; Dong, Y.; Wang, C.; Zhang, M.; Wang, H. Multi-scale memory network with separation training for hyperspectral anomaly detection. *Inf. Process. Manag.* **2026**, *63*, 104494. [CrossRef]
2. Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral Remote Sensing Data Analysis and Future Challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36. [CrossRef]
3. Shimoni, M.; Haelterman, R.; Perneel, C. Hyperspectral Imaging for Military and Security Applications: Combining Myriad Processing and Sensing Techniques. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 101–117. [CrossRef]

4. Ram, B.G.; Oduor, P.; Igathinathane, C.; Howatt, K.; Sun, X. A Systematic Review of Hyperspectral Imaging in Precision Agriculture: Analysis of Its Current State and Future Prospects. *Comput. Electron. Agric.* **2024**, *222*, 109037. [CrossRef]
5. Huo, Y.; Qian, X.; Li, C.; Wang, W. Multiple Instance Complementary Detection and Difficulty Evaluation for Weakly Supervised Object Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 6006505. [CrossRef]
6. Xiao, J.-L.; Huang, T.; Deng, L.; Wu, Z.; Vivone, G. A New Context-Aware Details Injection Fidelity with Adaptive Coefficients Estimation for Variational Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5408015. [CrossRef]
7. Dong, W.; Liu, S.; Xiao, S.; Qu, J.; Li, Y. ISPDiff: Interpretable Scale-Propelled Diffusion Model for Hyperspectral Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5519614. [CrossRef]
8. Qu, J.; Liu, X.; Dong, W.; Li, Y.; Meng, D. Progressive Multi-Iteration Registration-Fusion Co-Optimization Network for Unregistered Hyperspectral Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5519814. [CrossRef]
9. Huo, Y.; Cheng, X.; Lin, S.; Zhang, M.; Wang, H. Memory-Augmented Autoencoder With Adaptive Reconstruction and Sample Attribution Mining for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5518118. [CrossRef]
10. Buades, A.; Coll, B.; Morel, J.-M. A Non-Local Algorithm for Image Denoising. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Washington, DC, USA, 20–25 June 2005; IEEE Computer Society; Volume 2, pp. 60–65. [CrossRef]
11. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [CrossRef]
12. Gu, S.; Zhang, L.; Zuo, W.; Feng, X. Weighted Nuclear Norm Minimization with Application to Image Denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2862–2869. Greater Columbus Convention Center, Columbus, OH, USA, 23–28 June 2014; IEEE: New York, NY, USA, 2014. [CrossRef]
13. Maggioni, M.; Katkovnik, V.; Egiazarian, K.; Foi, A. Nonlocal Transform-Domain Filter for Volumetric Data Denoising and Reconstruction. *IEEE Trans. Image Process.* **2013**, *22*, 119–133. [CrossRef] [PubMed]
14. Maggioni, M.; Boracchi, G.; Foi, A.; Egiazarian, K. Video Denoising, Deblocking, and Enhancement Through Separable 4-D Nonlocal Spatiotemporal Transforms. *IEEE Trans. Image Process.* **2012**, *21*, 3952–3966. [CrossRef] [PubMed]
15. Yuan, Q.; Zhang, Q.; Li, J.; Shen, H.; Zhang, L. Hyperspectral Image Denoising Employing a Spatial–Spectral Deep Residual Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1205–1218. [CrossRef]
16. Wei, K.; Fu, Y.; Huang, H. 3-D Quasi-Recurrent Neural Network for Hyperspectral Image Denoising. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 2613–2627. [CrossRef] [PubMed]
17. Fu, G.; Xiong, F.; Lu, J.; Zhou, J.; Qian, Y. Spatial–Spectral Recurrent Transformer U-Net for Hyperspectral Image Denoising. *arXiv* **2023**, arXiv:2401.03885.
18. Pang, L.; Gu, W.; Cao, X. TRQ3DNet: A 3D Quasi-Recurrent and Transformer Based Network for Hyperspectral Image Denoising. *Remote Sens.* **2022**, *14*, 4598. [CrossRef]
19. Ye, J.; Xiong, F.; Zhou, J.; Qian, Y. Iterative Low-Rank Network for Hyperspectral Image Denoising. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5528015. [CrossRef]
20. Zhang, H.; He, W.; Zhang, L.; Shen, H.; Yuan, Q. Hyperspectral Image Restoration Using Low-Rank Matrix Recovery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4729–4743. [CrossRef]
21. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear Total Variation Based Noise Removal Algorithms. *Phys. Nonlinear Phenom.* **1992**, *60*, 259–268. [CrossRef]
22. He, W.; Zhang, H.; Zhang, L.; Shen, H. Total-Variation-Regularized Low-Rank Matrix Factorization for Hyperspectral Image Restoration. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 178–188. [CrossRef]
23. Wang, Y.; Peng, J.; Zhao, Q.; Leung, Y.; Zhao, Xi.; Meng, D. Hyperspectral Image Restoration via Total Variation Regularized Low-Rank Tensor Decomposition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1227–1243. [CrossRef]
24. Yang, F.; Chen, X.; Chai, L. Hyperspectral Image Destriping and Denoising Using Stripe and Spectral Low-Rank Matrix Recovery and Global Spatial–Spectral Total Variation. *Remote Sens.* **2021**, *13*, 827. [CrossRef]
25. Wang, H.; Peng, J.; Qin, W.; Wang, J.; Meng, D. Guaranteed Tensor Recovery Fused Low-Rankness and Smoothness. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 10990–11007. [CrossRef] [PubMed]
26. Chang, Y.; Yan, L.; Fang, H.; Luo, C. Anisotropic Spectral-Spatial Total Variation Model for Multispectral Remote Sensing Image Destriping. *IEEE Trans. Image Process.* **2015**, *24*, 1852–1866. [CrossRef] [PubMed]
27. Cai, W.; Jiang, J.; Ouyang, S. Hyperspectral Image Denoising Using Adaptive Weight Graph Total Variation Regularization and Low-Rank Matrix Recovery. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
28. Chen, Y.; Cao, W.; Pang, L.; Peng, J.; Cao, X. Hyperspectral Image Denoising via Texture-Preserved Total Variation Regularizer. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5516114. [CrossRef]
29. Li, D.; Chu, D.; Guan, X.; He, W.; Shen, H. Adaptive Hyper-Laplacian Regularized Low-Rank Tensor Decomposition for Hyperspectral Image Denoising and Destriping. *arXiv* **2024**, arXiv:2401.05682.

30. Xu, S.; Qiao, K.; Peng, J.; Zhao, Z. Hyperspectral Image Denoising by Low-Rank Models with Hyper-Laplacian Total Variation Prior. *Signal Process.* **2024**, *201*, 108733. [CrossRef]
31. Peng, J.; Xie, Q.; Zhao, Q.; Wang, Y.; Leung, Y.; Meng, D. Enhanced 3DTV Regularization and Its Applications on HSI Denoising and Compressed Sensing. *IEEE Trans. Image Process.* **2020**, *29*, 7889–7903. [CrossRef]
32. Peng, J.; Wang, H.; Cao, X.; Liu, X.; Rui, X.; Meng, D. Fast Noise Removal in Hyperspectral Images via Representative Coefficient Total Variation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5546017. [CrossRef]
33. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust Principal Component Analysis? *J. ACM* **2011**, *58*, 1–37. [CrossRef]
34. Fan, H.; Chen, Y.; Guo, Y.; Zhang, H.; Kuang, G. Hyperspectral Image Restoration Using Low-Rank Tensor Recovery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4589–4604. [CrossRef]
35. Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; Yan, S. Tensor Robust Principal Component Analysis with a New Tensor Nuclear Norm. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 925–938. [CrossRef] [PubMed]
36. Xue, J.; Zhao, Y.; Bu, Y.; Chan, J.C.; Kong, S.G. When Laplacian Scale Mixture Meets Three-Layer Transform: A Parametric Tensor Sparsity for Tensor Completion. *IEEE Trans. Cybern.* **2022**, *52*, 13887–13901. [CrossRef]
37. Garrigues, P.J.; Olshausen, B.A. Group Sparse Coding with a Laplacian Scale Mixture Prior. *Adv. Neural Inf. Process. Syst.* **2010**, *23*, 676–684.
38. Zuo, W.; Meng, D.; Zhang, L.; Feng, X.; Zhang, D. A Generalized Iterated Shrinkage Algorithm for Non-Convex Sparse Coding. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 1–8 December 2013; pp. 217–224. [CrossRef]
39. Zheng, Y.-B.; Huang, T.; Zhao, X.; Chen, Y.; He, W. Double-Factor-Regularized Low-Rank Tensor Factorization for Mixed Noise Removal in Hyperspectral Image. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8450–8464. [CrossRef]
40. Xu, S.; Yu, C.; Peng, J.; Chen, S.; Cao, X.; Meng, D. Haar Nuclear Norms With Applications to Remote Sensing Imagery Restoration. *IEEE Trans. Image Process.* **2025**, *34*, 6879–6894. [CrossRef]
41. Zhuang, L.; Bioucas-Dias, J.M. Fast Hyperspectral Image Denoising and Inpainting Based on Low-Rank and Sparse Representations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 730–742. [CrossRef]
42. He, W.; Yao, Q.; Li, C.; Yokoya, N.; Zhao, Q. Non-local Meets Global: An Integrated Paradigm for Hyperspectral Denoising. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 6868–6877. [CrossRef]
43. Huynh-Thu, Q.; Ghanbari, M. Scope of Validity of PSNR in Image/Video Quality Assessment. *Electron. Lett.* **2008**, *44*, 800. [CrossRef]
44. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
45. Mittal, A.; Moorthy, A.K.; Bovik, A.C. Blind/Referenceless Image Spatial Quality Evaluator. In Proceedings of the 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, USA, 6–9 November; pp. 723–727. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

remote sensing

MDPI

Article

SATRNet: Self-Attention-Aided Deep Unfolding Tensor Representation Network for Robust Hyperspectral Anomaly Detection

Jing Yang [1,2], Jianbin Zhao [1], Lu Chen [2,3,*], Haorui Ning [1] and Ying Li [4]

[1] School of Automation and Software Engineering, Shanxi University, Taiyuan 030031, China; yangjing199002@sxu.edu.cn (J.Y.); zhaojianbin@sxu.edu.cn (J.Z.); ninghaorui@sxu.edu.cn (H.N.)
[2] Key Laboratory of Evolutionary Science Intelligence of Shanxi Province, Shanxi University, Taiyuan 030006, China
[3] Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, China
[4] School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China; lybyp@nwpu.edu.cn
* Correspondence: chenlu@sxu.edu.cn

Abstract: Hyperspectral anomaly detection (HAD) aims to separate subtle anomalies of a given hyperspectral image (HSI) from its background, which is a hot topic as well as a challenging inverse problem. Despite the significant success of the deep learning-based HAD methods, they are hard to interpret due to their black-box nature. Meanwhile, deep learning methods suffer from the identity mapping (IM) problem, referring to the network excessively focusing on the precise reconstruction of the background while neglecting the appropriate representation of anomalies. To this end, this paper proposes a self-attention-aided deep unfolding tensor representation network (SATRNet) for interpretable HAD by solving the tensor representation (TR)-based optimization model within the framework of deep networks. In particular, a Self-Attention Learning Module (SALM) was first designed to extract discriminative features of the input HSI. The HAD problem was then formulated as a tensor representation problem by exploring both the low-rankness of the background and the sparsity of the anomaly. A Weight Learning Module (WLM) exploring local details was also generated for precise background reconstruction. Finally, a deep network was built to solve the TR-based problem through unfolding and parameterizing the iterative optimization algorithm. The proposed SATRNet prevents the network from learning meaningless mappings, making the network interpretable to some extent while essentially solving the IM problem. The effectiveness of the proposed SATRNet is validated on 11 benchmark HSI datasets. Notably, the performance of SATRNet against adversarial attacks is also investigated in the experimentation, which is the first work exploring adversarial robustness in HAD to the best of our knowledge.

Keywords: hyperspectral anomaly detection; deep unfolding; tensor representation; self-attention; interpretability; adversarial robustness

1. Introduction

Hyperspectral anomaly detection (HAD), which enables the identification of subtle, unknown materials within hyperspectral images (HSIs), represents a fundamental task in hyperspectral data analysis [1]. HAD typically operates in an unsupervised manner without requiring prior spectral knowledge of targets, aiming to detect objects whose spectral characteristics significantly differ from their background [2]. This unique capability

renders HAD particularly valuable for various surveillance applications, where specific targets are unknown beforehand, including but not limited to the following: detecting camouflaged personnel in battlefield scenarios [3], identifying pesticide residues during food safety inspections [4], and locating tumors in medical imaging [5].

In the nascent stage of HAD development, statistical-based methods predominated in this field. These approaches, rooted in signal detection theory, were primarily employed for background characterization. The Reed–Xiaoli (RX) detector [6], derived from a binary hypothesis testing framework under a multivariate Gaussian distribution assumption, subsequently became the most widely adopted HAD method due to its algorithmic simplicity and computational efficiency. Its numerous variants rapidly proliferated, including local-RX (LRX) [7], which exclusively utilizes pixels from dual concentric sliding windows to maintain uncontaminated background sets, weighted-RX (WRX) [8], which strategically implements weighting schemes to enhance background representation and improve the detection performance of RX, and kernel-RX (KRX) [9], which projects original data into higher-dimensional feature spaces through nonlinear mapping to amplify separability between background and anomalies. Further methodological innovations introduced using the signal-to-noise ratio (SNR) as an alternative optimization criterion for detector design [10]. Notably, the formulation of a hyperspectral target detection (HTD) dual theory specifically for HAD applications marked the first comprehensive exploration of their inherent duality relationships.

However, statistical assumption-based approaches for HAD algorithm design encounter a fundamental limitation: the unknown probability density function in practical applications. This challenge has led to the widespread investigation and application of sparse coding (SC) [11,12] and matrix low-rank representation (MLRR) [13–15] techniques in HAD research. These methods rely on carefully designed regularization terms that employ various norms to capture essential data characteristics, particularly low-rankness and sparsity properties. SC-based HAD approaches utilize over-complete dictionaries with multiple atoms to represent test pixels through ℓ_1-norm optimization, as demonstrated in background joint SC [11] and archetypal analysis with structured SC [12]. At the matrix level, common optimization objectives incorporate the Frobenius norm, $\ell_{2,1}$-norm, and nuclear norm to enforce desired structural properties. The low-rank and sparse representation (LRASR) framework [13] unifies these concepts by jointly modeling low-rank structures in coefficient matrices and sparse structures in residual matrices. By imposing nuclear norm constraints on the coefficient matrix, LRASR effectively captures the intrinsic HSI structure while isolating anomalies. Further developments, like graph and total variation regularized low-rank representation (GTVLRR) [14], augment the capabilities of LRASR through dual regularization terms that better preserve spatial structures. Recent advances include nonconvex-based HAD [15], where generalized shrinkage mapping achieves three objectives: approximating $\ell_{2,0}$ penalties through group sparsity constraints, simulating ℓ_0 TV penalties via TV regularization, and enforcing low-rank properties with nuclear norm penalties. This integrated approach enables comprehensive spatial correlation modeling.

The intrinsic data structure of HSI is fundamentally tensor-based, while most existing methods convert the HSI data into matrix form. This will potentially compromise the inherent spatial characteristics of HSI. In contrast, tensor representation (TR) methods preserve spatial information completely, consequently achieving enhanced HAD performance. The TR algorithms focus on tensor rank minimization, where (unlike matrix rank) multiple definitions exist, including but not limited to the following: CANDECOMP/PARAFAC (CP) rank [16], Tucker rank [17], ring rank [18], average rank [19], and tubal rank [20]. A CP rank-based TR model [16] decomposes backgrounds via mode-3 products of abundance tensors and endmember matrices while preserving spatial-spectral integrity. By

leveraging the inherent sparsity of core tensors generated through Tucker decomposition of gradient tensors, a specialized regularization term [17] is developed. This unified term simultaneously encodes both the low-rank and local smoothness constraints of background components, consequently improving HAD detection performance. The tensor ring decomposition (TRD) method [18] effectively exploits low-rank characteristics inherent in HSI background tensors. This approach decomposes a third-order background tensor through multilinear tensor products of three factor tensors, systematically uncovering low-rank features across different dimensionalities of the background tensor. A TR model [19] incorporates both low average rank and piecewise smooth constraints, employing the tensor nuclear norm (TNN) derived from tensor singular value decomposition (T-SVD) to induce low-rank properties on the background tensor. The tensor low-rank and sparse representation (TLRSR) framework [20] extends LRASR to the tensor domain, employing the weighted tensor nuclear norm (WTNN) as the convex relaxation for tensor tubal rank. This formulation enables computationally efficient Fourier-domain implementation while maintaining strong theoretical guarantees. The TLRSR further incorporates the $\ell_{F,1}$ norm to enforce structured group sparsity constraints. As a tensor dictionary representation model, the spatial invariant tensor self-representation (SITSR) model [21] merges the unfolded matrices of two representative coefficients and constrains them in a low-dimensional subspace. This strategy fully preserves the multi-dimensional structure. Beyond convex formulations, some nonconvex TR methods [22–24] have been explored as alternatives to unbiased estimators. These nonconvex formulations can be replaced to promote sparsity, group sparsity, and low-rankness, which have shown obvious advantages in HAD.

More recently, the deep learning (DL) technique has achieved significant success in multiple real-world tasks [25–28]. This rapid development in DL also prompted numerous researchers to apply it to HAD, as evidenced by the growing body of work in this field [29–35]. The well-known deep learning method is the autonomous HAD (Auto-AD) network [29], which uses a fully convolutional autoencoder (AE) network to reconstruct the background and uses the reconstruction error as the anomaly score. Auto-AD applies an adaptive weighted loss function to further reduce the reconstruction of anomalies by lowering the weight of anomalous pixels during training. The robust graph autoencoder (RGAE) [30] incorporates a graph regularization term to preserve the spatial geometric structure during the reconstruction process. RGAE utilizes a superpixel segmentation-based graph regularization term instead of constructing a graph Laplacian matrix to reduce the computational complexity. In addition, the $\ell_{2,1}$ norm is introduced into the loss function to improve the robustness to anomalies. The transformer-based AE framework (TAEF) [31] customizes the multilinear transformation decoder, which aims to improve the background learning capability of vanilla AE. In the preprocessing phase, TAEF uses a clustering-based method to suppress anomalies. The deep feature aggregation network (DFAN) [32] discovers the anomalies by enhancing the representation of various background patterns and calculating the Mahalanobis distance of the residual results. The gated transformer Network for HAD (GT-HAD) [33] is a dual-branch framework consisting of the anomaly-focused branch and the gated transformer block to accurately reconstruct the background while suppressing the anomalies. To further improve HAD performance, some studies have built deep networks based on the dual-window idea [34,35]. The spatial-spectral dual window mask transformer (S2DWMTrans) [34] makes full use of spatial-spectral information. Specifically, S2DWMTrans develops the local shallow feature extraction module to integrate adjacent spatial information, and the encoder block based on dual-window mask multihead self-attention further eliminates anomalies. Moreover, DirectNet [35] incorporates a sliding dual-window model to adaptively reconstruct the background, which increases the reconstruction error of anomalies.

Although the DL-based methods have become promising approaches in HAD, there are still some challenges. As mentioned in [36–38], the DL-based HAD methods widely suffer from the identity mapping (IM) problem, which refers to the phenomenon that the network reconstructs the whole HSI completely so that the anomalies are represented with imperceptible errors and, thus, are hard to detect. Moreover, the black-box nature of DL-based HAD methods makes them lack interpretability, which aggravates the IM issue to some extent. In fact, most of the existing DL-based methods, no matter how the architecture varies, learn meaningless mappings more or less due to the lack of interpretability and finally drive the network towards IM. On the contrary, traditional methods such as TR algorithms are fully interpretable and do not suffer from the IM problem, despite their limited generalization capability. Based on this observation, we naturally think of combining the TR methods with DL-based methods. Inspired by the deep unfolding technology, we propose a self-attention-aided deep unfolding tensor representation network called SATRNet for HAD. Specifically, SATRNet introduces Deep Unfolding Tensor Representation (DUTR), the architecture of which is designed by unfolding a customized TR algorithm. Due to the global modeling nature of DUTR, a Weight Learning Module (WLM) was developed to inject local details, enhancing the accuracy of background reconstruction. Moreover, a Self-Attention Learning Module (SALM) was introduced before DUTR to extract discriminative features, weakening the subsequent network's ability to reconstruct the anomalies. In this way, the working principles of the key components in SATRNet are consistent with the iterative algorithms of TR models, which avoids the meaningless feature mappings of the black-box models, and the network is guided towards the explicit objective. The IM problem can be effectively alleviated and thus enhance HAD performance and generalization capability. The main contributions of this work can be summarized as follows:

1. An interpretable network, the deep unfolding tensor representation network, is developed for HAD, solving the traditional TR model within a deep network. The combination of these two techniques integrates the interpretability of the TR model and the efficiency of the deep network for reliable anomaly detection. Moreover, a Weight Learning Module (WLM) is designed to learn local details to accurately reconstruct the background.
2. A Self-Attention Learning Module (SALM) is introduced to extract discriminative features, which avoids reconstructing the anomalies during the background reconstruction process. This also helps to effectively alleviate the identity mapping (IM) problem.
3. Apart from systematic evaluation against 11 HSI benchmark datasets, the adversarial attack experiment for HAD is performed in this work for the first time, indicating the adversarial robustness of the proposed SATRNet.

The rest of this paper is organized as follows. Notations, basic definitions, and related works are introduced in Section 2. Section 3 presents the proposed SATRNet model in detail. Comprehensive experimental results are given in Section 4. Discussions on the adversarial robustness of the proposed method are provided in Section 5. Finally, Section 6 concludes the paper.

2. Preliminaries

This section provides notations, basic definitions, and reviews of related works.

2.1. Notations and Basic Definitions

In this work, scalars, vectors, matrices, and tensors are represented by lowercase, bold lowercase, bold uppercase, and bold calligraphic uppercase letters, respectively, e.g., a, $\boldsymbol{a}$, $\boldsymbol{A}$, and $\boldsymbol{\mathcal{A}}$. For a third-order tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\boldsymbol{\mathcal{A}}_{(i)}$ is the ith frontal slice of $\boldsymbol{\mathcal{A}}$.

In addition, unfold($\boldsymbol{\mathcal{A}}$) converts $\boldsymbol{\mathcal{A}}$ into a matrix with dimensions $n_1 n_3 \times n_2$, and fold$(\cdot)$ is its inverse operator. The block circulant matrix bcirc($\boldsymbol{\mathcal{A}}$) is defined as

$$\text{bcirc}(\boldsymbol{\mathcal{A}}) = \begin{bmatrix} \boldsymbol{\mathcal{A}}_{(1)} & \boldsymbol{\mathcal{A}}_{(n_3)} & \cdots & \boldsymbol{\mathcal{A}}_{(2)} \\ \boldsymbol{\mathcal{A}}_{(2)} & \boldsymbol{\mathcal{A}}_{(1)} & \cdots & \boldsymbol{\mathcal{A}}_{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mathcal{A}}_{(n_3)} & \boldsymbol{\mathcal{A}}_{(n_3-1)} & \cdots & \boldsymbol{\mathcal{A}}_{(1)} \end{bmatrix}. \tag{1}$$

For unary operations, the ℓ_1-norm $\|\cdot\|_1$ and Frobenius norm $\|\cdot\|_F$ of $\boldsymbol{\mathcal{A}}$ are defined as

$$\|\boldsymbol{\mathcal{A}}\|_1 = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} |\boldsymbol{\mathcal{A}}_{i_1 i_2 i_3}|, \tag{2}$$

$$\|\boldsymbol{\mathcal{A}}\|_F = \sqrt{\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} |\boldsymbol{\mathcal{A}}_{i_1 i_2 i_3}|^2}. \tag{3}$$

For binary operations, the inner product $\langle \cdot, \cdot \rangle$ of tensors $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{C}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as

$$\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{C}} \rangle = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} \boldsymbol{\mathcal{A}}_{i_1 i_2 i_3} \boldsymbol{\mathcal{C}}_{i_1 i_2 i_3}. \tag{4}$$

Furthermore, the tensor–tensor product (t-product) [39] of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\boldsymbol{\mathcal{C}} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ is defined as

$$\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{C}} = \text{fold}(\text{bcirc}(\boldsymbol{\mathcal{A}}) \cdot \text{unfold}(\boldsymbol{\mathcal{C}})). \tag{5}$$

2.2. Related Works

(1) Alternating direction method of multiplier (ADMM) [40]: The ADMM method has been proven to be effective in dealing with constrained optimization problems. This algorithm decomposes the original problem into several manageable subproblems by introducing auxiliary variables and is able to derive closed-form solutions for each subproblem, thereby significantly improving computational efficiency. The algorithmic structure of ADMM is naturally suitable for handling structured sparsity constraints in high-dimensional data, a characteristic that makes it particularly valuable in the field of remote sensing image analysis [41]. Specifically, in the HAD task, ADMM is one of the preferred tools for solving optimization problems involved in MLRR [15] and TR [20] approaches.

(2) Deep unfolding: Recently, the deep unfolding technique has made significant progress in fusion [42], super-resolution [43], pansharpening [44], and so on. The deep unfolding network is designed according to a model-based iterative algorithm by cascading basic units, where each basic unit corresponds to a single iteration of the algorithm, and multiple basic units are connected into a deep network. By combining the interpretability of the iterative algorithm with the data-driven learning ability of the deep network, deep unfolding benefits from the advantages of both methods, i.e., better interpretability and generalization. As stated in [45,46], deep unfolding usually can achieve better performance compared with model-driven versus data-driven methods.

In the field of HAD, deep unfolding technology has also been explored recently. The basic idea of these approaches is to correspond the operators of traditional models with the network structure and transform the regularization parameters into learnable deep network parameters. For example, LRR-Net [46] is a kind of generalized deep unfolding network, in which the traditional MLRR model is unfolded into an unsupervised interpretable network. On this basis, by embedding sparsity, the LRR-Net framework

also demonstrates certain scalability. LRSR-I2Net-TV [47] is also an unfolding network based on MLRR. Different from LRR-Net, LRSR-I2Net-TV introduces the TV regularization term in an effort to compensate for the spatial correlation ignored in the MLRR model. However, these methods still have obvious limitations. Firstly, they are matrix-based methods that only realize the mining of spectral information. Although LRSR-I2Net-TV introduces TV regularization, it only provides limited utilization of spatial information, but inevitably increases the computational burden. Secondly, these approaches usually rely on original image features, which may hinder the accurate separation of anomalies and their backgrounds due to spectral variation and noise pollution. Therefore, it is necessary to develop more advanced solutions for deep unfolding-based HAD.

(3) Self-attention mechanism: Self-attention mechanism, as the core of transformer, is outstanding in constructing global spatial dependencies. In HAD, residual self-attention-based autoencoder (RSAAE) [48] exploits self-attention mechanism to highlight the features with high attention weight values, i.e., background features. Meanwhile, a low-rank constraint loss function is designed to preserve the low-rank property of the background in the latent space. In [49], a self-attention suppression (SAS) module is developed to reduce the impact of anomalies. By combining the SAS module with a multi-scale feature extraction strategy, the network can better minimize Please check that intended meaning has been retained the essential data features, thus improving the effect of background reconstruction.

3. Methodology

For an input HSI tensor $\mathcal{M} \in \mathbb{R}^{h \times w \times c}$, where h, w, and c express the height, width, and bands, respectively, the proposed SATRNet finally outputs the corresponding anomaly map $\boldsymbol{B} \in \mathbb{R}^{h \times w}$. Figure 1 shows a schematic of the proposed SATRNet. SATRNet primarily consists of two parts: Self-Attention Learning Module (SALM) and Deep Unfolding Tensor Representation (DUTR), which are introduced in Section 3.1 and Section 3.2, respectively.

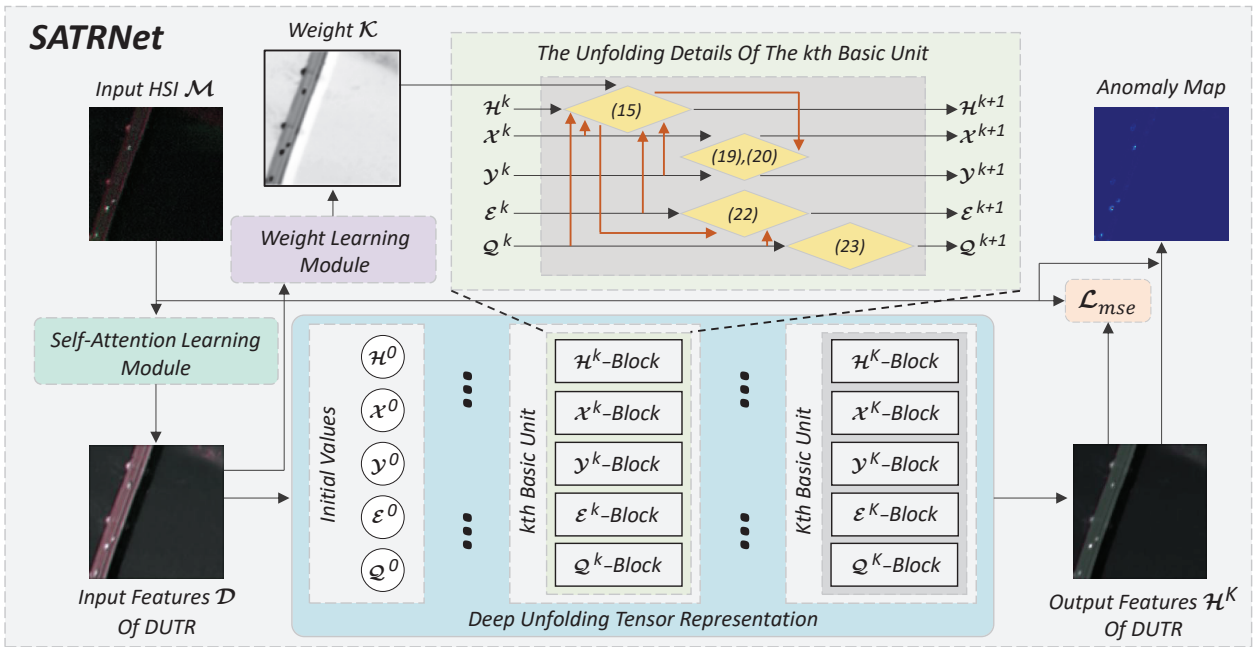


Figure 1. Framework of the proposed SATRNet. SALM first extracts discriminative features from input HSI $\mathcal{M}$. DUTR then takes the feature $\mathcal{D}$ as input and unfolds a tensor representation algorithm into an interpretable network. WLM injects local details via weight $\mathcal{K}$ to refine background reconstruction. The final output is the anomaly map. Note that the numbers in the yellow diamond shaped blocks represent relevant equations.

3.1. SALM: Self-Attention Learning Module

The raw pixels of the input HSI usually contain redundant information and are easily disturbed by noise; therefore, the learned features are generally used in the literature [50–52]. However, due to the limited receptive field of standard convolutions, global contexts are usually not adequately modeled. While dilated convolutions, deeper networks, or equivalent strategies can expand the receptive field, they risk overfitting. The attention mechanism enables the model to dynamically focus on important parts, which has been proven to be a powerful tool for feature extraction [53]. Thus, before the deep unfolding process, we introduce a Self-Attention Learning Module (SALM) to capture the discriminative and representative feature of the input HSI.

As shown in Figure 2, SALM includes a self-attention layer and a feed-forward network (FFN) layer. The FFN consists of 1×1 convolution, 3×3 convolution, and leaky rectified linear unit layers, following the common practice in [54]. In the self-attention layer, given the input tensor $\boldsymbol{\mathcal{M}}$, the output of this layer is $\boldsymbol{\mathcal{M}}^*$. The $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in \mathbb{R}^{hw \times c}$ are obtained through feature projection and patch embedding. A kernel operation based on Mercer's theorem [55] is employed to reconstruct $\boldsymbol{Q}$ and $\boldsymbol{K}$ for robust feature extraction. The kernel operation $\psi(\cdot)$ is expressed as follows (taking $\boldsymbol{Q}$ as an example):

$$\psi(\boldsymbol{Q}) = (1, \frac{(\boldsymbol{Q} - \bar{\boldsymbol{Q}})^2}{\sigma^{\frac{1}{2}}}, \frac{(\boldsymbol{Q} - \bar{\boldsymbol{Q}})^4}{\sigma^{\frac{1}{2}}}, \cdots, \frac{(\boldsymbol{Q} - \bar{\boldsymbol{Q}})^{2i}}{\sigma^{\frac{1}{2}}}). \tag{6}$$

where $\bar{\boldsymbol{Q}}$ represents the mean of $\boldsymbol{Q}$, and σ controls the scale of kernel operation $\psi(\cdot)$. Furthermore, the computational efficiency is improved by changing the order of multiplication in self-attention, i.e., first multiplying $\boldsymbol{K}$ and $\boldsymbol{V}$. The formula for calculating the self-attention matrix SA is as follows:

$$\mathrm{SA}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \mathrm{softmax}\left(\frac{\psi(\boldsymbol{K})^T \cdot \boldsymbol{V}}{\sqrt{d_s}}\right)\psi(\boldsymbol{Q}), \tag{7}$$

where d_s is the scaling factor. The theoretical justification and mathematical details for changing the order of multiplication can be found in [56]. In SATRNet, SALM not only effectively extracts distinctive features but also simplifies the subsequent process of separating background and anomalies. The feature extracted by SALM is represented as $\boldsymbol{\mathcal{D}}$.

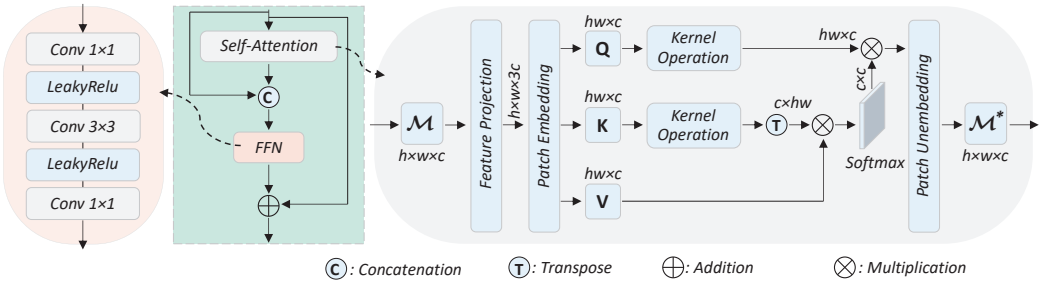


Figure 2. The detailed architecture of SALM, mainly including a self-attention layer and an FFN layer, aiming to extract distinctive and representative features.

3.2. DUTR: Deep Unfolding Tensor Representation

DUTR aims to design an effective deep architecture to solve the tensor-based HAD problem. The solving steps of TR-based models are essentially interpretable, which avoids producing meaningless mappings in deep models. Moreover, by incorporating physical priors (such as low-rankness and sparsity) in the tensor representation, the network can be guided to achieve explicit goals, which effectively alleviates the IM problem. In the follow-

ing, we first customize a TR-based optimization problem, then design an ADMM-based iterative algorithm for model solving, and finally, unfold the algorithm into a deep network.

3.2.1. TR Model Formulation and Optimization

We consider the robust principal component analysis (RPCA) problem for tensors, in which an input tensor $\boldsymbol{\mathcal{D}}$ is decomposed into a low-rank background tensor $\boldsymbol{\mathcal{H}}$ and a sparse anomaly tensor $\boldsymbol{\mathcal{E}}$. The tensor RPCA formulation for HAD is given by

$$\begin{aligned} &\min_{\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{E}}} \operatorname{rank}(\boldsymbol{\mathcal{H}}) + \lambda ||\boldsymbol{\mathcal{E}}||_0 \\ &\text{s.t.} \boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{E}}, \end{aligned} \tag{8}$$

where $\operatorname{rank}(\cdot)$ represents various tensor low-rank modeling strategies, $\boldsymbol{\mathcal{D}}$ is obtained through SALM, and λ is a trade-off parameter.

Tensor low-rank modeling strategies depend on the definition of tensor rank. The tubal rank, which is used in this study, is a good extension of the matrix rank and can characterize the low-rank property more accurately than other definitions. However, inducing low tubal rank usually uses TNN, which involves T-SVD computation and has been proven to be computationally expensive [19]. To avoid T-SVD computation, the t-product-based low-rank modeling strategy [39] is utilized to develop the fidelity term. Moreover, the ℓ_1-norm, which is the convex relaxation of ℓ_0-norm, is used, and the model is formulated as follows:

$$\begin{aligned} &\min_{\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}},\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{E}}} \frac{1}{2}\|\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{H}}\|_F^2 + \lambda \|\boldsymbol{\mathcal{E}}\|_1 \\ &\text{s.t.} \boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{E}}, \end{aligned} \tag{9}$$

where the tensor $\boldsymbol{\mathcal{H}}$ is decomposed into the t-product of two tensors $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$, i.e., $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{Y}}$, avoiding T-SVD calculation.

Now the model (9) can effectively model the background and obtain an interpretable mathematical expression. However, this modeling strategy still has limitations. Specifically, low-rank modeling is essentially based on global feature correlation, which lacks the ability to capture local details and is likely to lead to incomplete background learning. In addition, SALM tends to learn global spatial dependency and also ignores the local spatial features. It is not difficult to notice that the low-rank term in (9) directly processes the original data, rather than analyzing singular values in the transformed domain. This makes the low-rank part more susceptible and controllable. Therefore, we further introduce a Weight Learning Module (WLM) to capture local details and generate a weight $\boldsymbol{\mathcal{K}}$ to inject into the low-rank background term $\frac{1}{2}\|\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{H}}\|_F^2$. As shown in Figure 3, WLM is a simple convolutional neural network (CNN), which first uses an expanded depth-wise convolution with kernel size 3×3 to learn local information in $\boldsymbol{\mathcal{D}}$. Then, two 1×1 convolutions with a hidden Gaussian error linear unit (GELU) activation function [57] are used to generate enhanced local features. Finally, the sigmoid activation function is used to generate weight tensor $\boldsymbol{\mathcal{K}}$. The specific implementation of WLM is as follows:

$$\boldsymbol{\mathcal{K}} = \operatorname{Sigmoid}(\operatorname{Conv}_{1\times1}(\theta(\operatorname{Conv}_{1\times1}(\operatorname{DWConv}_{3\times3}(\boldsymbol{\mathcal{D}}))))), \tag{10}$$

where $\theta(\cdot)$ represents the GELU activation function.

For the weight tensor $\boldsymbol{\mathcal{K}}$, the optimization problem in (9) is further written as

$$\begin{aligned} &\min_{\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}},\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{E}}} \frac{1}{2}\|\sqrt{\boldsymbol{\mathcal{K}}} \odot (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{H}})\|_F^2 + \lambda \|\boldsymbol{\mathcal{E}}\|_1 \\ &\text{s.t.} \boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{E}}. \end{aligned} \tag{11}$$

where $\odot$ denotes the element-wise multiplication. Now, $\frac{1}{2}\|\sqrt{\mathcal{K}} \odot (\mathcal{X} * \mathcal{Y} - \mathcal{H})\|_F^2$ is injected with local details learned by WLM. This not only reduces the computational burden but also provides finer feature learning, which the traditional regularization terms struggle to do.

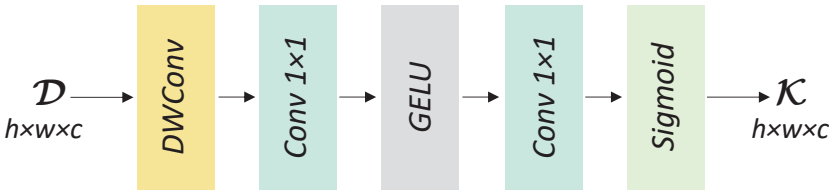


Figure 3. The detailed architecture of the WLM. This module aims to learn the enhanced local features and generate the refined weight of injecting local details into background reconstruction.

3.2.2. Solution to the TR Model

In this section, we will solve the optimization problem (11) by using ADMM. Under the ADMM framework, the penalty parameter $\mu^{k+1} = \rho\,\mu^k > 0$ is introduced to accelerate convergence, where $\rho > 1$, k is the number of iterations. Empirically, a larger value of ρ will lead to fewer iterations but lower accuracy, while a smaller ρ will lead to a higher number of iterations and higher accuracy. Next, we define the augmented Lagrangian function as follows:

$$\begin{aligned}\mathcal{L}(\mathcal{X},\mathcal{Y},\mathcal{H},\mathcal{E},\mathcal{Q}) = &\frac{1}{2}\|\sqrt{\mathcal{K}} \odot (\mathcal{X} * \mathcal{Y} - \mathcal{H})\|_F^2 + \lambda\|\mathcal{E}\|_1 \\ &+ \frac{\mu}{2}\|\mathcal{D} - \mathcal{H} - \mathcal{E}\|_F^2 + \langle \mathcal{Q}, \mathcal{D} - \mathcal{H} - \mathcal{E}\rangle,\end{aligned} \tag{12}$$

where $\mathcal{Q}$ represents the Lagrange multiplier.

We minimize the augmented Lagrangian function in (12) to obtain the optimal solution to the optimization problem (11) as follows:

$$(\mathcal{X}^\star,\mathcal{Y}^\star,\mathcal{H}^\star,\mathcal{E}^\star) = \arg\min_{\mathcal{X},\mathcal{Y},\mathcal{H},\mathcal{E}} \mathcal{L}(\mathcal{X},\mathcal{Y},\mathcal{H},\mathcal{E},\mathcal{Q}). \tag{13}$$

The joint optimization problem in (13) is difficult to solve directly; we decompose it into multiple subproblems in the kth iteration and solve them separately as follows:

(1) Update $\mathcal{H}$:

$$\begin{aligned}\mathcal{H}^{k+1} = \arg\min_{\mathcal{H}} &\frac{1}{2}\|\sqrt{\mathcal{K}} \odot (\mathcal{X}^k * \mathcal{Y}^k - \mathcal{H})\|_F^2 \\ &+ \frac{\mu^k}{2}\|\mathcal{D} - \mathcal{H} - \mathcal{E}^k\|_F^2 + \langle \mathcal{Q}^k, \mathcal{D} - \mathcal{H} - \mathcal{E}^k\rangle.\end{aligned} \tag{14}$$

The closed-form solution to this subproblem is given by

$$\mathcal{H}^{k+1} = \frac{\mathcal{K} \odot (\mathcal{X}^k * \mathcal{Y}^k) + \mathcal{Q}^k + \mu^k\mathcal{D} - \mu^k\mathcal{E}^k}{\mathcal{K} + \mu^k\mathcal{N}}, \tag{15}$$

where $\mathcal{N}$ is a tensor of all versions. Moreover, the involved t-product in (15) is computed following [39] to save computational cost.

(2) Update $\mathcal{X}$ and $\mathcal{Y}$:

$$\min_{\mathcal{X},\mathcal{Y}} \frac{1}{2}\|\sqrt{\mathcal{K}} \odot (\mathcal{X} * \mathcal{Y} - \mathcal{H})\|_F^2. \tag{16}$$

Let $\widehat{\boldsymbol{\mathcal{A}}}^{k+1}$ be the Fourier transform on $\boldsymbol{\mathcal{A}}^{k+1}$ along the third mode, i.e., $\widehat{\boldsymbol{\mathcal{A}}}^{k+1}(i,j,:) = \text{FFT}\{\boldsymbol{\mathcal{A}}^{k+1}(i,j,:)\}$. In the Fourier domain, the closed-form solutions of $\widehat{\boldsymbol{\mathcal{X}}}^{k+1}$ and $\widehat{\boldsymbol{\mathcal{Y}}}^{k+1}$ are given by

$$\widehat{\boldsymbol{\mathcal{X}}}^{k+1} = \widehat{\boldsymbol{\mathcal{H}}}_{(i)}^{k+1}\left(\widehat{\boldsymbol{\mathcal{Y}}}_{(i)}^{k}\right)^{H}\left(\widehat{\boldsymbol{\mathcal{Y}}}_{(i)}^{k}\left(\widehat{\boldsymbol{\mathcal{Y}}}_{(i)}^{k}\right)^{H}\right)^{\dagger}, \tag{17}$$

$$\widehat{\boldsymbol{\mathcal{Y}}}^{k+1} = \left(\left(\widehat{\boldsymbol{\mathcal{Y}}}_{(i)}^{k}\right)^{H}\widehat{\boldsymbol{\mathcal{Y}}}_{(i)}^{k}\right)^{\dagger}\left(\widehat{\boldsymbol{\mathcal{Y}}}_{(i)}^{k}\right)^{H}\widehat{\boldsymbol{\mathcal{H}}}_{(i)}^{k+1}, \tag{18}$$

where $(\cdot)^H$ and $(\cdot)^\dagger$ are the conjugate transpose and the pseudo-inverse operations. Then, $\boldsymbol{\mathcal{X}}^{k+1}$ and $\boldsymbol{\mathcal{Y}}^{k+1}$ are obtained using inverse Fourier transform as follows:

$$\boldsymbol{\mathcal{X}}^{k+1}(i,j,:) = \text{IFFT}\{\widehat{\boldsymbol{\mathcal{X}}}^{k+1}(i,j,:)\}, \tag{19}$$

$$\boldsymbol{\mathcal{Y}}^{k+1}(i,j,:) = \text{IFFT}\{\widehat{\boldsymbol{\mathcal{Y}}}^{k+1}(i,j,:)\}. \tag{20}$$

(3) Update $\boldsymbol{\mathcal{E}}$:

$$\begin{aligned}\boldsymbol{\mathcal{E}}^{k+1} = \arg\min_{\boldsymbol{\mathcal{E}}} \lambda^k\|\boldsymbol{\mathcal{E}}\|_1 + \frac{\mu^k}{2}\|\boldsymbol{\mathcal{D}} - \boldsymbol{\mathcal{H}}^{k+1} - \boldsymbol{\mathcal{E}}\|_F^2 \\ + \langle \boldsymbol{\mathcal{Q}}^k, \boldsymbol{\mathcal{D}} - \boldsymbol{\mathcal{H}}^{k+1} - \boldsymbol{\mathcal{E}}\rangle.\end{aligned} \tag{21}$$

The closed-form solution to this subproblem is given by

$$\boldsymbol{\mathcal{E}}^{k+1} = \boldsymbol{\mathcal{S}}_\tau\left(\frac{\boldsymbol{\mathcal{Q}}^k}{\mu^k} + \boldsymbol{\mathcal{D}} - \boldsymbol{\mathcal{H}}^k\right), \tag{22}$$

where $\boldsymbol{\mathcal{S}}_\tau$ is the element-wise soft-thresholding operator [35], i.e., $\boldsymbol{\mathcal{S}}_\tau = \text{sign}(\boldsymbol{\mathcal{S}}_{ijk})(|\boldsymbol{\mathcal{S}}_{ijk}| - \tau)_+$, where $(x)_+ = x$ if $x > 0$; otherwise, it is $(x)_+ = 0$, $\tau = \frac{\lambda^k}{\mu^k}$.

(4) Update $\boldsymbol{\mathcal{Q}}$:

$$\boldsymbol{\mathcal{Q}}^{k+1} = \boldsymbol{\mathcal{Q}}^k + \mu^k(\boldsymbol{\mathcal{D}} - \boldsymbol{\mathcal{H}}^{k+1} - \boldsymbol{\mathcal{E}}^{k+1}). \tag{23}$$

3.2.3. Deep Unfolding

Now, we unfold the above iterative algorithm to build an effective deep network. As shown in Figure 1, DUTR first receives the tensor $\boldsymbol{\mathcal{D}}$ obtained by SALM and then learns local details to produce the weight tensor $\boldsymbol{\mathcal{K}}$ via WLM. Finally, the tensors $\boldsymbol{\mathcal{D}}$ and $\boldsymbol{\mathcal{K}}$ are input into the deep network consisting of K consecutive basic units; the value of K is typically smaller than that of the iteration number for traditional algorithms. Each basic unit in DUTR consists of $\boldsymbol{\mathcal{H}}$-block, $\boldsymbol{\mathcal{X}}$-block, $\boldsymbol{\mathcal{Y}}$-block, $\boldsymbol{\mathcal{E}}$-block, and $\boldsymbol{\mathcal{Q}}$-block, where each block corresponds to a subproblem-solving step in the iterative algorithm. Such a network architecture design gives each module of the network explicit physical meaning, therefore enhancing the network interpretability and essentially alleviating the IM problem. Figure 4 shows the relationship between the optimization algorithms and the design of network blocks more clearly. Each block is regarded as the interpretable mapping of the optimization algorithm, and the basic units formed by the connection of these network blocks form an interpretable network. In addition, unlike traditional ADMM, in which most of the parameters are predefined or calculated manually, all the parameters are learnable and can be updated automatically by the network in the proposed SATRNet. For example, the update of penalty parameter μ is independent of ρ but is adjusted by backpropagation during the network training process. Similarly, λ does not need to be configured manually but is learned adaptively from the network.

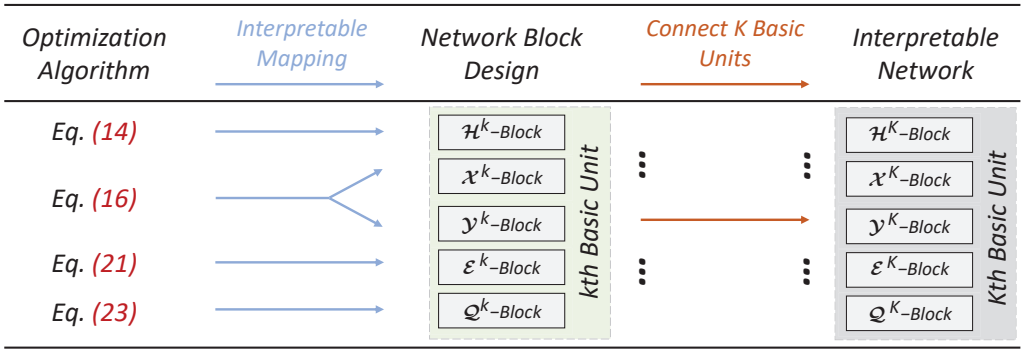


Figure 4. Relationship diagram between optimization algorithms and network blocks.

3.3. SATRNet for HAD

SATRNet is optimized by the widely used Adam optimizer and the backpropagation strategy. The mean squared error (MSE) is a simple and effective loss function commonly used in HAD [32]. As the objective function for optimization of SATRNet, MSE is given by

$$\mathcal{L}_{mse} = \frac{1}{hw}\sum_{i=1}^{h}\sum_{j=1}^{w}\|(\mathbf{x}_{i,j} - \tilde{\mathbf{x}}_{i,j})\|_2^2, \tag{24}$$

where $\mathbf{x}_{i,j}$ and $\tilde{\mathbf{x}}_{i,j}$ represent the pixels at location (i,j) of $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{H}}^K$. Finally, the anomaly detection map $\boldsymbol{B}$ can be obtained by

$$\boldsymbol{B}(h,w) = \sqrt{\sum_{i=1}^{c}|\boldsymbol{\mathcal{P}}(h,w,c)|^2}. \tag{25}$$

where $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{M}} - \boldsymbol{\mathcal{H}}^K$.

The SATRNet for HAD is summarized in Algorithm 1.

Algorithm 1 SATRNet for HAD

Input: HSI tensor $\boldsymbol{\mathcal{M}}$, parameter K.

Initialize: $\boldsymbol{\mathcal{H}}^0 = \boldsymbol{\mathcal{E}}^0 = \boldsymbol{\mathcal{Q}}^0 = 0, \boldsymbol{\mathcal{X}}^0 = \boldsymbol{\mathcal{Y}}^0 = 1e-1$.
Step 1: Training Process:
while *training process* **do**
  Obtain $\boldsymbol{\mathcal{D}}$ by Self-Attention Learning Module;
  while $i < K$ **do**
    Update $\boldsymbol{\mathcal{H}}$ by (15);
    Update $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{Y}}$ by (19) and (20);
    Update $\boldsymbol{\mathcal{E}}$ by (22);
    Update $\boldsymbol{\mathcal{Q}}$ by (23);
    $i = i + 1$.
  end while
  Compute the loss function by (24);
  Perform network backward process.
end while
Step 2: Detection Process:
Obtain the low-rank background tensor $\boldsymbol{\mathcal{H}}^K$.
Obtain the anomaly detection map $\boldsymbol{B}$ by (25).

Output: Anomaly detection map $\boldsymbol{B}$.

4. Results

In this section, we assess the efficacy of the proposed SATRNet in comparison with a number of advanced HAD methods across 11 HSI datasets.

4.1. Datasets

To systematically evaluate the performance of the proposed SATRNet approach, one synthetic and 10 real HSI datasets were employed, where the synthetic Salinas-I dataset was provided by Tan et al. [58]. The Salinas-II dataset was collected in the Salinas Valley, CA, USA, where farmhouses are regarded as anomalies [59]. The Urban-I, Urban-II, and Urban-IV datasets were collected from the widely used ABU dataset [22]. In the Urban-I dataset, multiple storage tanks within an oil refinery are labeled as anomalies. In the Urban-II and Urban-IV datasets, buildings are considered as anomalies. The UHAD dataset [60] is provided by Liu et al. and represents a park scene where pedestrians on the grass are regarded as anomalous, tiny targets. The HAD-100 dataset is a test scene from a large-scale HAD dataset [61] proposed by Li et al., where man-made objects are considered as anomalies. The Pavia dataset [32] was collected in Pavia, Italy, where three cars on the bridge and two patches of bare soil near the bridge-pier are labeled as anomalies. The HYDICE dataset [32] was acquired using the hyperspectral digital imagery collection experiment (HYDICE), and the anomalies are composed of vehicles and roofs under different backgrounds. The anomalies in the Segundo dataset [62] are storage tanks and towers, and the anomaly pixels account for 6.45% of the whole image, which is significantly higher than the above datasets. The background of the SpecTIR dataset [63] is mainly roads and vegetation, while anomalies are composed of man-made colored square textiles of different sizes. The details of these datasets are listed in Table 1, and the pseudo-color images and ground truth images are exhibited in Figure 5.

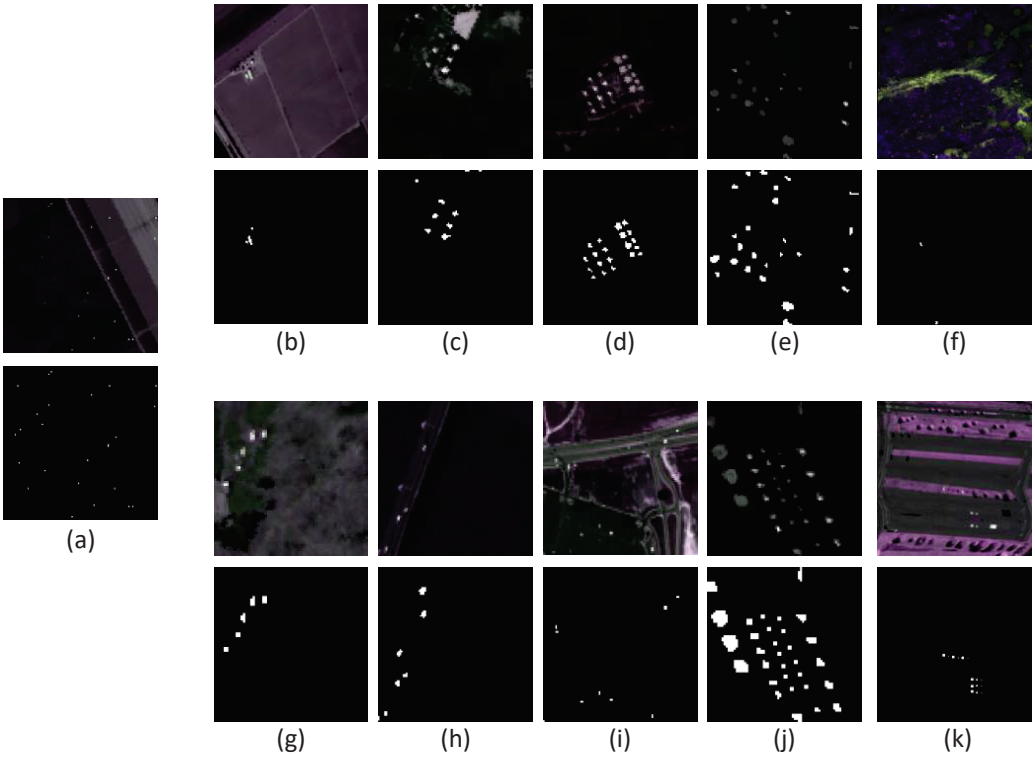


Figure 5. Pseudo-color images and ground truth of 11 datasets. (**a**) Salinas-I. (**b**) Salinas-II. (**c**) Urban-I. (**d**) Urban-II. (**e**) Urban-IV. (**f**) UHAD. (**g**) HAD-100. (**h**) Pavia. (**i**) HYDICE. (**j**) Segundo. (**k**) SpecTIR.

Table 1. Detailed information for 11 HSI datasets.

Datasets	Location	Spatial Size	Number of Bands	Sensors	Size of Targets	Proportion
Salinas-I [58]	Salinas Valley	120 × 120	204	AVIRIS	Small	0.17%
Salinas-II [59]	Salinas Valley	150 × 200	204	AVIRIS	Small and large	0.14%
Urban-I [22]	Texas Coast	100 × 100	204	AVIRIS	Small and large	0.67%
Urban-II [22]	Texas Coast	100 × 100	207	AVIRIS	Small and large	1.55%
Urban-IV [22]	Los Angeles	100 × 100	205	AVIRIS	Large	2.72%
UHAD [60]	Park	250 × 250	61	HyperspecI	Small	0.05%
HAD-100 [61]	Grassland	64 × 64	276	AVIRIS-NG	Large	0.65%
Pavia [32]	Pavia	100 × 100	102	ROSIS	Small and large	0.68%
HYDICE [32]	Urban	80 × 100	162	HYDICE	Small	0.23%
Segundo [62]	Segundo	75 × 75	224	AVIRIS	Small and large	6.45%
SpecTIR [63]	New York	180 × 180	120	SHARE	Small and large	0.22%

4.2. Compared Methods and Parameter Settings

In the experiments, 7 advanced compared methods were selected to evaluate the detection performance of the developed SATRNet method. These compared methods include two statistical theory-based methods, i.e., RX [6] and 2S-GLRT [64]; one MLRR-based method, i.e., LRASR [13]; three TR-based methods, i.e., PCA-TLRSR [20], CF2-GNBRL [24], and LARTVAD [19]; and two DL-based methods, i.e., Auto-AD [29] and DFAN [32]. All parameters involved in the above compared methods were used according to the authors' recommended values or tuned to the optimal settings. As shown in Tables 2 and 3, we give the parameter values of all compared algorithms for each dataset.

The proposed SATRNet was computed on NVIDIA GPUs, GeForce RTX 4060, and Intel Core i7 processors, implemented in the Pytorch framework on the Windows 10 operating system. The network was trained using the Adam optimizer, and the hyperparameters were set as follows: the learning rate was 0.001, the batch size was 1, and the training epochs were 300 and 500 for the HYDICE and SpecTIR datasets, respectively, and 100 for the other datasets. In addition, SATRNet involves a key parameter, i.e., the number of basic units K, which is discussed in detail in Section 4.5.

Table 2. Parameter values for the compared methods on the first 6 datasets.

Methods	Parameters	Salinas-I	Salinas-II	Urban-I	Urban-II	Urban-IV	UHAD
RX	/	/	/	/	/	/	/
2S-GLRT	(w_{in}, w_{out})	(1, 3)	(11, 15)	(11, 13)	(11, 13)	(11, 13)	(11, 13)
LRASR	(p, k, λ, β)	(20, 15, 0.1, 0.1)	(20, 15, 0.1, 0.1)	(20, 15, 0.1, 0.01)	(20, 15, 0.1, 0.1)	(20, 15, 0.1, 0.1)	(20, 15, 0.1, 0.1)
PCA-TLRSR	$(d, \lambda_1, \lambda_2, \omega)$	(10, 0.06, 0.1, 5)	(5, 0.1, 0.5, 5)	(5, 0.06, 0.1, 5)	(5, 0.06, 0.1, 5)	(10, 0.06, 0.1, 5)	(5, 0.06, 0.01, 5)
LARTVAD	(K, λ, β)	(60, 0.1, 2)	(60, 0.1, 2)	(80, 0.1, 0.1)	(80, 0.1, 0.1)	(80, 0.1, 0.1)	(80, 0.1, 0.1)
CF2-GNBRL	$(x_{PCA}, \lambda_1, \lambda_2, \beta)$	(2, 0.01, 0.01, 0.05)	(2, 0.01, 0.01, 0.05)	(2, 0.01, 0.01, 0.05)	(2, 0.01, 0.01, 0.05)	(2, 0.01, 0.01, 0.05)	(5, 0.01, 0.01, 0.05)
Auto-AD	/	/	/	/	/	/	/
DFAN	$(MinPts, Eps, Epochs, LR)$	(3, 0.5, 30, 1×10^{-5})	(3, 0.5, 30, 1×10^{-5})	(4, 0.5, 20, 1×10^{-4})	(3, 0.5, 30, 1×10^{-5})	(6, 0.5, 20, 1×10^{-2})	(3, 0.5, 30, 1×10^{-5})

Table 3. Parameter values for the compared methods on the last 5 datasets.

Methods	Parameters	HAD-100	Pavia	HYDICE	Segundo	SpecTIR
RX	/	/	/	/	/	/
2S-GLRT	(w_{in}, w_{out})	(11, 13)	(11, 13)	(11, 13)	(11, 13)	(11, 13)
LRASR	(p, k, λ, β)	(20, 15, 0.01, 0.1)	(20, 15, 0.01, 0.1)	(20, 15, 0.01, 0.1)	(20, 15, 0.01, 0.1)	(20, 15, 0.1, 0.1)
PCA-TLRSR	(d, λ_1, λ_2, ω)	(10, 0.06, 0.1, 5)	(10, 0.06, 0.01, 5)	(10, 0.06, 0.01, 5)	(5, 0.06, 0.1, 5)	(5, 0.06, 0.1, 5)
LARTVAD	(K, λ, β)	(40, 0.1, 10)	(80, 0.1, 0.1)	(80, 3, 2)	(80, 0.1, 0.1)	(80, 0.1, 0.1)
CF2-GNBRL	(λ_1, λ_2, β)	(5, 0.01, 0.01, 0.05)	(2, 0.01, 0.01, 0.05)	(2, 0.01, 0.01, 0.05)	(2, 0.01, 0.01, 0.05)	(2, 0.01, 0.01, 0.05)
Auto-AD	/	/	/	/	/	/
DFAN	($MinPts$, Eps, $Epochs$, LR)	(3, 0.5, 30, 1×10^{-5})	(6, 0.6, 20, 1×10^{-4})	(3, 0.5, 30, 1×10^{-2})	(3, 0.5, 30, 1×10^{-5})	(3, 0.5, 30, 1×10^{-5})

4.3. Detection Performance

The detection performance of each method was evaluated using the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) [65]. In addition, 3D-ROCs [66], generated by ROCs and separability box plots, were also introduced to further evaluate detection performance effectively.

Figures 6–8 show the visual detection results of different HAD methods on 11 HSI datasets. The RX detector can detect most anomalies, but it has a dark brightness and slight false alarms. Its performance on the HAD-100 dataset is commendable, with obvious anomalies detected. Overall, the performance of RX is relatively balanced. For the 2S-GLRT algorithm, a few anomalies are correctly detected. The detection map shows erroneous blocky anomalies; this is unacceptable, although the background is relatively pure. This result may be attributed to the sliding dual-window strategy of 2S-GLRT to reduce false alarms, which also makes it unable to handle features of different scales. As a milestone of the MLRR-based method, the detection results of LRASR show that it has poor ability in the degree of background suppression, and on the Urban-I dataset, it suffers serious interference from noise. Owing to the powerful representation capacity of TR, the detection performance of PCA-TLRSR is better than the above algorithms, and more anomalies are detected, but it also suffers from eliminating background information. In terms of visual effects, LARTVAD has good performance, similar to PCA-TLRSR, but its background suppression is slightly better than PCA-TLRSR. CF2-GNBRL has the same limitation on false detections as PCA-TLRSR and LARTVAD, which is more serious on some datasets, such as Urban-I and HAD-100. The detection results of Auto-AD on all datasets show relatively pure backgrounds, but they are affected by the IM issue, and significantly fewer anomalies are detected. Although DFAN can highlight anomalies well, there are lots of false alarms on all datasets, which makes it more difficult to locate anomalies. On the contrary, the proposed SATRNet can greatly reduce the erroneous detection on the background and maintain an appearance similar to the ground truth maps.

Figure 9 shows the ROC curves of different HAD methods on all datasets. The red curve regarding the proposed SATRNet has both high detection probability P_D and low false alarm probability P_F. Therefore, it is closer to the upper left corner than other HAD methods, which stands for excellent HAD performance. Figure 10 shows the 3D-ROC curves obtained by each method on 11 datasets, respectively. As can be seen from Figure 10, the 3D-ROC curves of the SATRNet are closer to the upper right corner than those of other

detectors, which means that SATRNet has the best detection performance. Furthermore, Figure 11 shows the box plots of the separation of backgrounds and anomalies by different methods on all datasets. The background boxes (i.e., purple boxes) and anomaly boxes (i.e., black boxes) are displayed by 10% to 90% of the images, and the height of the boxes indicates the degree of background suppression and anomaly detection of different algorithms. It can be seen from Figure 11 that although the gap between the background and anomaly boxes of the proposed method is not always the largest, it achieves the best background suppression without overlapping. In summary, SATRNet exhibits a stronger ability to separate background from anomalies on 11 datasets.

To quantitatively measure the accuracy of the proposed method, the AUC and running time of each method on 11 datasets are listed in Table 4, where the best results are shown in bold. From Table 4, it can be seen that SATRNet achieves the highest AUC values on all datasets. For the Salinas-II and HAD-100 datasets, the RX method is tied for first place with our method, but SATRNet is highly competitive due to better background suppression and detection rates in visualization, which is more conducive to accurate anomaly identification. In terms of running time, the RX algorithm has the shortest running time, but the detection results are not satisfactory. Although the time efficiency of the proposed SATRNet ranks middle among all methods, such time cost is acceptable given the excellent HAD ability.

Table 4. Quantitative results over different HAD algorithms on 11 datasets.

Datasets	Metrics	RX	2S-GLRT	LRASR	PCA-TLRSR	LARTVAD	CF2-GNBRL	Auto-AD	DFAN	SATRNet
Salinas-I	AUC	0.8073	0.9296	0.9269	0.9884	0.9780	0.9833	0.9876	0.9913	**0.9924**
	Time (s)	**0.6036**	45.9294	60.3541	4.0059	31.1208	150.4626	9.9758	30.9004	13.3504
Salinas-II	AUC	**0.9994**	0.9754	0.9983	0.9974	0.9945	0.9881	0.9965	0.9677	**0.9994**
	Time (s)	**1.0782**	626.3742	58.0547	19.7640	86.0172	212.4481	26.6702	60.4936	26.3504
Urban-I	AUC	0.9907	0.9206	0.9621	0.9892	0.9893	0.9474	0.9844	0.9898	**0.9909**
	Time (s)	**0.4242**	134.8021	20.0812	2.2647	24.3573	170.7258	19.7661	32.7611	9.6824
Urban-II	AUC	0.9946	0.7354	0.9482	0.9947	0.9983	0.9968	0.9993	0.9652	**0.9994**
	Time (s)	**0.3719**	191.4367	24.3287	2.4378	26.4354	170.4741	9.9373	28.4253	12.5209
Urban-IV	AUC	0.9887	0.9099	0.6120	0.9856	0.9908	0.9952	0.9924	0.9832	**0.9965**
	Time (s)	**0.4025**	201.6623	38.7214	2.4194	23.2054	213.9266	6.9560	15.6662	6.4984
UHAD	AUC	0.9759	0.7509	0.9241	0.9194	0.8953	0.9367	0.9640	0.9677	**0.9902**
	Time (s)	**0.0968**	823.9238	297.4990	69.4288	231.2874	840.1632	33.8099	98.5544	25.7368
HAD-100	AUC	**0.9977**	0.8920	0.9816	0.9920	0.9965	0.9931	0.9924	0.9971	**0.9977**
	Time (s)	**0.1885**	112.6620	11.4568	1.9148	6.0656	84.7245	21.0011	8.6736	13.5849
Pavia	AUC	0.9887	0.9084	0.9910	0.9665	0.9896	0.9873	0.9732	0.9914	**0.9971**
	Time (s)	**0.0538**	140.1739	12.4761	4.7736	24.5325	87.5847	20.8602	27.6034	32.1914
HYDICE	AUC	0.9824	0.8843	0.9742	0.9963	0.9974	0.9216	0.9694	0.9982	**0.9986**
	Time (s)	**0.3047**	129.0832	21.2849	4.2619	22.7051	101.1674	23.4645	14.8354	15.6793
Segundo	AUC	0.9753	0.7570	0.9719	0.9576	0.9305	0.9249	0.8701	0.9714	**0.9900**
	Time (s)	**0.2937**	115.2845	13.9249	2.4278	15.3215	48.6454	18.4637	16.5910	5.5105
SpecTIR	AUC	0.9914	0.8440	0.9491	0.9889	0.9959	0.9774	0.9719	0.9910	**0.9988**
	Time (s)	**1.1375**	410.0467	56.3814	34.3501	175.2243	420.6973	26.2397	43.2219	36.5105

Bold values indicate the best performance results.

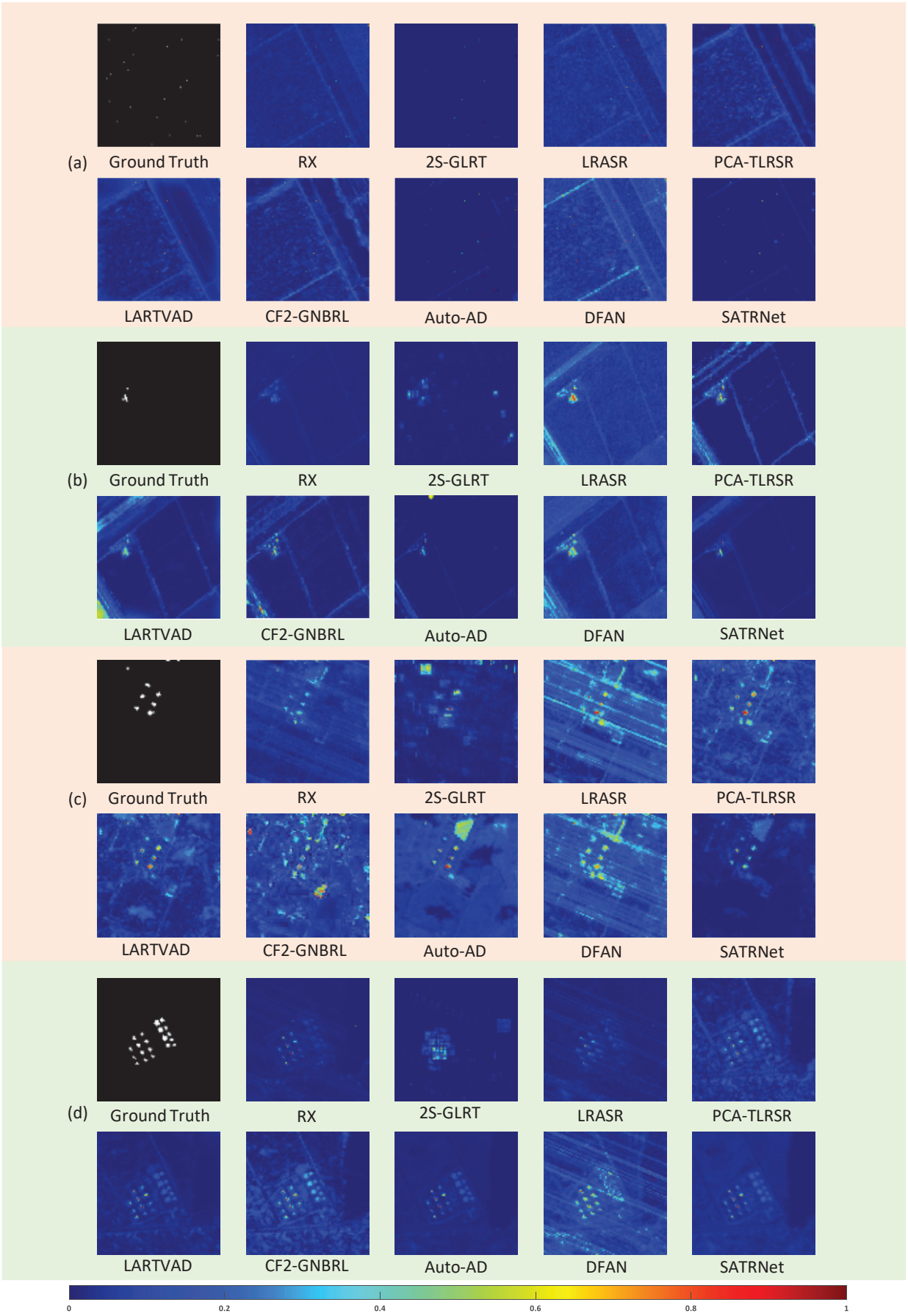


Figure 6. Detection results of different HAD methods. (**a**) Salinas-I. (**b**) Salinas-II. (**c**) Urban-I. (**d**) Urban-II.

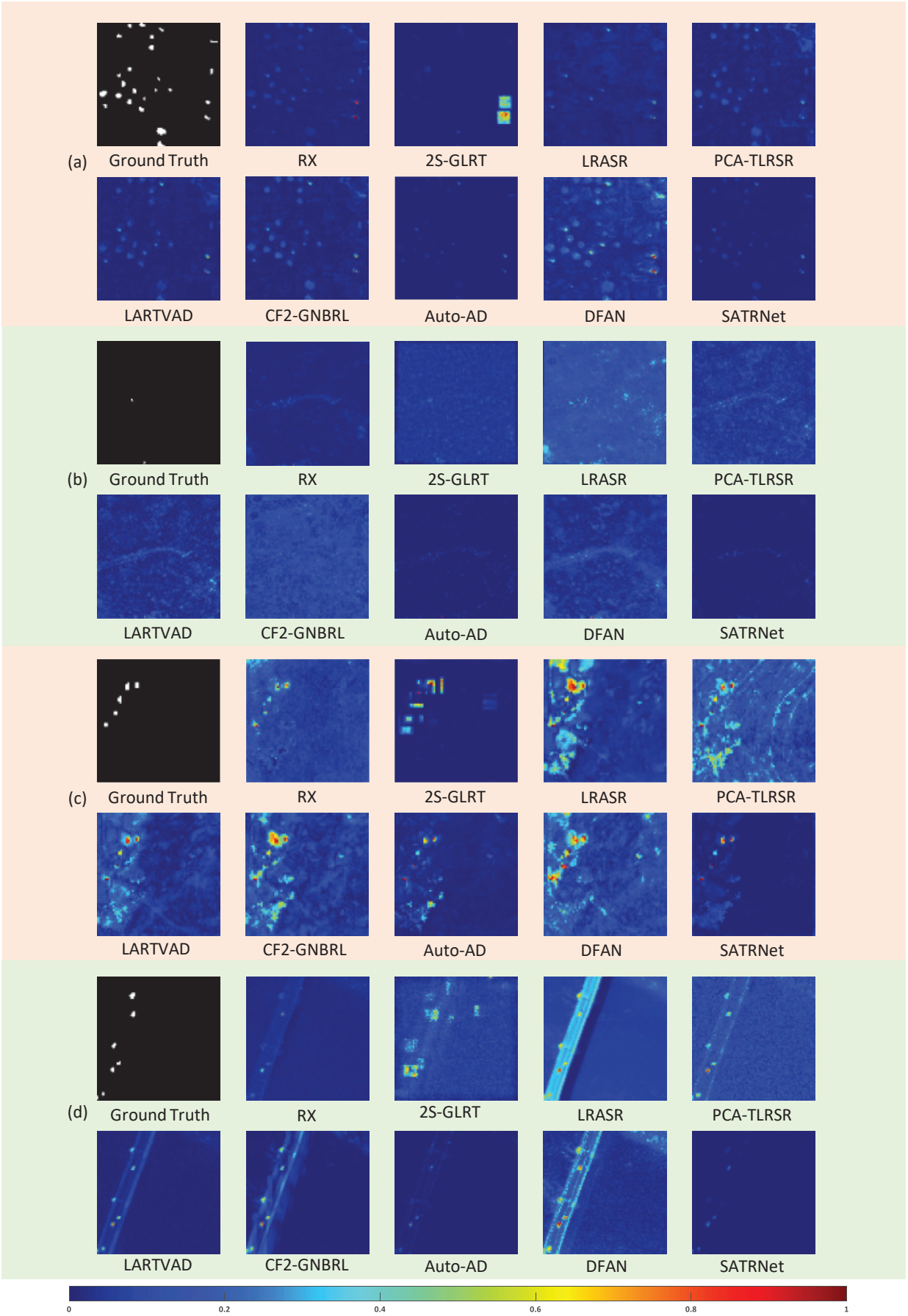


Figure 7. Detection results of different HAD methods. (**a**) Urban-IV. (**b**) UHAD. (**c**) HAD-100. (**d**) Pavia.

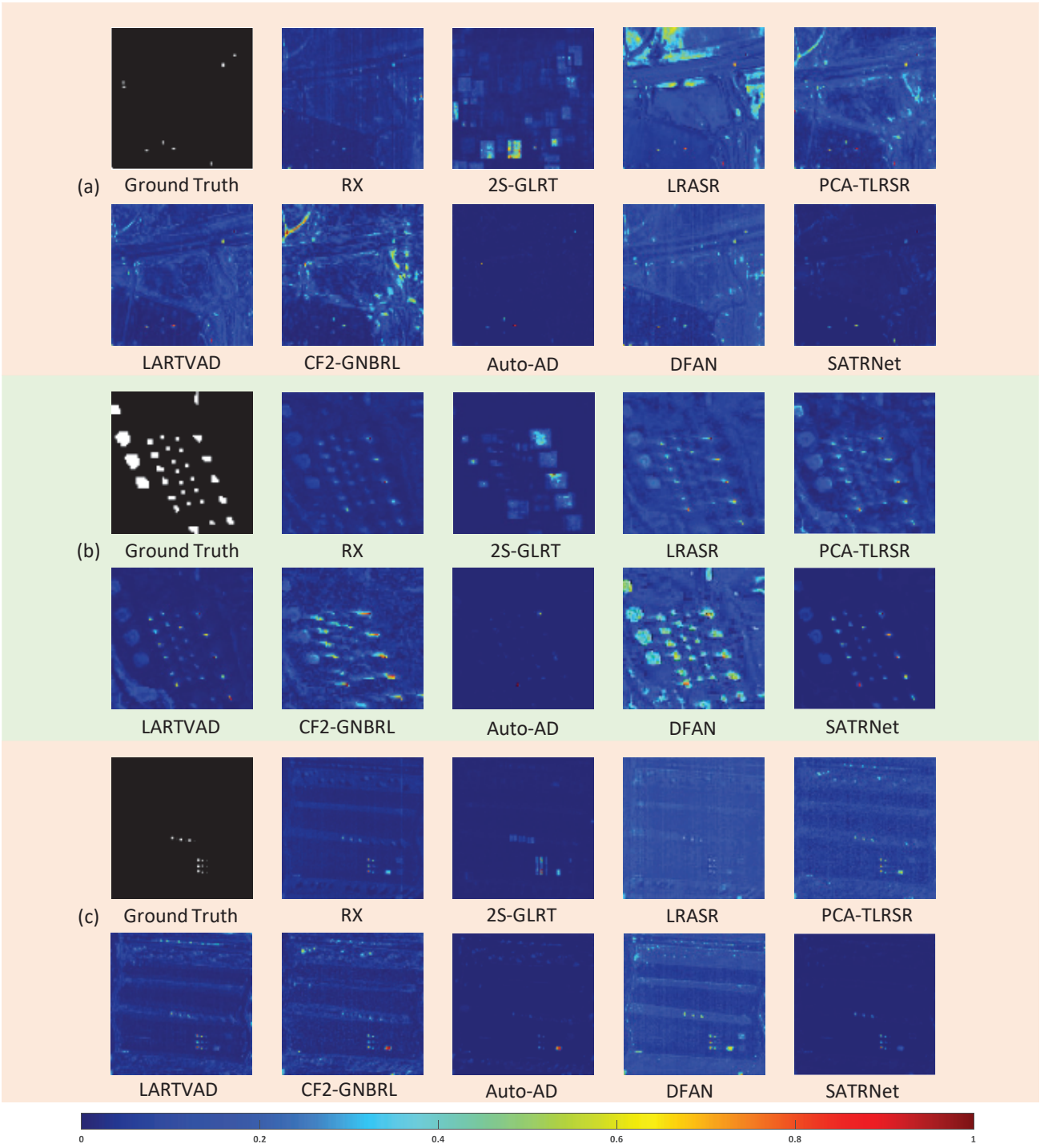


Figure 8. Detection results of different HAD methods. (**a**) HYDICE. (**b**) Segundo. (**c**) SpecTIR.

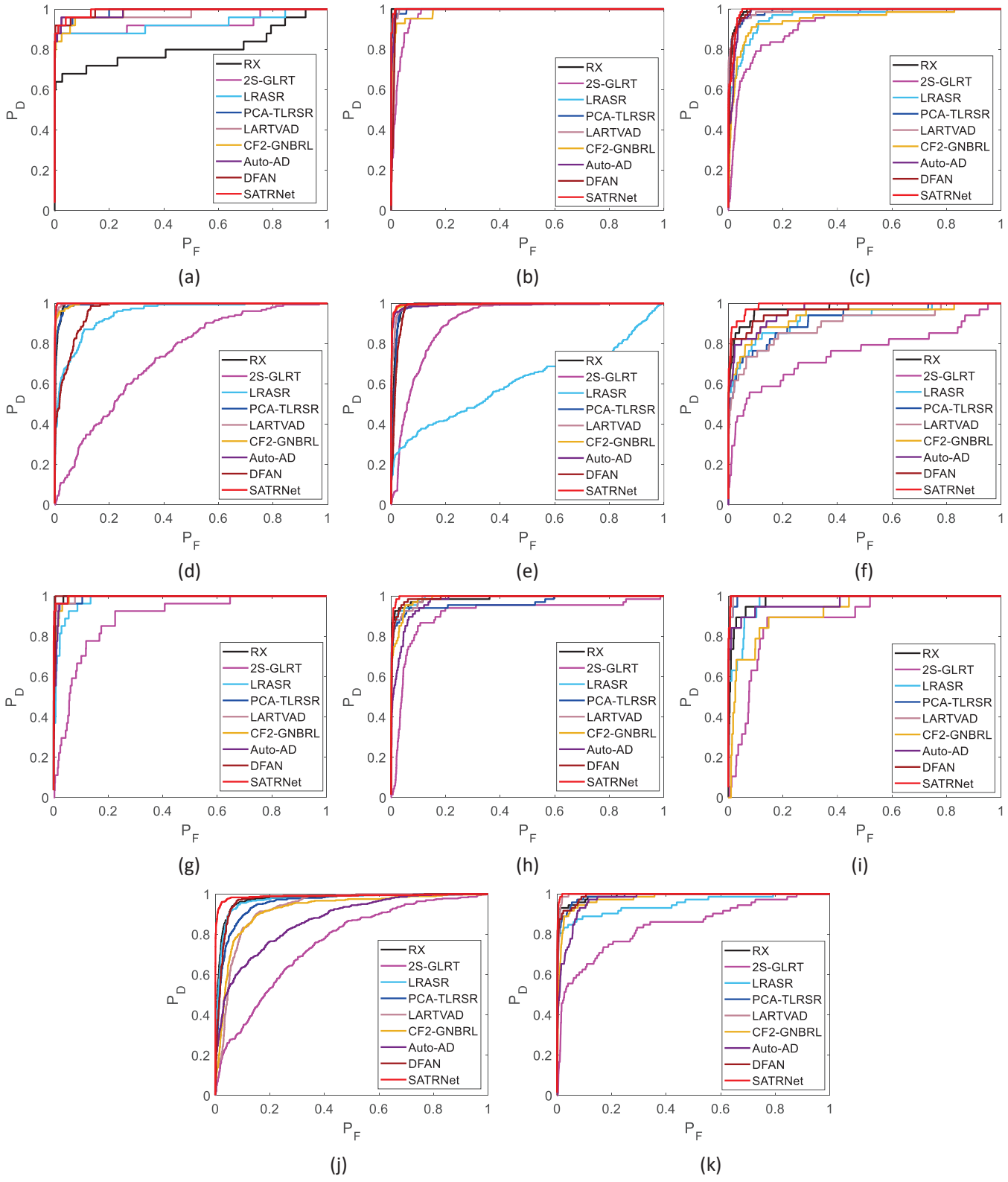


Figure 9. ROC curves of all HAD methods on 11 datasets. (**a**) Salinas-I. (**b**) Salinas-II. (**c**) Urban-I. (**d**) Urban-II. (**e**) Urban-IV. (**f**) UHAD. (**g**) HAD-100. (**h**) Pavia. (**i**) HYDICE. (**j**) Segundo. (**k**) SpecTIR.

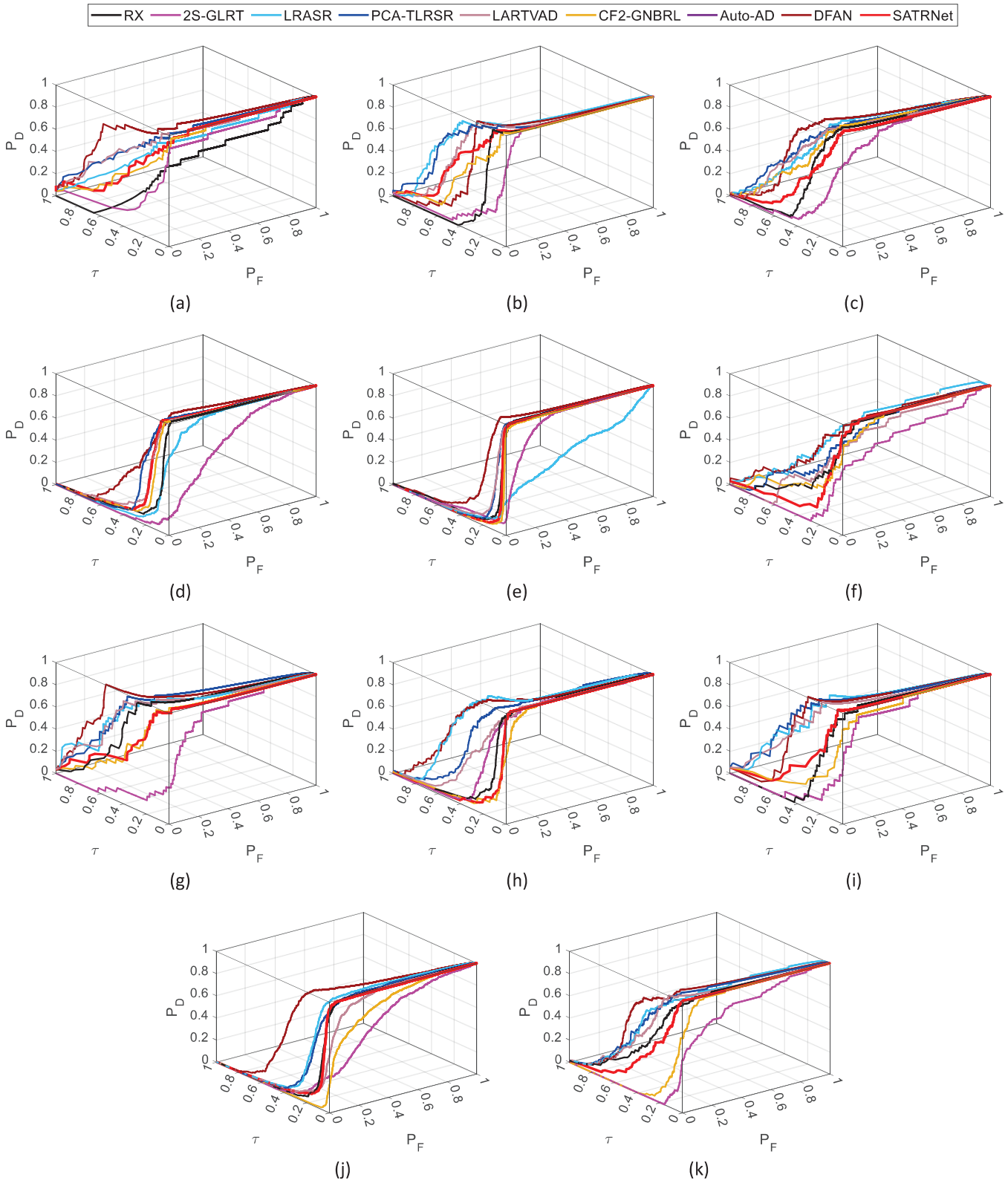


Figure 10. 3D-ROC curves of all HAD methods on 11 datasets. (**a**) Salinas-I. (**b**) Salinas-II. (**c**) Urban-I. (**d**) Urban-II. (**e**) Urban-IV. (**f**) UHAD. (**g**) HAD-100. (**h**) Pavia. (**i**) HYDICE. (**j**) Segundo. (**k**) SpecTIR.

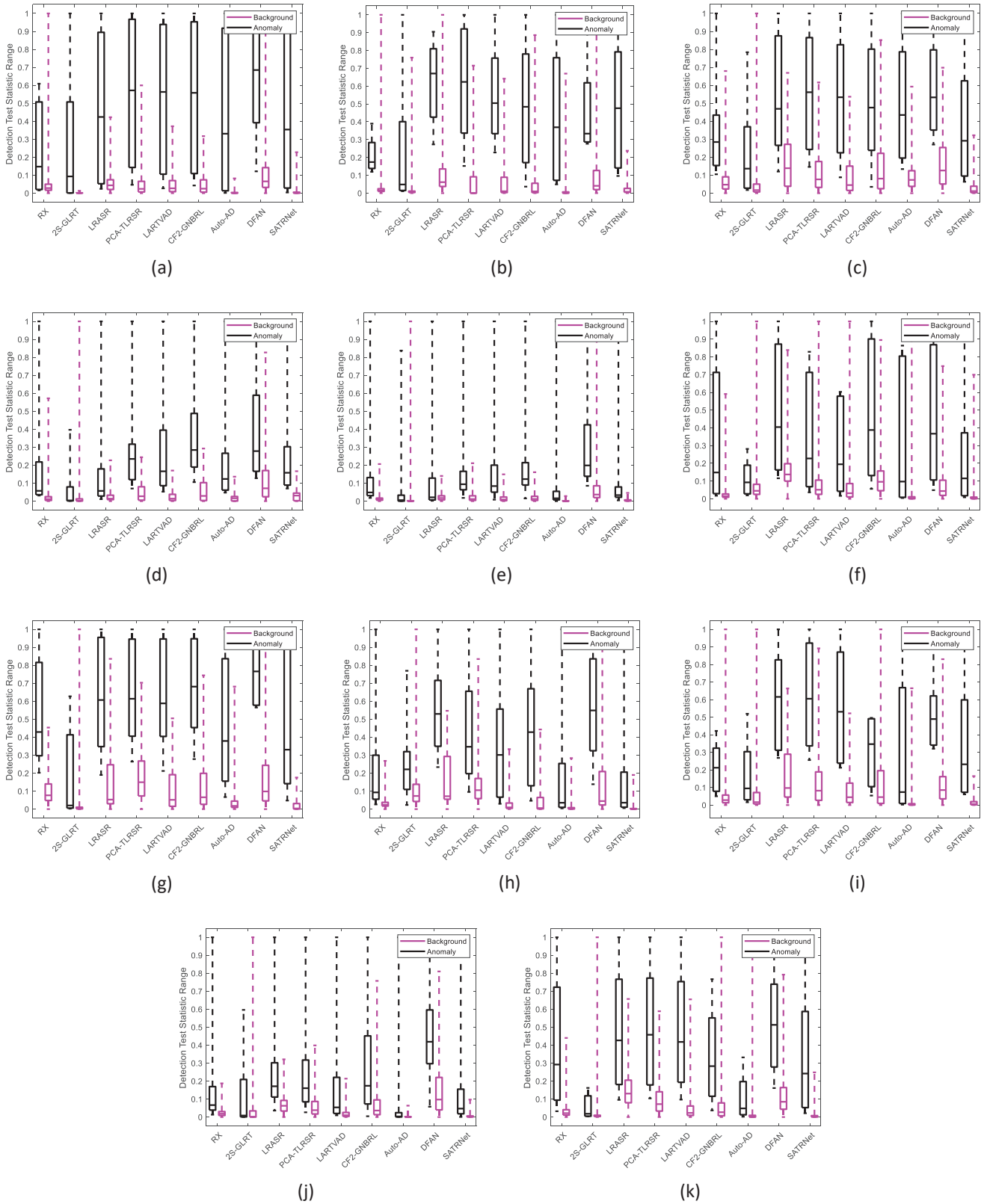


Figure 11. Separability box plots of all HAD methods on 11 datasets. (**a**) Salinas-I. (**b**) Salinas-II. (**c**) Urban-I. (**d**) Urban-II. (**e**) Urban-IV. (**f**) UHAD. (**g**) HAD-100. (**h**) Pavia. (**i**) HYDICE. (**j**) Segundo. (**k**) SpecTIR.

4.4. Ablation Study

This section studies the effectiveness of the two key components of SATRNet, i.e., SALM and WLM. Table 5 shows the detection performance of the proposed method with SALM and WLM removed. As can be seen from Table 5, except for the Salinas-II dataset, the detection performance on other datasets has declined to varying degrees after removing SALM. The performance degradation is more significant on Salinas-I, Urban-IV, UHAD, HAD-100, HYDICE, and SpecTIR after removing SALM, among which HAD completely failed on the UHAD dataset; this sufficiently illustrates the importance of SALM in extracting discriminant features for subsequent separation of background and anomalies. Then, we fixed all values in the weight tensor $\mathcal{K}$ of SATRNet to 1 for the method after removing WLM. As shown in Table 5, there is still a decline in detection performance on all datasets except for Salinas-II. This indicates that DUTR may not be able to completely and accurately learn the background features when it loses the local details injected by WLM, thus affecting HAD performance. We provide qualitative results on the Salinas-I and Pavia datasets, as shown in Figure 12. On the Salinas-I dataset, removing SALM makes the method detect fewer anomalies. When removing WLM, the background is falsely detected, which means that WLM is beneficial for more accurate background learning. For the Pavia dataset, removing either SALM or WLM makes the model wrongly detect roads as anomalies. Such results are not surprising and correspond well to the AUC scores. Therefore, we can conclude that both SALM and WLM are crucial for improving detection performance. When incorporating SALM and WLM, SATRNet can achieve excellent HAD performance on 11 datasets, demonstrating the effectiveness and rationality of the proposed method.

Table 5. Quantitative results of ablation experiments on 11 datasets.

Datasets	Metrics	SATRNet w/o SALM	SATRNet w/o WLM	SATRNet
Salinas-I	AUC	0.7599	0.9441	**0.9924**
	Time (s)	**9.9418**	12.9921	13.3504
Salinas-II	AUC	**0.9998**	**0.9998**	0.9994
	Time (s)	**21.2195**	23.4119	26.3504
Urban-I	AUC	0.9718	0.9832	**0.9909**
	Time (s)	**7.7756**	7.9701	9.6824
Urban-II	AUC	0.9989	0.9993	**0.9994**
	Time (s)	**4.8408**	10.9294	12.5209
Urban-IV	AUC	0.3948	0.9964	**0.9965**
	Time (s)	**4.7376**	5.5326	6.4984
UHAD	AUC	0.0180	0.9257	**0.9902**
	Time (s)	**20.3288**	24.5836	25.7368
HAD-100	AUC	0.7491	0.7457	**0.9977**
	Time (s)	11.5895	**11.1662**	13.5849
Pavia	AUC	0.9903	0.9907	**0.9971**
	Time (s)	29.4589	**29.1680**	32.1914
HYDICE	AUC	0.7549	0.8463	**0.9986**
	Time (s)	**11.6275**	14.7218	15.6793
Segundo	AUC	0.9899	0.9808	**0.9900**
	Time (s)	4.4982	**4.4902**	5.5105
SpecTIR	AUC	0.6042	0.8561	**0.9988**
	Time (s)	**39.8812**	43.1194	46.5105

Bold values indicate the best performance results.

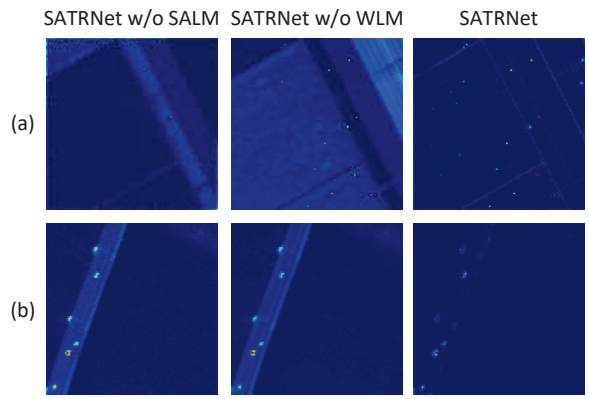


Figure 12. Visualization of ablation study on two datasets. (**a**) Salinas-I. (**b**) Pavia.

4.5. Parameter Analysis

In the proposed SATRNet, the main DUTR part consists of K successive basic units to progressively separate the background and anomalies, corresponding to the iteration times of model-based algorithms. This section investigates the impacts of the number of basic units K on detection performance. Figure 13 shows the detection performance of the proposed method with various values of K on 11 datasets, where the value of K is set as 5, 10, 15, 20, 25, 30, 35, and 40 for each dataset. As shown in Figure 13, the best performance on most of the datasets is produced at $K = 5$ and $K = 10$, which illustrates that the required iteration times of SATRNet are much less than those of model-based algorithms, showing high efficiency. Therefore, for Salinas-I, Salinas-II, Urban-I, Urban-II, Urban-IV, HYDICE, Segundo, and SpecTIR datasets, the value of K is set as 5; for the UHAD and HAD-100 datasets, the value of K is 10; and for the Pavia dataset, the value of K is set as 20. For future users, to balance time efficiency and detection performance, we suggest setting K as 5.

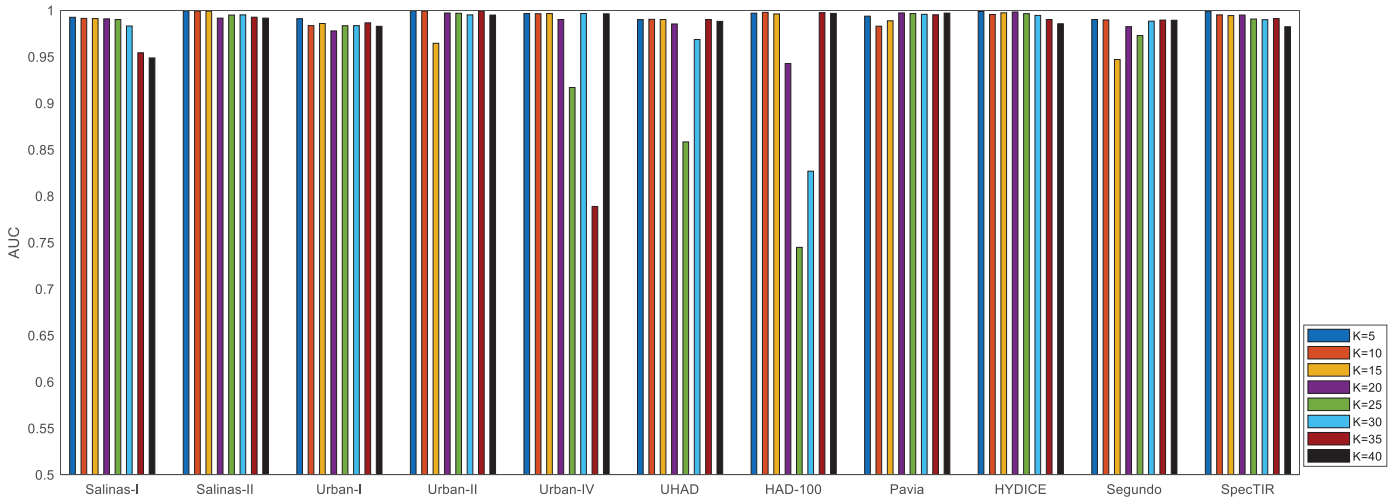


Figure 13. Impact of the parameter K on 11 datasets.

5. Discussion

Despite the great success that deep learning-based models have achieved in different fields, recent studies reveal that deep learning is vulnerable under adversarial attacks [67]. Szegedy et al. [67] first found that deep neural networks are fragile when facing adversarial examples. Adversarial examples, also called adversarial noises, are generated by adding subtle adversarial perturbations to the original images and showing a similar appearance to the original images visually. In remote sensing, adversarial examples have also been proven to exist and draw people's attention to the adversarial robustness of networks [68].

As far as we know, this topic in HAD has not been studied yet. Different from traditional noises that bring about visible degradation to images, adversarial noise can make networks output wrong results without being perceived by humans. This may cause serious security problems in some application scenarios [68]. Therefore, it is essential to analyze the adversarial robustness of the HAD methods.

To examine the adversarial robustness of our proposed SATRNet, we initially generated adversarial samples. These samples necessitate the utilization of adversarial attacks against a well-trained model. Goodfellow et al. [69] introduced a potent adversarial attack method, the fast gradient sign method (FGSM). Notably, FGSM is an established and effective attack method capable of theoretically attacking any loss function [69]. Therefore, we used FGSM as the method for obtaining adversarial samples. The samples x_{adv} are obtained through FGSM as follows:

$$x_{adv} = \text{clip}(x + \epsilon \cdot \text{sign}(\nabla_x \boldsymbol{\mathcal{L}}_{mse}(\theta, x, f(x)))). \tag{26}$$

where $\text{clip}(\cdot)$ clips the pixel values in the image, $\nabla_x \boldsymbol{\mathcal{L}}_{mse}(\theta, x, f(x))$ calculates the gradients of the loss function $\boldsymbol{\mathcal{L}}_{mse}(\cdot)$ with respect to the input sample x, $\text{sign}(\cdot)$ is the sign function, and parameter ϵ is used to control the attack strength. Further details of the method can be found in [69].

We took 6 HSI datasets to test the adversarial robustness of the proposed method. Figure 14 visualizes the adversarial examples on 6 datasets with different ϵ values. As shown in Figure 14, on 6 datasets, the changes in adversarial examples with $\epsilon = 0.01$ and $\epsilon = 0.04$ are hard to perceive compared to clean data. When $\epsilon = 0.1$ and $\epsilon = 0.4$, the quality degradation becomes obvious on five datasets, except for Urban-IV. The small-sized anomalies are completely submerged in the background, especially on the Salinas-I, HYDICE, and SpecTIR datasets. For the Urban-IV dataset, as the value of ϵ increases, the quality degradation is relatively not obvious. Combining the above observations, in order to balance attack strength with visual imperceptibility, we selected $\epsilon = 0.04$ to evaluate adversarial robustness. It is worth mentioning that these observations are basically consistent with those findings in [68,70], where the ϵ value was also set as 0.04.

Quantitative performance of the proposed method under different ϵ values on 6 datasets is exhibited in Table 6. As shown in Table 6, on the Salinas-I, HYDICE, SpecTIR, and HAD-100 datasets, the detection performance of the proposed method experienced a continuous decline with the increase in attack strength. Compared to the detection performance on clean images, the performance degradation on the four datasets was not obvious at $\epsilon = 0.01$ and $\epsilon = 0.04$. Among them, detection performance on the Salinas-I dataset at $\epsilon = 0.01$ even exceeded the performance on clean images, which may be caused by the fact that adversarial attacks highlighted small-sized anomalies to some extent while interfering with background learning. At $\epsilon = 0.1$, the detection performance on the HYDICE, HAD-100, and SpecTIR datasets began to show a large degree of degradation, and at $\epsilon = 0.4$, the detection results became unsatisfactory. Different from the above results, on the two urban datasets Urban-II and Urban-IV, the proposed method showed almost consistently reliable detection performance under different ϵ values, with only a slight performance degradation on the Urban-II dataset at $\epsilon = 0.4$, but this is still reliable. The difference in the results can be attributed to the diversity of the surface features of the datasets. Studies have shown that the adversarial robustness of deep neural networks is also related to remote sensing datasets [71]. In general, datasets with higher diversity tend to provide higher robustness when suffering adversarial attacks. The Urban-II and Urban-IV datasets collect urban scenes, which have higher diversity compared to the large patches of grassland in other datasets such as HAD-100. This can be regarded as one possible reason for the difference in performance. Apart from the AUC index,

further comprehensive quantitative analysis is performed using two metrics: AUC_{BS} [23] and AUC_{ODP} [23]. AUC_{BS} focuses on reflecting the background suppression level of the detector, whereas AUC_{ODP} measures the overall performance of the detection rate and false alarm rate. Tables 7 and 8 show the AUC_{BS} and AUC_{ODP} of the proposed method with different ϵ values, respectively. From Tables 7 and 8, it is not difficult to see that adversarial attacks have a negative impact on the background suppression capability and overall performance of the detector, which is similar to what is reflected by the AUC values. To investigate the adversarial robustness of the proposed method, we further studied the detection performance of $\epsilon = 0.04$. Figure 15 shows the visual detection results of the 6 datasets when $\epsilon = 0.04$. As shown in Figure 15, the visual detection effects of the Urban-II and Urban-IV datasets remain basically unchanged, and on other datasets, compared with the original detection results, a small number of false detections appear in the background area of the detection map. However, under the adversarial attack, the visual effect still maintains strong discrimination ability. The above results indicate that, despite the difficulties brought by adversarial attacks to the HAD problem, the proposed SATRNet method maintains strong detection performance under reasonable attack intensity and presents good visual effects.

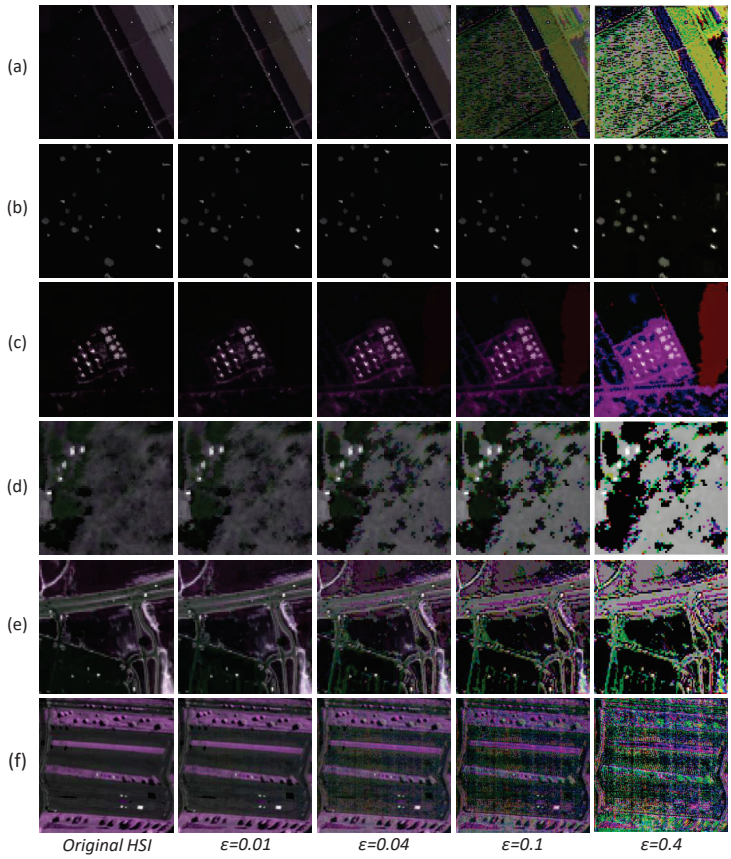


Figure 14. Adversarial examples generated with different ϵ values on 6 datasets. (**a**) Salinas-I. (**b**) Urban-IV. (**c**) Urban-II. (**d**) HAD-100. (**e**) HYDICE. (**f**) SpecTIR.

To more comprehensively evaluate the robustness of the proposed algorithm, we compared SATRNet with S^{3}ANet [70], a recently proposed classification network with strong adversarial robustness. The original S^{3}ANet is designed for hyperspectral image classification, which is a supervised method and is inherently different from the unsupervised HAD problem in this work. In order to fit its structure and learning learning paradigm to the HAD task, we made the following modifications. Firstly, in the original S^{3}ANet, the number of output channels is the number of categories; we adjusted it to

the number of bands of the input image so as to maintain the consistency of input and output dimensions. Secondly, the loss function was replaced by the MSE regression loss to optimize the reconstruction quality. This adjustment enables S³ANet to reconstruct the background of hyperspectral images in an unsupervised manner and, thus, achieve anomaly detection through residual operation. In the training phase, the batch size was 1, the learning rate was 0.001, and the iteration was 100 rounds. Figure 16 shows the performance of S³ANet on the Salinas-I and Urban-IV datasets when $\epsilon = 0.04$. It can be seen from Figure 16 that S³ANet experiences a significant decline in various detection indicators on the Salinas-I and Urban-IV datasets, where the performance degradation on the Urban-IV dataset is more significant. As shown in Figures 17–19, the AUC, AUC_{BS}, and AUC_{ODP} metrics under different ϵ values were tested on the Salinas-I and Urban-IV datasets, respectively. The experimental results show that the proposed method performs better in most cases when compared with S³ANet in terms of different AUC metrics on the Salinas-I dataset. On the Urban-IV dataset, the performance of S³ANet declines seriously as the attack strength increases, while the proposed SATRNet method maintains stability in all metrics, showing its significant superiority to the performance of S³ANet. Detection maps of the two methods at $\epsilon = 0.04$ are given in Figure 20, which also shows that the proposed SATRNet exhibits better detection accuracy. Combining all observations, we can consider that the proposed method has certain adversarial robustness. This benefits from the capture of spatial correlation by self-attention learning and the inherent robustness of the tensor model [72].

Table 6. Detection performance (AUC) of the proposed method with different ϵ values.

Datasets	Test Without FGSM	$\epsilon = 0.01$	$\epsilon = 0.04$	$\epsilon = 0.1$	$\epsilon = 0.4$
Salinas-I	0.9924	**0.9933**	0.9912	0.9861	0.8761
Urban-IV	**0.9965**	**0.9965**	**0.9965**	**0.9965**	**0.9965**
Urban-II	**0.9994**	**0.9994**	**0.9994**	**0.9994**	0.9987
HAD-100	**0.9977**	0.9970	0.9907	0.9467	0.7150
HYDICE	**0.9986**	0.9980	0.9892	0.9480	0.7257
SpecTIR	**0.9988**	0.9981	0.9918	0.9516	0.7139

Bold values indicate the best performance results.

Table 7. Background suppression level (AUC_{BS}) of the proposed method with different ϵ values.

Datasets	Test Without FGSM	$\epsilon = 0.01$	$\epsilon = 0.04$	$\epsilon = 0.1$	$\epsilon = 0.4$
Salinas-I	**0.9894**	**0.9894**	0.9834	0.9751	0.8761
Urban-IV	**0.9927**	**0.9927**	**0.9927**	**0.9927**	**0.9927**
Urban-II	0.9701	0.9700	0.9696	0.9666	**0.9987**
HAD-100	**0.9864**	0.9842	0.9702	0.9021	0.6965
HYDICE	**0.9892**	0.9864	0.9678	0.9315	0.7253
SpecTIR	**0.9950**	0.9923	0.9820	0.9516	0.7138

Bold values indicate the best performance results.

Table 8. Overall performance (AUC_{ODP}) of the proposed method with different ϵ values.

Datasets	Test Without FGSM	$\epsilon = 0.01$	$\epsilon = 0.04$	$\epsilon = 0.1$	$\epsilon = 0.4$
Salinas-I	**1.4078**	1.4063	1.4071	1.2712	0.8761
Urban-IV	**1.4085**	**1.4085**	**1.4085**	**1.4085**	**1.4085**
Urban-II	1.1681	1.1683	1.1698	**1.1727**	1.0084
HAD-100	**1.4414**	1.4300	1.3905	1.2861	0.7210
HYDICE	**1.3212**	1.3157	1.2814	1.0531	0.7256
SpecTIR	**1.2826**	1.2788	1.1708	0.9517	0.7139

Bold values indicate the best performance results.

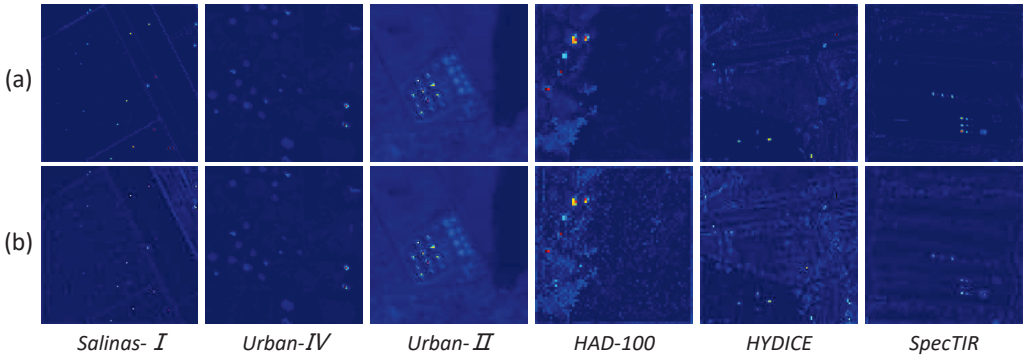


Figure 15. Visualization of detection performance for $\epsilon = 0.04$ versus detection performance without adversarial attacks. (**a**) Detection performance without adversarial attacks. (**b**) Detection performance with $\epsilon = 0.04$.

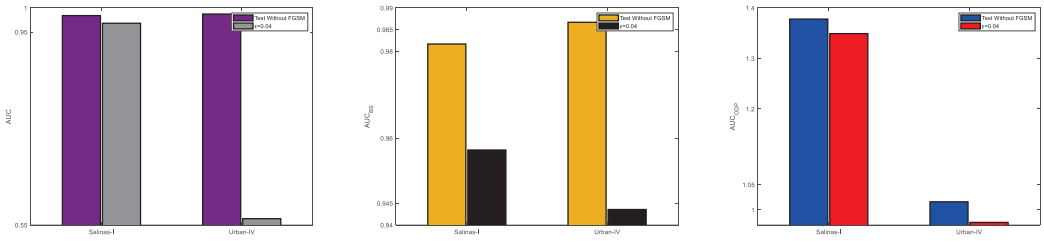


Figure 16. Performance of S³ANet with and without adversarial attacks ($\epsilon = 0.04$) on Salinas-I and Urban-IV datasets using different metrics. (**a**) AUC. (**b**) AUC_{BS}. (**c**) AUC_{ODP}.

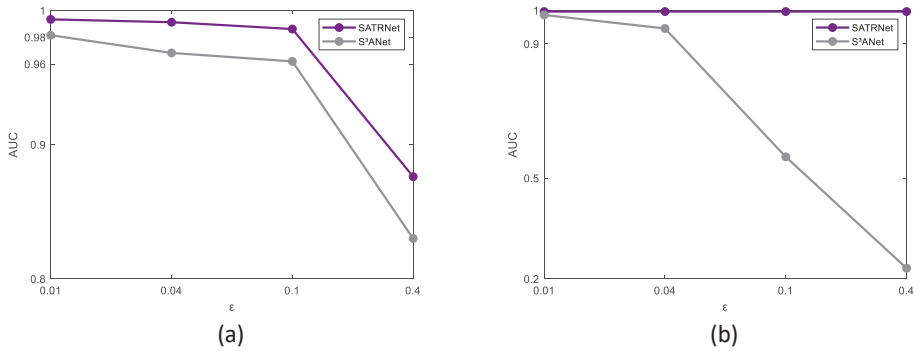


Figure 17. AUC values of the Salinas-I and Urban-IV datasets for different ϵ values. (**a**) Salinas-I. (**b**) Urban-IV.

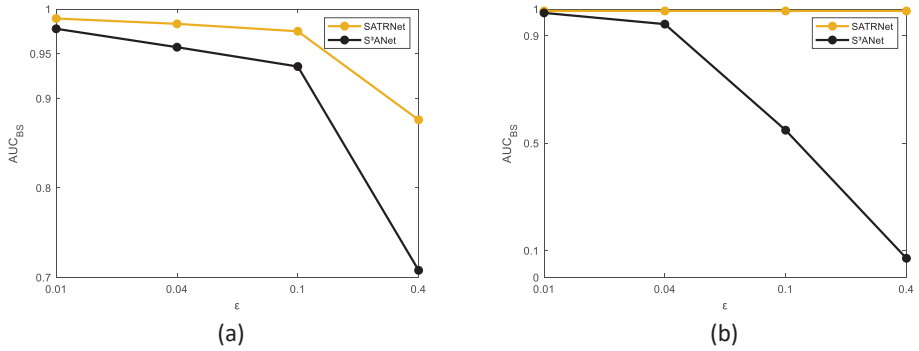


Figure 18. AUC_{BS} values of the Salinas-I and Urban-IV datasets for different ϵ values. (**a**) Salinas-I. (**b**) Urban-IV.

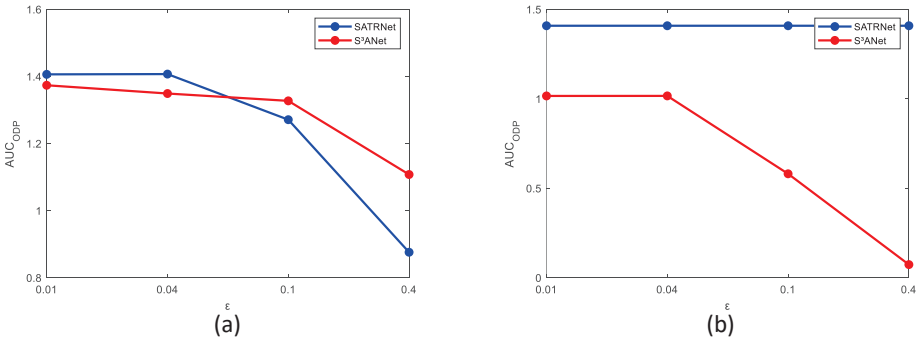


Figure 19. AUC_{ODP} values of the Salinas-I and Urban-IV datasets for different ϵ values. (**a**) Salinas-I. (**b**) Urban-IV.

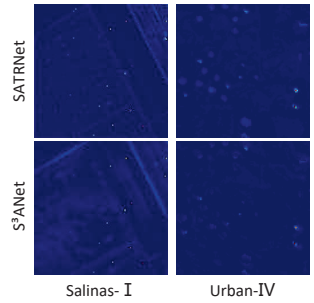


Figure 20. Detection maps of the two methods at $\epsilon = 0.04$.

6. Conclusions

This paper presents a self-attention-aided deep unfolding tensor representation network (SATRNet) for HAD. SATRNet introduces SALM to extract discriminative features to reduce the difficulty of the following separation of background and anomalies. Then, we formulated the HAD problem as the TR problem, which is further guided by the weight tensor $\mathcal{K}$ generated by WLM to inject local details into the model. By unfolding and parameterizing this TR algorithm, DUTR has K consecutive basic units with precise correspondence to the iterative algorithm, ensuring seamless integration of model-based methods and learning-based methods. Combining the merits of model-based and learning-based paradigms, SATRNet can effectively alleviate the IM problem, leading to the best performance on 11 datasets. Note that SATRNet is resistant to adversarial attacks, which also represents the first work to investigate adversarial robustness in HAD.

Author Contributions: Conceptualization, J.Y. and J.Z.; methodology, J.Y.; software, J.Z.; validation, J.Y., L.C. and Y.L.; formal analysis, L.C.; writing—original draft preparation, J.Z.; writing—review and editing, L.C. and Y.L.; visualization, J.Z. and H.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was jointly funded by the National Natural Science Foundation of China (No. 62406181, 62373233), the Research Project Supported by Shanxi Scholarship Council of China (No. 2023-023), the Fundamental Research Program of Shanxi Province (No. 202203021222010), the Science and Technology Major Project of Shanxi Province (No. 202201020101006), and the special fund for Science and Technology Innovation Teams of Shanxi Province.

Data Availability Statement: The data used in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, X.; Wang, Z.; Wang, K.; Jia, H.; Han, Z.; Tang, Y. Multi-dimensional low-rank with weighted Schatten *p*-norm minimization for hyperspectral anomaly detection. *Remote Sens.* **2023**, *16*, 74. [CrossRef]
2. Zhao, R.; Yang, Z.; Meng, X.; Shao, F. A novel fully convolutional auto-encoder based on dual clustering and latent feature adversarial consistency for hyperspectral anomaly detection. *Remote Sens.* **2024**, *16*, 717. [CrossRef]
3. Yin, C.; Lv, X.; Zhang, L.; Ma, L.; Wang, H.; Zhang, L.; Zhang, Z. Hyperspectral UAV images at different altitudes for monitoring the leaf nitrogen content in cotton crops. *Remote Sens.* **2022**, *14*, 2576. [CrossRef]
4. Farrar, M.B.; Wallace, H.M.; Brooks, P.; Yule, C.M.; Tahmasbian, I.; Dunn, P.K.; Hosseini Bai, S. A performance evaluation of Vis/NIR hyperspectral imaging to predict curcumin concentration in fresh turmeric rhizomes. *Remote Sens.* **2021**, *13*, 1807. [CrossRef]
5. Zhang, X.; Han, L.; Dong, Y.; Shi, Y.; Huang, W.; Han, L.; González-Moreno, P.; Ma, H.; Ye, H.; Sobeih, T. A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral UAV images. *Remote Sens.* **2019**, *11*, 1554. [CrossRef]
6. Reed, I.S.; Yu, X. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Trans. Acoust. Speech Signal Process.* **1990**, *38*, 1760–1770. [CrossRef]
7. Borghys, D.; Kåsen, I.; Achard, V.; Perneel, C. Comparative evaluation of hyperspectral anomaly detectors in different types of background. In Proceedings of the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, Baltimore, MD, USA, 23–27 April 2012; pp. 803–814.
8. Guo, Q.; Zhang, B.; Ran, Q.; Gao, L.; Li, J.; Plaza, A. Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2351–2366. [CrossRef]
9. Kwon, H.; Nasrabadi, N.M. Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 388–397. [CrossRef]
10. Chang, C.-I. Hyperspectral anomaly detection: A dual theory of hyperspectral target detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–20. [CrossRef]
11. Li, J.; Zhang, H.; Zhang, L.; Ma, L. Hyperspectral anomaly detection by the use of background joint sparse representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2523–2533. [CrossRef]
12. Zhao, G.; Li, F.; Zhang, X.; Laakso, K.; Chan, J.C.-W. Archetypal analysis and structured sparse representation for hyperspectral anomaly detection. *Remote Sens.* **2021**, *13*, 4102. [CrossRef]
13. Xu, Y.; Wu, Z.; Li, J.; Plaza, A.; Wei, Z. Anomaly detection in hyperspectral images based on low-rank and sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 1990–2000. [CrossRef]
14. Cheng, T.; Wang, B. Graph and total variation regularized low-rank representation for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 391–406. [CrossRef]
15. Ren, L.; Gao, L.; Wang, M.; Sun, X.; Chanussot, J. HADGSM: A unified nonconvex framework for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5503415. [CrossRef]
16. Shang, W.; Jouni, M.; Wu, Z.; Xu, Y.; Dalla Mura, M.; Wei, Z. Hyperspectral anomaly detection based on regularized background abundance tensor decomposition. *Remote Sens.* **2023**, *15*, 1679. [CrossRef]
17. Shang, W.; Peng, J.; Wu, Z.; Xu, Y.; Jouni, M.; Dalla Mura, M.; Wei, Z. Hyperspectral anomaly detection via sparsity of core tensor under gradient domain. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5517816. [CrossRef]

18. Feng, M.; Chen, W.; Yang, Y.; Shu, Q.; Li, H.; Huang, Y. Hyperspectral anomaly detection based on tensor ring decomposition with factors TV regularization. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5514114. [CrossRef]
19. Sun, S.; Liu, J.; Chen, X.; Li, W.; Li, H. Hyperspectral anomaly detection with tensor average rank and piecewise smoothness constraints. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 8679–8692. [CrossRef]
20. Wang, M.; Wang, Q.; Hong, D.; Roy, S.K.; Chanussot, J. Learning tensor low-rank representation for hyperspectral anomaly detection. *IEEE Trans. Cybern.* **2022**, *53*, 679–691. [CrossRef] [PubMed]
21. Sun, S.; Liu, J.; Li, W. Spatial invariant tensor self-representation model for hyperspectral anomaly detection. *IEEE Trans. Cybern.* **2023**, *54*, 3120–3131. [CrossRef] [PubMed]
22. Qin, H.; Shen, Q.; Zeng, H.; Chen, Y.; Lu, G. Generalized nonconvex low-rank tensor representation for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5526612. [CrossRef]
23. Yang, J.; Zhao, J.; Chen, L.; Geng, H.; Zhang, P. Learning nonconvex tensor representation with generalized reweighted sparse regularization for hyperspectral anomaly detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 14718–14737. [CrossRef]
24. Yu, Q.; Bai, M. Generalized nonconvex hyperspectral anomaly detection via background representation learning with dictionary constraint. *SIAM J. Imaging Sci.* **2024**, *17*, 917–950. [CrossRef]
25. Chen, L.; Li, Z.; Lu, Z.; Wang, Y.; Nie, H.; Yang, C. Domain-invariant feature learning via margin and structure priors for robotic grasping. *IEEE Robot. Autom. Lett.* **2024**, *10*, 1313–1320. [CrossRef]
26. Li, J.; Zheng, K.; Gao, L.; Han, Z.; Li, Z.; Chanussot, J. Enhanced deep image prior for unsupervised hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5504218. [CrossRef]
27. Li, J.; Zheng, K.; Gao, L.; Ni, L.; Huang, M.; Chanussot, J. Model-informed multistage unsupervised network for hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5516117. [CrossRef]
28. Li, Q.; Gong, M.; Yuan, Y.; Wang, Q. Symmetrical feature propagation network for hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5536912. [CrossRef]
29. Wang, S.; Wang, X.; Zhang, L.; Zhong, Y. Auto-AD: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5503314. [CrossRef]
30. Fan, G.; Ma, Y.; Mei, X.; Fan, F.; Huang, J.; Ma, J. Hyperspectral anomaly detection with robust graph autoencoders. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5511314.
31. Wu, Z.; Wang, B. Transformer-based autoencoder framework for nonlinear hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5508015. [CrossRef]
32. Cheng, X.; Huo, Y.; Lin, S.; Dong, Y.; Zhao, S.; Zhang, M.; Wang, H. Deep feature aggregation network for hyperspectral anomaly detection. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5033016. [CrossRef]
33. Lian, J.; Wang, L.; Sun, H.; Huang, H. GT-HAD: Gated transformer for hyperspectral anomaly detection. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *36*, 3631–3645. [CrossRef]
34. Xiao, S.; Zhang, T.; Xu, Z.; Qu, J.; Hou, S.; Dong, W. Anomaly detection of hyperspectral images based on transformer with spatial–spectral dual-window mask. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 1414–1426. [CrossRef]
35. Wang, D.; Zhuang, L.; Gao, L.; Sun, X.; Zhao, X.; Plaza, A. Sliding dual-window-inspired reconstruction network for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5504115. [CrossRef]
36. Yang, Z.; Zhao, R.; Meng, X.; Yang, G.; Sun, W.; Zhang, S.; Li, J. A multi-scale mask convolution-based blind-spot network for hyperspectral anomaly detection. *Remote Sens.* **2024**, *16*, 3036. [CrossRef]
37. Wang, D.; Zhuang, L.; Gao, L.; Sun, X.; Huang, M.; Plaza, A.J. PDBSNet: Pixel-shuffle downsampling blind-spot reconstruction network for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5511914. [CrossRef]
38. Wang, D.; Zhuang, L.; Gao, L.; Sun, X.; Huang, M.; Plaza, A. BockNet: Blind-block reconstruction network with a guard window for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5531916. [CrossRef]
39. Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; Yan, S. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 925–938. [CrossRef]
40. Chan, S.H.; Wang, X.; Elgendy, O.A. Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. *IEEE Trans. Comput. Imaging* **2016**, *3*, 84–98. [CrossRef]
41. He, X.; Wu, J.; Ling, Q.; Li, Z.; Lin, Z.; Zhou, S. Anomaly detection for hyperspectral imagery via tensor low-rank approximation with multiple subspace learning. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5509917. [CrossRef]
42. Khader, A.; Yang, J.; Xiao, L. NMF-DuNet: Nonnegative matrix factorization inspired deep unrolling networks for hyperspectral and multispectral image fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5704–5720. [CrossRef]
43. Qu, J.; Dong, W.; Li, Y.; Hou, S.; Du, Q. An interpretable unsupervised unrolling network for hyperspectral pansharpening. *IEEE Trans. Cybern.* **2023**, *53*, 7943–7956. [CrossRef]
44. Wang, K.; Liao, X.; Li, J.; Meng, D.; Wang, Y. Hyperspectral image super-resolution via knowledge-driven deep unrolling and transformer embedded convolutional recurrent neural network. *IEEE Trans. Image Process.* **2023**, *32*, 4581–4594. [CrossRef]

45. Yang, Y.; Sun, J.; Li, H.; Xu, Z. ADMM-CSNet: A deep learning approach for image compressive sensing. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *42*, 521–538. [CrossRef] [PubMed]
46. Li, C.; Zhang, B.; Hong, D.; Yao, J.; Chanussot, J. LRR-Net: An interpretable deep unfolding network for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5513412. [CrossRef]
47. Lin, S.; Cheng, X.; Zeng, Y.; Huo, Y.; Zhang, M.; Wang, H. Low-rank and sparse representation inspired interpretable network for hyperspectral anomaly detection. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5033116. [CrossRef]
48. Wang, L.; Wang, X.; Vizziello, A.; Gamba, P. RSAAE: Residual self-attention-based autoencoder for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5510614. [CrossRef]
49. Tu, B.; Zhou, T.; Liu, B.; He, Y.; Li, J.; Plaza, A. Multi-scale autoencoder suppression strategy for hyperspectral image anomaly detection. *IEEE Trans. Image Process.* 2023, *online ahead of print*. [CrossRef]
50. Li, Q.; Yuan, Y.; Jia, X.; Wang, Q. Dual-stage approach toward hyperspectral image super-resolution. *IEEE Trans. Image Process.* **2022**, *31*, 7252–7263. [CrossRef]
51. Li, Q.; Yuan, Y.; Wang, Q. Multiscale factor joint learning for hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5523110. [CrossRef]
52. Li, J.; Zheng, K.; Li, Z.; Gao, L.; Jia, X. X-shaped interactive autoencoders with cross-modality mutual learning for unsupervised hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5518317. [CrossRef]
53. Chen, L.; Niu, M.; Yang, J.; Qian, Y.; Li, Z.; Wang, K.; Yan, T.; Huang, P. Robotic grasp detection using structure prior attention and multiscale features. *IEEE Trans. Syst. Man Cybern. Syst.* **2024**, *54*, 7039–7053. [CrossRef]
54. Li, Y.; Zhang, K.; Cao, J.; Timofte, R.; Van Gool, L. Localvit: Bringing locality to vision transformers. *arXiv* **2021**, arXiv:2104.05707.
55. Mercer, J. Xvi. functions of positive and negative type, and their connection the theory of integral equations. *Philos. Trans. R. Soc. Lond. Ser. A Contain. Pap. A Math. Phys. Character* **1909**, *209*, 415–446.
56. Zhang, M.; Zhang, C.; Zhang, Q.; Guo, J.; Gao, X.; Zhang, J. ESSAformer: Efficient transformer for hyperspectral image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–3 October 2023; pp. 23073–23084.
57. Hendrycks, D.; Gimpel, K. Gaussian error linear units (gelus). *arXiv* **2016**, arXiv:1606.08415.
58. Tan, K.; Hou, Z.; Wu, F.; Du, Q.; Chen, Y. Anomaly detection for hyperspectral imagery based on the regularized subspace method and collaborative representation. *Remote Sens.* **2019**, *11*, 1318. [CrossRef]
59. Zhang, W.; Lu, X.; Li, X. Similarity constrained convex nonnegative matrix factorization for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4810–4822. [CrossRef]
60. Liu, S.; Peng, L.; Chang, X.; Wen, G.; Zhu, C. Adaptive dual-domain learning for hyperspectral anomaly detection with state space models. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5503719. [CrossRef]
61. Li, Z.; Wang, Y.; Xiao, C.; Ling, Q.; Lin, Z.; An, W. You only train once: Learning a general anomaly enhancement network with random masks for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–18. [CrossRef]
62. Sun, B.; Zhao, Z.; Liu, D.; Gao, X.; Yu, T. Tensor decomposition-inspired convolutional autoencoders for hyperspectral anomaly detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4990–5000. [CrossRef]
63. Li, L.; Li, W.; Qu, Y.; Zhao, C.; Tao, R.; Du, Q. Prior-based tensor approximation for anomaly detection in hyperspectral imagery. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *33*, 1037–1050. [CrossRef]
64. Liu, J.; Hou, Z.; Li, W.; Tao, R.; Orlando, D.; Li, H. Multipixel anomaly detection with unknown patterns for hyperspectral imagery. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 5557–5567. [CrossRef]
65. Chang, C.-I. Comprehensive analysis of receiver operating characteristic (roc) curves for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5541124. [CrossRef]
66. Chang, C.-I. An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5131–5153. [CrossRef]
67. Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; Fergus, R. Intriguing properties of neural networks. *arXiv* **2013**, arXiv:1312.6199.
68. Xu, Y.; Du, B.; Zhang, L. Self-attention context network: Addressing the threat of adversarial attacks for hyperspectral image classification. *IEEE Trans. Image Process.* **2021**, *30*, 8671–8685. [CrossRef] [PubMed]
69. Goodfellow, I.J.; Shlens, J.; Szegedy, C. Explaining and harnessing adversarial examples. *arXiv* **2014**, arXiv:1412.6572.
70. Xu, Y.; Xu, Y.; Jiao, H.; Gao, Z.; Zhang, L. S^{3}ANet: Spatial–spectral self-attention learning network for defending against adversarial attacks in hyperspectral image classification. *IEEE Trans. Image Process.* **2024**, *62*, 5512913.

71. Chen, L.; Zhu, G.; Li, Q.; Li, H. Adversarial example in remote sensing image recognition. *arXiv* **2019**, arXiv:1910.13222.
72. Zhang, J.; Hong, Y.; Cheng, D.; Zhang, L.; Zhao, Q. Defending adversarial attacks in Graph Neural Networks via tensor enhancement. *Pattern Recognit.* **2025**, *158*, 110954. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

remote sensing

MDPI

Article

Reliable Task-Constrained Band Selection Method for Hyperspectral Anomaly Detection

Genrui Zhang, Wenzheng Wang *, Yuqi Han, Chenwei Deng and Xingshi Luo

School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China; genrui_zhang@bit.edu.cn (G.Z.); yuqi_han@tsinghua.edu.cn (Y.H.); cwdeng@bit.edu.cn (C.D.); xingshi_luo@bit.edu.cn (X.L.)

* Correspondence: wang_wenzheng@bit.edu.cn or wenzheng_wang@pku.edu.cn

Abstract: Hyperspectral band selection utilizes a crucial band subset to represent original data. In hyperspectral anomaly detection tailored for specific tasks, detection performance can be enhanced by pre-selecting a subset of bands that are more representative. However, existing methods remain constrained in modeling spatial–spectral dependencies and simultaneously extracting distinct bands' contribution from the established model, thus struggling to balance effectiveness and stability. To address these issues, we propose a reliable band selection method for anomaly detection. Concretely, we conduct a convolution–transformer hybrid autoencoder architecture to fully exploit the local and global spatial–spectral interdependencies. Next, we design an anomaly–background separability constraint to seamlessly integrate the task priors of anomaly detection into network optimization. Furthermore, we design a spectral attention module to quantify the contribution of different bands during network optimization. Simultaneously, an adaptive band allocation method is designed to optimize the internal structure of the selected band subset. Extensive experiments demonstrate that the proposed method achieves more robust band selection results compared to existing related methods.

Keywords: band selection; anomaly detection; autoencoder; spectral reconstruction

1. Introduction

Hyperspectral images (HSIs) consist of hundreds of narrow spectral channels, providing both two-dimensional spatial information and detailed spectral information of objects [1,2]. The fine spectral information reflects the intrinsic characteristics of the targets of interest. Hyperspectral anomaly detection is one of the most significant research areas in HSI processing. It aims to identify anomalous pixels whose spectral features significantly differ from those of neighboring background pixels in an unsupervised manner. This has critical implications in areas such as smart agriculture, mineral exploration, and military reconnaissance [3–5].

In the real world, people often have specific expectations regarding anomalies, such as ships in the sea, vehicles in the forest, and airplanes in the sky [6]. In such practical cases, anomaly detection is often task specific, which facilitates obtaining coarse prior information about background and anomalies [7,8]. To leverage this information for the enhancement of anomaly detection performance tailored for specific tasks, this paper focuses on developing a band pre-selection model to provide a reliable band subset.

Concretely, band selection is an effective and widely applied approach, used to select a subset of necessary bands from the original spectrum, effectively retaining the spectral significance of spectral channels [9–11]. This is because HSI exhibits high inter-band

correlation, leading to substantial data redundancy. It will obscure spectral features, thereby severely limiting the performance of anomaly detection [12,13]. Many studies have indicated that anomaly detection using the full spectrum does not always yield optimal results. High-performance anomaly detection can be achieved using a smaller subset of selected discriminatory bands, enhancing spectral separation between the anomalies and background, and achieving background suppression [14–16].

Many commonly used band selection methods focus on redundancy reduction, without considering how to relate them to the anomaly detection task itself [17–19]. They utilize data statistics such as variance, entropy, and information divergence for evaluation, which can be categorized into rank-based, clustering-based, and search-based methods. Rank-based methods assess the priority of each band according to predefined criteria and select the top-ranked bands, such as maximum-variance principal component analysis (MVPCA) and maximum-SNR principal component analysis (MSNRPCA) [14], as well as methods constrained by energy minimization for band selection [20]. Clustering-based methods divide the original bands into clusters and select representative bands from each cluster to form a band subset, such as [21], which uses different statistical features of hyperspectral data like interquartile range and covariance for k-means-based band clustering. Ref. [22] uses affinity propagation (AP) clustering based on samples, considering the correlation between bands. Search-based methods transform the band selection problem into an optimization problem for an objective function, searching for the optimal solution under the objective function, such as the recently proposed MCBS method that treats band selection as a multi-criteria decision-making (MCDM) problem, utilizing the cuckoo search algorithm for band selection [11]. Optimal Neighborhood Reconstruction (ONR) [23] proposes a band selection method that models the neighborhood correlation of hyperspectral images (HSIs) and suppresses noise. However, these methods generally lack adaptability to anomaly detection, and their band selection results have limited effectiveness in enhancing subsequent detection performance.

Consequently, prior constraints related to anomaly–background anomaly detection have been incorporated into band selection algorithms to improve detection accuracy. This involves selecting band subsets with high separation between the anomalies and background to highlight anomalies while suppressing the background. By using entropy and histograms for coarse spectral band classification (EHBS), a band subset is screened to suppress the background and highlight anomalies in [24]. Hyperspectral anomaly detection-oriented unsupervised band selection (HADUBS) [25] integrates anomaly distribution characteristics with band selection for the first time by separating anomaly subspaces through spectral–spatial out-of-distribution (OOD) detection. Ref. [26] introduces the multiband target–background ratio (MBTBR) prior to identify discriminative band subsets that maximize the separability between anomalies and backgrounds. Ref. [27] employs density peak clustering to partition anomalies and background sets, combined with residual projection for band selection (RDBSMaxSR), effectively preserving the highly discriminative features between anomalies and backgrounds. However, these methods rely on linear assumptions, insufficiently model the nonlinear spectral–spatial relationships, and fail to explicitly account for inter-band redundancy.

Recently, deep learning models have demonstrated strong feature extraction capabilities on hyperspectral data, prompting researchers to explore deep learning models for solving the band selection problem [28]. However, most existing band selection methods have been developed for tasks such as classification and target detection [29], with few focused on anomaly detection. Ref. [15] proposes a VAE band selection network based on structural similarity (SSBS). It retains bands highly similar to a fused latent feature map, improving anomaly–background separability for anomaly detection, yet it does not

explicitly control inter-band redundancy, which may lead to a concentrated distribution of the selected subset.

In summary, current band selection methods for anomaly detection often lack sufficient modeling of the multi-scale spatial–spectral features and constraints from anomaly–background priors, which results in a larger number of bands in the optimal subset and poor representation of the selected band subset. Additionally, the selection results are easily influenced by original data and detection algorithms, making it challenging to achieve stable improvements in anomaly detection performance.

To address these issues, this paper proposes a reliable band selection method for anomaly detection based on an autoencoder spectral reconstruction network. The method utilizes an autoencoder spectral reconstruction model to characterize complex backgrounds and employs a structure combining 1D-CNN and Transformer to fully explore the interdependencies between spectral bands [30]. From the perspective of anomaly detection task characteristics, an anomaly–background separability constraint is set in the reconstruction loss using a small number of anomalous spectral samples. A spectral attention mechanism is used to quantify the contribution of different bands in the anomaly–background separation process, identifying the positions of discriminative bands. Considering the strong correlation between adjacent hyperspectral bands, an adaptive band allocation algorithm is also designed to optimize the output band subset.

1. A 1D-CNN and Transformer combined AE network is designed to model the background spectral space. By adding local and global receptive fields, the ability of the AE to represent the internal dependencies of spectral sequences is enhanced, further improving the representational completeness of the selected band subset.
2. An anomaly–background separability constraint is designed to seamlessly integrate the anomaly detection task prior into network optimization. By balancing the minimization of input data reconstruction constraint and anomaly–background separability constraint, dual needs of background modeling and differentiated feature prominence are achieved.
3. Spectral contribution vectors are used to modulate training data, quantifying the spectral contribution distribution from network optimization. An adaptive band allocation algorithm is integrated to overcome issues of neighborhood concentration and unevenness in band contribution distribution, enhancing the effectiveness and stability of the selected band subset.

The rest of this article is organized as follows. Section 2 briefly introduces the research background. Section 3 explains the motivation. Section 4 details the specific structure of the proposed network. Section 5 conducts extensive testing of the proposed network from multiple perspectives, compares it with existing advanced methods, and discusses the experimental results. Section 6 provides a summary.

2. Preliminaries

2.1. Related Work

2.1.1. CNN–Transformer Autoencoder Network

The autoencoder network (AE) excels in characterizing complex backgrounds by extracting hidden features of HSI in a low-dimensional space, which makes it highly effective for HSI dimensionality reduction and reconstruction tasks [25,31,32]. An AE consists of two main components: an encoder and a decoder. It estimates parameters by minimizing the loss between the input data and the reconstructed data [26]. For the band selection task, which fundamentally involves extracting internal correlations of spectral features, we use one-dimensional spectral sequences as input data. To further enhance the

representational capability of the AE structure, this study employs a combination of CNN and Transformer within the AE framework.

Primarily, local features are critical for discriminative band selection. The 1D-CNN has demonstrated precise and efficient performance in extracting one-dimensional hyperspectral features [33–35]. Embedding a 1D-CNN into the autoencoder structure allows for projecting local features into a lower dimension, adding local receptive fields to the network and capturing the shape features of spectral sequences.

Secondarily, after transforming the data into embedded features via 1D-CNN, a Transformer module is employed to perform positional encoding on the one-dimensional spectral sequences. In recent years, many approaches have leveraged the attention mechanism of Transformers to mine joint spatial–spectral information. For example, ref. [36] proposes a Fourier conditional mask-based mixed attention (FCMMA) method for HAD, whose performance is more effective and superior compared to other similar methods. Ref. [37] applies it to cross-domain small-object detection, which can effectively detect targets in HSIs. The attention focuses on learning long-range dependencies to obtain better feature representation [38,39], enhancing the model's interpretability in the band selection task:

$$Attention(Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (1)$$

where Q, K, V is a set of query, key, and value vectors generated for the input data through linear transformations, and d_k represents the dimension of the input.

2.1.2. Band Selection Based on Deep-Learning

The key to using deep learning for band selection lies in ensuring that the abstract, high-level semantic features extracted by the network are accurately reflected in the positional information of the input bands [40]. To enhance the interpretability of deep learning-based band selection and clarify the specific role of each band, several approaches have been proposed. For instance, the contextual multi-scale convolutional neural network (CM-CNN) band selection model [41] employs weighted feature maps extracted by a CNN to generate a band contribution map, thereby determining the priority of bands at different positions. Methods such as [42,43] attempt to train spectral attention mechanism weights to reflect the contribution of each spectral band.

While these methods have achieved commendable results, they often overlook the strong correlation between hyperspectral bands and the discontinuous distribution of noise bands. This oversight limits the effectiveness and stability of the band selection results. To address this, our study designs an adaptive band allocation algorithm based on quantified band contributions from the spectral attention mechanism, optimizing the internal structure of the selected band subset.

2.2. Motivation

The background is the majority of pixels in the image, featuring highly similar and continuous spectra with a broad and complex distribution. Accurately representing the background's spectral feature space is crucial for selecting the most representative band subset [44,45]. However, band selection is an NP-hard problem [46], and no prior information exists about the optimal band subset for training deep learning models. To tackle this challenge, we reformulate background modeling as a spectral reconstruction problem under the assumption of hyperspectral sparsity. Specifically, we design a self-supervised background reconstruction network to capture complex background distributions. Considering the characteristics of hyperspectral anomaly data, we use a large volume of background

samples as training data to ensure the network captures consistent and accurate background spectral representations.

The anomalies are characterized by a small number of pixels and significant spectral differences from the background. Anomalies generally lack a precise definition, but in specific application scenarios—such as fire locations in forests [47] or vehicles in border patrol operations [48]—they may have a rough identifiable range. Therefore, we estimate the mean of anomalies from a priori spectral libraries to serve as an easily accessible semi-supervised signal. Band selection is used in the preprocessing stage to provide a smaller and more stable subset of selected bands for downstream anomaly detection tasks.

3. Methodology

In this section, we elaborate the proposed band selection model for anomaly detection (ADBS-Net). The specific flowchart is shown in Figure 1. Initially, we embed a spectral attention module at the input of the model to determine band contributions. Subsequently, the backbone AE network combines 1D-CNN and Transformer to achieve spectral reconstruction, trained under the constraint of anomaly–background separability. Finally, we design an adaptive band allocation algorithm to derive the final band subset.

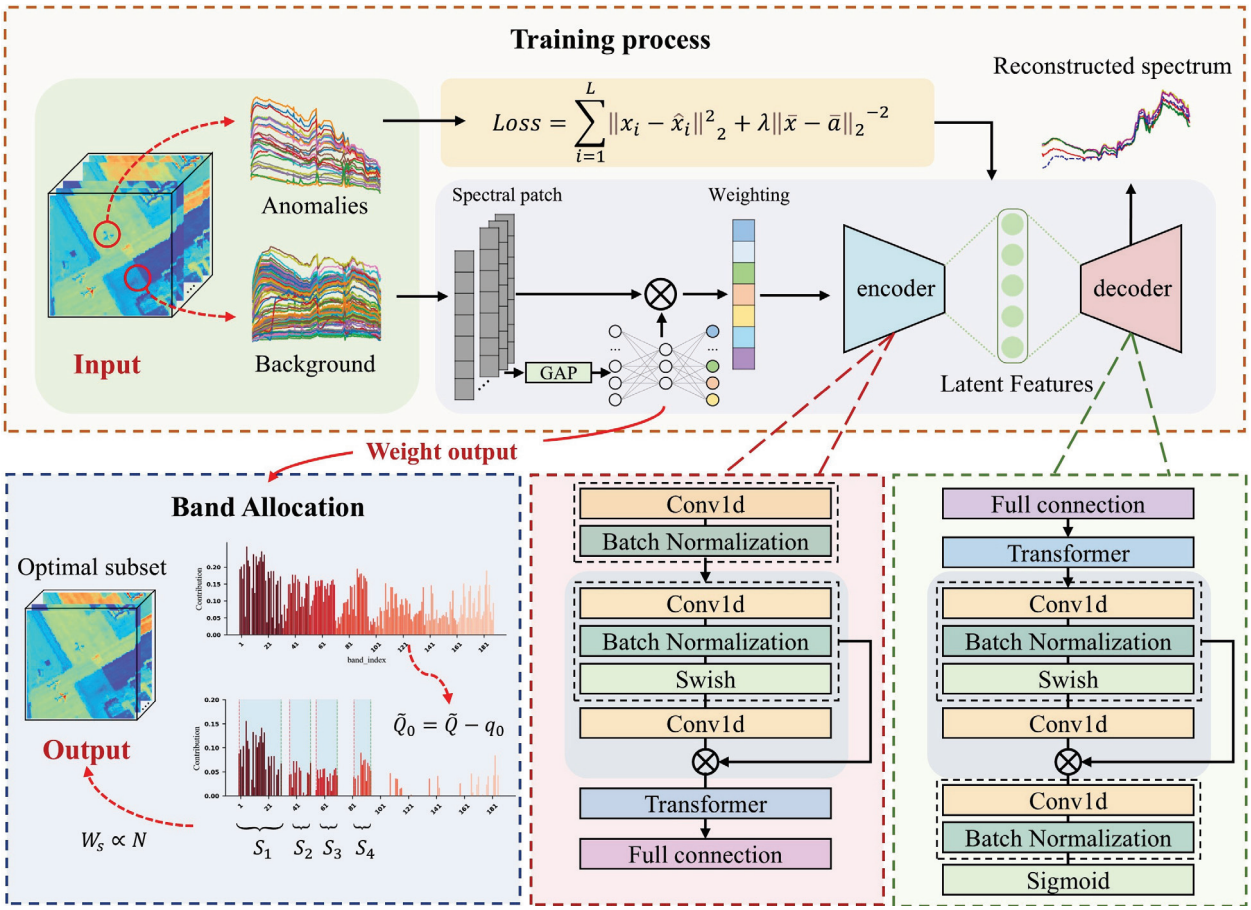


Figure 1. Architecture of the proposed ADBS-Net. ADBS-Net includes the training process and band allocation process. The figure shows the specific data flow and the structure of the AE network.

3.1. Spectral Attention Module

To fully exploit the contribution of different bands to the anomaly detection task, we embed a spectral channel attention module in the first layer of the overall network. The module is primarily divided into two parts: squeeze and excitation [49]. Initially, background spectral samples with dimensions $R^{P \times C}$ are fed into a global average pooling

layer, where P represents the number of samples per training batch, and C denotes the number of spectral channels, resulting in input data $x \in R^{1\times C}$.

To fully explore the interdependencies between spectral channels, the input spectral data is mapped through two cascaded fully connected layers. The mapping process can be expressed as

$$\tilde{Q} = \sigma(W_2\delta(W_1 x)) \tag{2}$$

Here, $W_1 \in R^{\frac{C}{r}\times C}$ and $W_2 \in R^{C\times\frac{C}{r}}$ denote the weights of the two fully connected (FC) layers. δ is a rectified linear unit (ReLU), and the sigmoid function σ is used to constrain the weight output of the spectral attention layer.

Finally, the adjusted C dimensional weight output $\tilde{Q}$ is applied to the original data x, effectively weighting the original input:

$$x^* = F_{scale}(x_{input}, \tilde{Q}) = x \cdot \tilde{Q} \tag{3}$$

where F_{scale} refers to spectral-wise multiplication, and x^* is the weighted output of the spectral attention module. Specifically, in the spectral-wise weight $\tilde{Q} = [q_1, q_2, \ldots, q_i, \ldots, q_C]$, each scalar element q_i reflects the contribution level of the corresponding spectral band. We reasonably assume that the larger the q_i, the higher the contribution of the $i\tilde{}th$ spectral band to the task. The structure of the spectral attention module is illustrated in Figure 2.

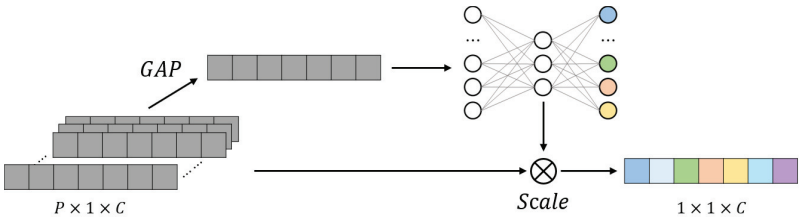


Figure 2. Schematic diagram of the spectral attention module structure.

3.2. Backbone Network

This paper employs an autoencoder-based backbone network to characterize background spectral features through the reconstruction process. It specifically comprises two parts: an encoder and a decoder. During training, the error between the input spectrum and the reconstructed spectrum is reduced via backpropagation. Additionally, a separability constraint between the background and anomalies is incorporated into the loss function, guiding the optimization of the network. This ensures that the beneficial spectral bands for anomaly detection can be highlighted through the spectral attention module.

The backbone network based on the autoencoder structure can be represented as follows:

$$\begin{aligned} L &= J_{encoder}(x, \theta_{encoder}) \\ \hat{x} &= J_{decoder}(L, \theta_{decoder}) \end{aligned} \tag{4}$$

where $J_{encoder}$ and $J_{decoder}$ denote the encoder and decoder functions, respectively, $\theta_{encoder}$ and $\theta_{decoder}$ are their corresponding parameters, L represents the spectral feature representation in the latent layer, and $x \in R^C$ is the reconstructed spectrum output by the network.

The encoder is composed of one-dimensional convolutional layers, convolutional residual blocks, Transformer layers, and a fully connected layer as illustrated in Figure 1. Each one-dimensional convolutional layer is followed by a batch normalization layer. Initially, the input spectral sample x is mapped to a higher dimension through a one-dimensional convolutional layer. Then, spectral features are extracted through cascaded convolutional residual blocks. Each convolutional residual block consists of four one-

dimensional convolutional layers, each followed by batch normalization and the *Swish* activation function, along with an additional one-dimensional convolutional layer, and is configured with skip connections. Subsequently, to fully explore the interdependencies between spectral bands, two Transformer modules are set up. The extracted features undergo dimensional transformation and are finally fed into a fully connected layer to obtain the encoder output L. The specific structure and parameters of the network are detailed in Table 1.

Table 1. ADBS-Net specific parameters table.

Block	Definition	Number	Input Size	Output Size
Input		2	$1 \times 1 \times C$	$1 \times 1 \times C$
Conv	$\begin{bmatrix} \text{Conv1d} & \begin{matrix} kernel\ size = 3 \\ stride = 1 \end{matrix} \\ \text{Batch Normalization} \end{bmatrix}$	2	$1 \times 1 \times C$	$1 \times 1 \times 64$
Res Conv	$\begin{bmatrix} \begin{bmatrix} \text{Conv1d} & \begin{matrix} kernel\ size = 3 \\ stride = 1 \end{matrix} \\ \text{Batch normalization} \\ \text{Swish} \\ \begin{bmatrix} \text{Conv1d} & \begin{matrix} kernel\ size = 3 \\ stride = 1 \end{matrix} \end{bmatrix} \\ \text{Residual Connection} \end{bmatrix} \times 4 \end{bmatrix}$	2	$1 \times 1 \times 64$	$1 \times 1 \times 64$
Transformer	$\begin{bmatrix} \text{Feed Forward} \\ \text{Layer Norm} \\ \text{Multi_head Attention} \quad 8 \times 32 \\ \text{Layer Norm} \end{bmatrix}$	1	$1 \times 1 \times 64$	$1 \times 1 \times 64$
Full connection				

The decoder maps the latent feature representation L to a higher-dimensional space, ultimately outputting the reconstructed spectrum $\hat{x}$. As illustrated in Figure 1, the decoder has a mirror-symmetric structure to the encoder, with a *Sigmoid* activation function applied at the output end.

3.3. Loss Function

The loss function integrates background reconstruction loss and anomaly–background separability regularization, establishing accurate background estimation and mining discriminative features between anomalies and the background class. Crucial, our goal is not to reconstruct more detailed HSI but rather to balance the trade-off between background modeling and anomaly–background separability. This trade-off is explicitly quantified via spectral attention weights, which guide interpretable band selection.

We express the reconstruction loss by minimizing the Mean Squared Error (MSE) function, where M is the total number of samples:

$$\min_{W,b,W',b'} \frac{1}{M}(x_i - \widehat{x}_i)^2 \tag{5}$$

Next, anomaly–background separability refers to the quantifiable difference between the anomalies and background class in the feature space, which we express using Euclidean distance. We apply a prior constraint in the form of the anomaly mean vector $\bar{a}$. Anomalies spectra can be based on a prior spectral library or manual visual inspection; after sparse

sampling, their mean value is taken. This aligns with the reality of scarce abnormal spectral samples, where anomaly samples are not utilized as training data:

$$\|\bar{x} - \bar{a}\|_2 \tag{6}$$

where $\bar{x}$ denotes the mean of the reconstructed spectra.

Applying a prior constraint in the form of $\bar{a}$ aligns with the reality of the scarcity of anomalous spectral samples. Based on the Lagrangian formulation, the anomaly–background separability constraint is incorporated into the reconstruction loss using a hyperparameter λ. This balances the dual requirements of minimizing the reconstruction loss of the input data and the loss function for the anomaly detection task constraints:

$$Loss = \frac{1}{M}\sum_{i=1}^{P}(x_i - \hat{x}_i)^2 + \frac{\lambda}{\|\bar{x} - \bar{a}\|_2} \tag{7}$$

In this context, to ensure that both terms remain on the same scale during model optimization, the constraint is applied in a reciprocal form. The goal is to minimize the first term, the reconstruction loss, while maximizing the second term, the anomaly–background separability constraint.

3.4. Adaptive Band Allocation

After model training, we obtain the band contribution vector $\tilde{Q}$. To highlight the variations, we first perform mean-centering, where q_0 is the average of bands contribution:

$$\tilde{Q_0} = \tilde{Q} - q_0 \tag{8}$$

For $q_i \in \tilde{Q_0}$, if $q_i > 0$, the $i\tilde{}th$ band is considered a beneficial band; conversely, if $q_i < 0$, the $i\tilde{}th$ band is considered a non-beneficial band. The band contribution obtained from training is not suitable for direct band selection due to the following two factors:

Primarily, numerous studies [23,50,51] have proven that HSIs have strong correlations between adjacent spectral bands. The spectral correlation matrix for Sandiego data is shown in Figure 3. We observe a strong correlation in neighborhood intervals within the HSI, which causes bands within the same interval to have similar contributions. Specifically, bands are divided into intervals $B_h(q_i > 0)$, where beneficial bands are continuously distributed, and intervals $B_l(q_i < 0)$, where non-beneficial bands are continuously distributed. If band selection is performed by directly sorting band contributions, the selection results will be concentrated within the same B_h, greatly reducing the information content of the selected subset. Therefore, we first locate these B_h, and then decide the number of bands to select from each B_h based on their contribution levels.

Secondarily, the band contribution quantified through the spectral attention module is not flat. Due to factors such as the discontinuous distribution of noise between HSI bands, band contributions may exhibit local jumps, resulting in lower confidence in single-band contribution results, which affects the effectiveness and stability of the selected band subset. To address this, we set a protection bandwidth τ to overcome the impact of local jumps in q_i on the delineation of B_h and to exclude high-contribution bands with low confidence. A schematic diagram of B_h delineation is shown in Figure 4.

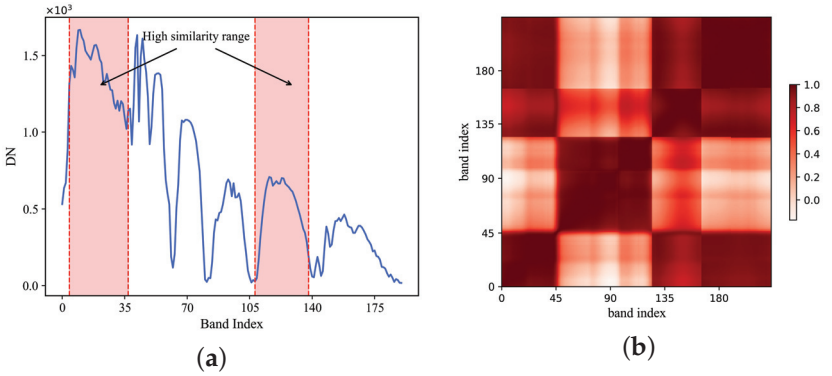


Figure 3. Neighborhood spectral correlation of the Sandiego dataset. (**a**) Schematic diagram of a strong correlation area in the spectral neighborhood. (**b**) Sandiego dataset spectral correlation intensity matrix.

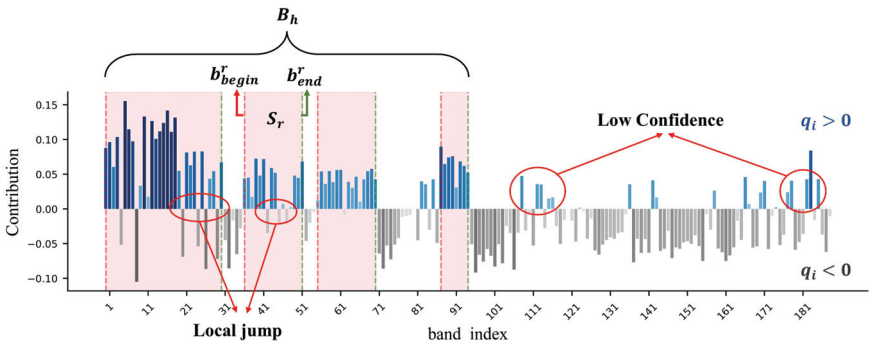


Figure 4. Band weight $\tilde{Q}_0$ distribution map.

Specifically, assuming the target selected band subset consists of L bands, and $B_h = [S_1, S_2, \ldots, S_r, \ldots, S_K]$, to determine the number of intervals K and the specific range of each interval S_r, follow these steps: to begin with, we traverse $\tilde{Q}_0$ and set a protection bandwidth τ. When the number of consecutive high-value bands exceeds the protection bandwidth, mark it as the start of a band interval b^r_{begin}. When the number of consecutive low-value bands exceeds the protection bandwidth, mark it as the end of a band interval b^r_{end}. And we define the interval $S_r = (b^r_{begin}, b^r_{end})$. Next, we calculate the interval weight W_r corresponding to S_r based on $\tilde{X}_0$:

$$W_r = \sum_{i=b^r_{begin}}^{b^r_{end}} q_i \tag{9}$$

The higher the interval weight W_r, the more bands are selected from the interval S_r:

$$N_r = \frac{W_r}{\sum_{i=1}^{K} W_i} \times L \tag{10}$$

Finally, we select the corresponding number of bands from each interval to form the optimal band subset $J_{optimal}$:

$$\begin{aligned} J_r &= argmax_{N_r}\{q^r_{begin}, \cdots, q_{b^r_{end}}\} \\ J_{optimal} &= \{J_1, J_2, \cdots, J_K\} \end{aligned} \tag{11}$$

Here, $argmax_{N_r}$ is used to denote finding the indices J_r of the top N_r maximum values in the set. The specific process of the post-processing algorithm is shown in Algorithm 1.

Algorithm 1: Post processing of band weights.

Input: Bands weights $\tilde{Q}_0 = [q_1, q_2, \ldots, q_C]$; Total Number of Bands C; Selected band size L; Guard band τ;
Output: The index of optimal selected bands

1. Valuable band count $m = 0$; B_h index $l = 1$;
2. Guard band $\tau = 3$;
3. Valuable bands section $B_h = [S_1, S_2, \ldots, S_r, \ldots]$, where $S_r = (b^r_{begin}, b^r_{end})$, b^r_{begin} is the beginning of the r-th section, b^r_{end} is the end of the r-th section;
4. B_h starting signal $s = 1$
5. for $i \leftarrow 1$ to C do
   If $m \leq \tau$ then:
   | If $q_i \times s > 0$ then $m \leftarrow m + 1$
   | else $m \leftarrow 0$
   else:
   | If $s = -1$ then:
   | | $b^l_{end} \leftarrow i$;$l \leftarrow l + 1$; $m = 0$; $s = 1$
   | else:
   | | $b^l_{beigin} \leftarrow i$; $m = 0$; $s = -1$
   | end
   end
6. Calculate bands weight of S_r according to Equation (9).
7. Calculate the number of selected bands for S_r according to Equation (10).
8. Select the corresponding number of bands from S_r according to Equation (11).
9. Obtain the index of optimal band subset.

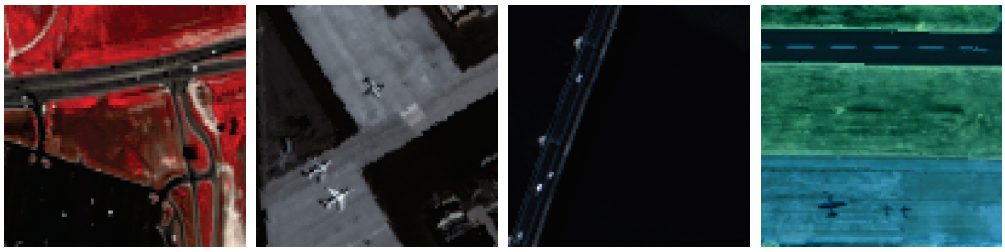
4. Experimental Analysis

In this section, we conduct comparative experiments on seven band selection methods involving anomaly detection across four real-world datasets to demonstrate the effectiveness and robustness of our proposed ADBS-Net. The pseudocolor images of the four datasets are shown in the Figure 5.

4.1. Dataset Description

With reference to previous studies [44,45,52,53], we select four datasets that cover different backgrounds, anomaly types, and anomaly sizes.

- HYDICE Urban Dataset: It was captured by the HYDICE sensor (Naval Research Laboratory, NRL) over an urban area, Washington, DC, USA, in 1995. A region with a size of 80 × 100 pixels covering 175 spectral bands is selected as the test image, where 21 pixels are identified as anomalies, which were roofs and cars.
- Sandiego Dataset: It was captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor (NASA Jet Propulsion Laboratory, JPL) from San Diego, CA, USA. After removing the low signal-to-noise bands, there are 189 spectral bands used for experiments. It covers an airport area with a spatial pixel size of 100 × 100. Three aircraft are considered anomalies of interest with 57 pixels.
- Pavia: It was captured by the reflective optics system imaging spectrometers (ROSIS-o3) sensor (University of Bremen) over a beach area of Pavia, Italy, in which the anomalies mainly consist of some vehicles. The spatial size is 100 × 100, and 102 bands are reserved for experiments.
- Gulfport: It was captured by the AVIRIS about the airport region of Gulfport, MS, USA. It contains 191 spectral bands and 100 × 100 spatial size. This image includes a highway, an airport runway and vegetation. Three airplanes in various scales with 60 pixels are identified as anomalies.



(**a**) (**b**) (**c**) (**d**)

Figure 5. Pseudocolor images of four datasets. (**a**) HYDICE_urban, (**b**) sandiego_plane, (**c**) Pavia, (**d**) Gulfport.

The data encompasses detection results in both complex and simple backgrounds, including scenarios with single and multiple anomalies.

4.2. Experimental Setup

4.2.1. Comparison Methods

We compared the proposed ADBS-Net with seven advanced band selection methods commonly used for anomaly detection, including ONR [23], RDBSMaxSR [27], MBTBR_BS [26], EHBS [24], HADUBS [25], and SSBS [15]. ONR is a search-based method. RDBSMaxSR, EHBS, and HADUBS are rank-based methods. MBTBR_BS is a clustering-based method, and SSBS is a method based on deep learning. To further compare the effectiveness of band selection methods, we also introduce uniform band selection (UBS) for comparison, in order to analyze the differences between the results obtained through random uniform selection and those achieved through deliberate selection. These parameters of the compared methods are set as the recommended parameters in accordance with the original literature. The specific parameters are shown in Table 2.

Table 2. Parameter settings table for state-of-the-art methods.

Method	Values	Significance
ONR	$\tau = 0.1$	reconstruction error threshold
RDBSMaxSR	$\rho = 1$ $d_c = 6\%$	density threshold cut off distance
EHBS	$t = 10$ $\alpha = 0.15$	band sampling interval separation threshold
HADUBS	$\epsilon = 96$ $\psi = 25$	anomaly filtering parameter pixel threshold
SSBS	Layer = 2, Nodes = 20 $\lambda = 0.05$	layers and hidden nodes settings spectral angle mapper weight

To further validate the effectiveness of ADBS-Net in practical applications, we refer to previous research [25,53] and integrate the RX [54] detector into the band selection methods for unified quantitative and qualitative analysis. The RX is widely applicable and considered the benchmark anomaly detection algorithm for HSI. We aim to evaluate the performance improvement of the band selection method in anomaly detection through this.

4.2.2. Evaluation Metrics

To quantitatively assess band selection performance, we employ three-dimensional Receiver Operating Characteristic (3D-ROC) curves and their projected Area Under the Curve (AUC) metrics [55]. The 3D-ROC surface, constructed using the detection threshold

τ, detection probability (P_D), and false alarm rate (P_F), comprehensively characterizes the anomaly detectors' capability to discriminate anomalies from background. Three 2D-ROC curves are generated by projecting the 3D surface onto the (P_D,P_F), (P_D,τ), and (P_F,τ) planes, yielding three corresponding AUC metrics: AUC(P_D,P_F), AUC(P_D,τ), and AUC(P_F,τ). Further generating five metrics based on these three AUCs, their formulas are

$$0 \leq AUC_{TD} = AUC_{(D,F)} + AUC_{(D,\tau)} \leq 2 \tag{12}$$

$$-1 \leq AUC_{BS} = AUC_{(D,F)} - AUC_{(F,\tau)} \leq 1 \tag{13}$$

$$-1 \leq AUC_{ODP} = AUC_{(D,F)} + AUC_{(D,\tau)} - AUC_{(F,\tau)} \leq 2 \tag{14}$$

$$-1 \leq AUC_{TD-BS} = AUC_{(D,\tau)} - AUC_{(F,\tau)} \leq 1 \tag{15}$$

$$0 \leq AUC_{SNPR} = \frac{AUC_{(D,\tau)}}{AUC_{(F,\tau)}}. \tag{16}$$

AUC_{TD} is introduced to quantify the ability to accurately recognize anomalies. AUC_{BS} is adopted to assess the capability to suppress false positives caused by background interference. The AUC_{ODP} provides an integrated evaluation of both anomalies detectability and background suppression performance. AUC_{TD-BS} focuses on the effectiveness in identifying anomalies within a noisy background, taking into consideration both the detection accuracy and background suppression. AUC_{SNPR} is used to characterize the ability to distinguish signals from noise.

4.3. Parameter Determination

In ADBS-Net, we analyze the impact of the anomaly–background separability constraint weight λ on the experimental results. We set the range of λ from 0.1 to 0.5, and the corresponding band selection results on the RX algorithm's AUC(P_D, P_F) score curve are shown in Figure 6. For the AUC(P_D, P_F), we take the average with a band subset size ranging from 15 to 50, in intervals of 5. The AUC(P_D, P_F) score illustrates the model's performance variations under different λ values. By conducting experiments on different datasets with controlled variables only changing the anomaly–background separability constrained loss function, it can be seen that the best experimental results are achieved when λ is set to approximately 0.19.

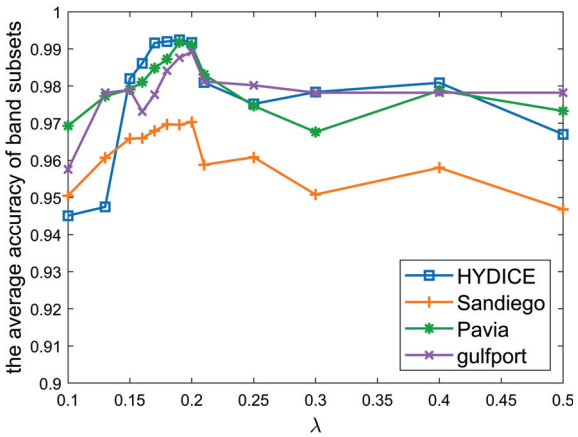


Figure 6. *AUC* versus λ variation curve.

All experiments are carried out using PyTorch (v1.6.0, Python 3.8.0, and CUDA 10.0) in a single NVIDIA RTX 3060 GPU with 12 GB memory. The parameters for the training process of our proposed ADBS-Net are detailed in Table 3.

Table 3. ADBS-Net specific training parameters table.

Hyperparameter	Value Used in Experiments
λ	0.19
Optimizer	Adam
Learning rate	0.001
Batch size	128
Training Epochs	100

4.4. Band Subset Size Analysis

First, we conduct a unified analysis of the performance of various band selection methods across different band subset sizes. Using RX as the baseline, we vary the band subset sizes from 2 to 50 and perform anomaly detection across different datasets while recording the variations in AUC(P_D,P_F).

As shown in Figure 7, ADBS-Net achieves superior stability and effectiveness across all datasets. Notably, the performance of ADBS-Net reaches a stable level when the band subset size exceeds 10. MBTBR_BS and HADUBS attain comparable performance to ADBS-Net with subset exceeding 20 bands. The deep learning-based SSBS requires larger band subset size compared to other methods for stable performance. Primarily, competing methods exhibit unstable adaptability to datasets with varying noise levels and anomaly–background separability. Secondarily, their performance fluctuates significantly with small subset sizes (<20 bands). In contrast, ADBS-Net can maintain stable and superior performance across all datasets, regardless of band subset size.

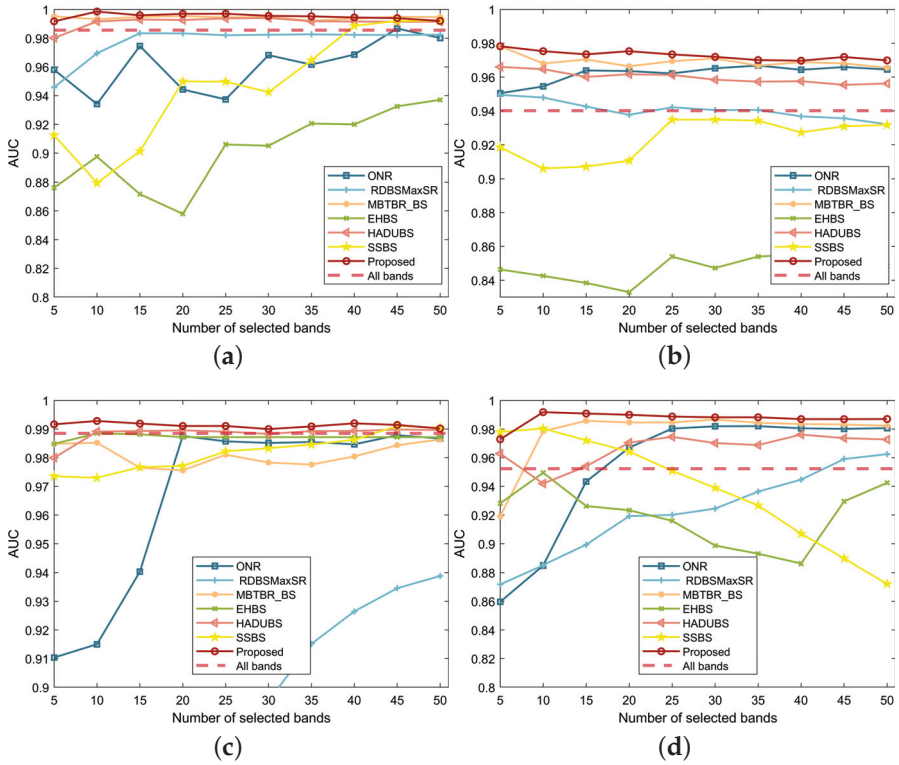


Figure 7. AUC variation under different band subset sizes: (**a**) HYDICE_urban; (**b**) Sandiego; (**c**) Pavia; (**d**) Gulfport.

Based on practical experience and experimental results, 20 bands may better meet the requirements of most competing band selection methods and anomaly detection tasks. To better observe the superiority of ADBS-Net, we cascade the RX detector with the band subset size set to 20, and compare the performance improvements of band selection methods.

4.5. Results

4.5.1. HYDICE_urban

The AUC results of RX anomaly detection on the HYDICE dataset using different band selection methods are shown in Table 4, where red represents the best and green represents the second best, and up and down arrows denote that higher and lower values are preferable, respectively. Their AUC curves are illustrated in Figure 8 and the specific band subset contents are shown in Table 5.

Table 4. AUC results of RX detector on HYDICE_urban.

Method	AUC (P_D, P_F) ↑	AUC (P_D, τ) ↑	AUC (P_F, τ) ↓	AUC_{TD} ↑	AUC_{BS} ↑	AUC_{ODP} ↑	AUC_{TD-BS} ↑	AUC_{SNPR} ↑
Full bands	0.9857	0.2404	0.0351	1.2261	0.9506	1.1910	0.2053	6.8476
ONR	0.9443	0.3122	0.0394	1.2565	0.9049	1.2171	0.2728	7.9282
RDBSMaxSR	0.9933	0.1850	0.0171	1.1783	0.9762	1.1612	0.1679	10.8187
MBTBR_BS	0.9954	0.3295	0.0210	1.3249	0.9744	1.3039	0.3085	15.6905
EHBS	0.8580	0.0525	0.0194	0.9105	0.8386	0.8911	0.0331	2.7041
HADUBS	0.9927	0.2217	0.0370	1.2144	0.9557	1.1774	0.1847	5.9932
SSBS	0.8461	0.0622	0.0204	0.9083	0.8257	0.8879	0.0418	3.0490
UBS	0.9956	0.3055	0.0219	1.3011	0.9737	1.2792	0.2836	13.9475
ADBS-Net	0.9969	0.3575	0.0190	1.3544	0.9779	1.3354	0.3385	18.8158

Table 5. The 20 selected band subsets on HYDICE_urban.

Methods	Index of the Selected Bands
Full bands	1:1:175
ONR	75, 80, 88, 91, 95, 104, 107, 112, 115, 118, 122, 126, 130, 134, 138, 142, 146, 150, 154, 158
RDBSMaxSR	12, 14, 16, 17, 18, 19, 23, 24, 38, 39, 40, 41, 42, 43, 44, 46, 47, 48, 49, 50
MBTBR_BS	1, 4, 7, 16, 20, 24, 28, 33, 37, 43, 47, 52, 53, 100, 102, 131, 138, 143, 155, 160
EHBS	57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95
HADUBS	1, 2, 9, 10, 18, 19, 37, 40, 45, 46, 47, 53, 77, 79, 84, 88, 118, 131, 145, 160
SSBS	58, 64, 65, 66, 67, 68, 69, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 84
UBS	1, 10, 19, 28, 37, 46, 55, 64, 73, 82, 91, 100, 109, 118, 127, 136, 145, 154, 163, 175
ADBS-Net	8, 23, 31, 37, 43, 50, 67, 68, 74, 97, 100, 101, 105, 113, 118, 155, 162, 168, 169, 171

It can be observed that, across all eight AUC metrics, ADBS-Net demonstrates the overall best performance, achieving the highest values in seven out of eight metrics. The methods AUC(P_F, τ) are the next best, with results being very close to the optimum. MBTBR_BS also yields competitive results, though it remains slightly inferior to ADBS-Net. These findings indicate that the band selection performed by ADBS-Net significantly enhances the anomaly detection capability of the detector. Furthermore, the substantial improvements in AUC_{TD-BS} and AUC_{SNPR} suggest that ADBS-Net provides the greatest benefit in terms of suppressing background interference and reducing false alarms.

Figure 9 visually presents the RX anomaly detection results for different band selection methods and compares them with the detection results using all bands. By evaluating these methods based on the degree of background suppression and the prominence of anomalies, it can be observed that the result of ADBS-Net is the closest to the GT. Compared to other band selection methods, ADBS-Net achieves the best overall performance by effectively highlighting the target while suppressing the background to maintain a low false alarm rate.

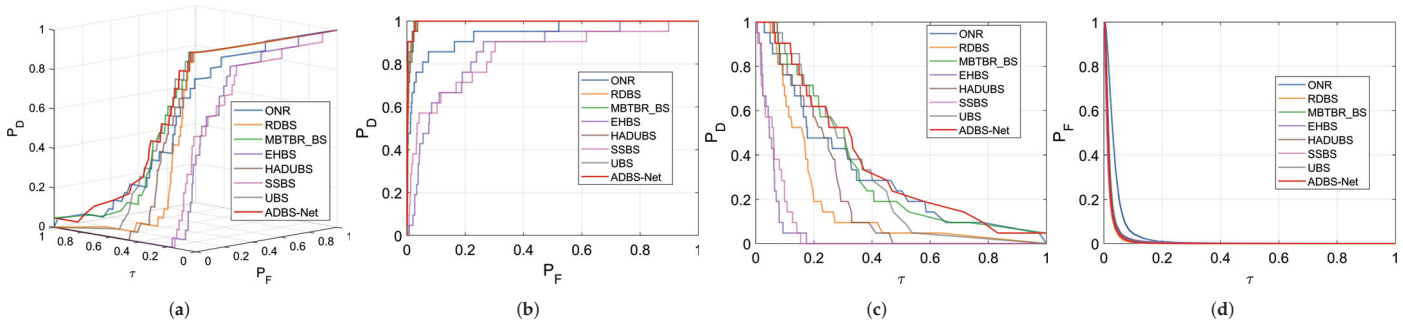


Figure 8. ROC curves of different methods with 20 bands in HYDICE_urban: (**a**) 3D ROC curves (P_D, P_F, τ). (**b**) Corresponding 2D ROC curves (P_D, P_F). (**c**) Corresponding 2D ROC curves (P_D, τ). (**d**) Corresponding 2D ROC curves (P_F, τ).

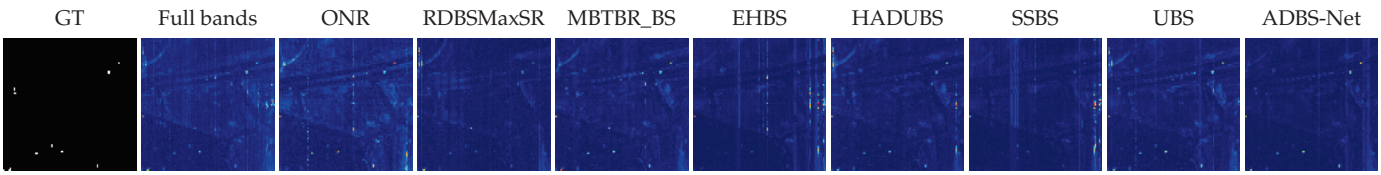


Figure 9. Detection results of GRX using different band selection methods on HYDICE_urban.

4.5.2. Sandiego_plane

As shown in Table 6, ADBS-Net achieves the best results in seven out of eight metrics. Although AUC(P_F, τ) is relatively high, a comprehensive evaluation based on AUC$_{BS}$, AUC$_{ODP}$, and AUC$_{TD-BS}$ demonstrates that ADBS-Net possesses the strongest background interference resistance capability. In terms of AUC(P_D, P_F), AUC(P_D, τ), and AUC$_{TD}$, ADBS-Net also provides the greatest improvement in the detector's anomaly detection performance. Additionally, the superior AUC$_{SNPR}$ indicates that ADBS-Net effectively distinguishes anomalies from noise. As illustrated in Figure 10, the ROC curve of ADBS-Net also occupies the optimal position in the upper-left corner. The specific band subset contents are shown in Table 7.

Table 6. AUC results of RX detector on sandiego_plane.

Method	AUC (P_D, P_F) ↑	AUC (P_D, τ) ↑	AUC (P_F, τ) ↓	AUC$_{TD}$ ↑	AUC$_{BS}$ ↑	AUC$_{ODP}$ ↑	AUC$_{TD-BS}$ ↑	AUC$_{SNPR}$ ↑
Full bands	0.9403	0.1778	0.0589	1.1181	0.8814	1.0592	0.1189	3.0195
ONR	0.9635	0.0871	0.0139	1.0506	0.9496	1.0367	0.0732	6.2669
RDBSMaxSR	0.9183	0.0767	0.0259	0.9950	0.8924	0.9691	0.0508	2.9633
MBTBR_BS	0.9663	0.1079	0.0177	1.0742	0.9486	1.0565	0.0902	6.0960
EHBS	0.8330	0.1780	0.0460	1.0110	0.7870	0.9650	0.1320	3.8696
HADUBS	0.9616	0.1079	0.0174	1.0695	0.9442	1.0521	0.0905	6.2011
SSBS	0.9106	0.0924	0.0308	1.0030	0.8798	0.9722	0.0616	2.9961
UBS	0.9627	0.1101	0.0161	1.0728	0.9466	1.0567	0.0940	6.8373
ADBS-Net	0.9753	0.1892	0.0230	1.1645	0.9523	1.1415	0.1662	8.2261

Table 7. The 20 selected band subsets on sandiego_plane.

Methods	Index of the Selected Bands
Full bands	1:1:189
ONR	4, 13, 21, 27, 36, 48, 58, 70, 74, 82, 90, 96, 100, 106, 118, 126, 133, 145, 152, 155
RDBSMaxSR	12, 14, 16, 17, 18, 19, 21, 22, 23, 24, 39, 41, 42, 43, 44, 46, 47, 48, 49, 50
MBTBR_BS	3, 7, 12, 17, 22, 121, 126, 131, 136, 138, 141, 147, 155, 159, 163, 168, 173, 178, 182, 186
EHBS	150, 151, 152, 153, 154, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170
HADUBS	1, 4, 14, 36, 40, 46, 47, 49, 58, 61, 62, 80, 84, 94, 95, 107, 109, 111, 127, 167
SSBS	6, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75
UBS	1, 10, 19, 28, 37, 46, 55, 64, 73, 82, 91, 100, 109, 118, 127, 136, 145, 154, 163, 189
ADBS-Net	4, 8, 12, 21, 23, 28, 53, 64, 101, 105, 118, 124, 134, 143, 147, 153, 165, 169, 175, 185

As visualized in Figure 11, ADBS-Net exhibits remarkable ability in highlighting anomalies. The three airplanes in the image have been significantly enhanced compared to other methods, especially the airplane at the top. However, it can also be observed that the slightly higher AUC(P_F, τ) of ADBS-Net has resulted in a false alarm on the rooftop in the lower-right corner of the background.

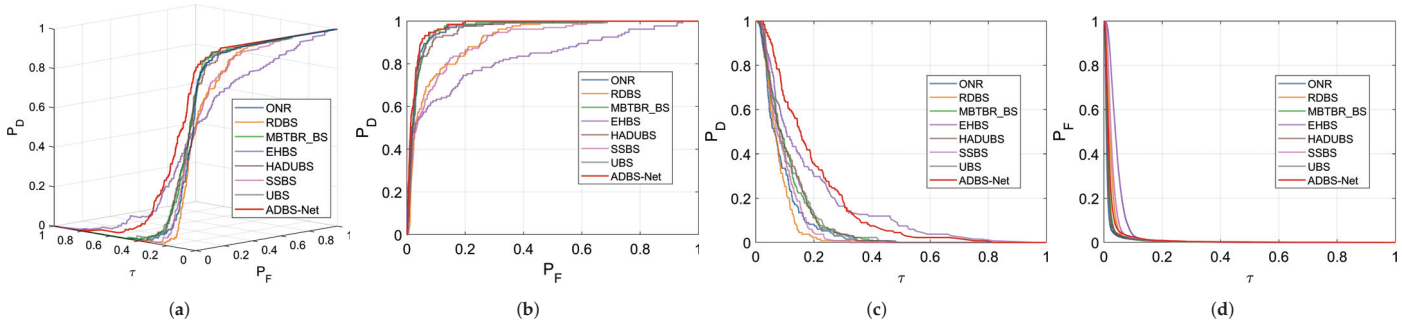


Figure 10. ROC curves of different methods with 20 bands in sandiego_plane: (**a**) 3D ROC curves (P_D, P_F, τ). (**b**) Corresponding 2D ROC curves (P_D, P_F). (**c**) Corresponding 2D ROC curves (P_D, τ). (**d**) Corresponding 2D ROC curves (P_F, τ).

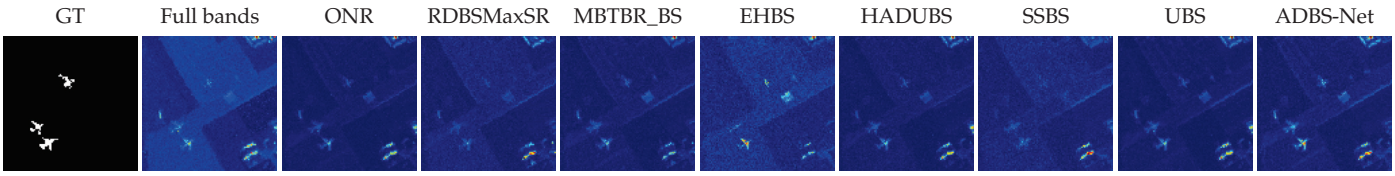


Figure 11. Detection results of GRX using different band selection methods on sandiego_plane.

4.5.3. Pavia

As illustrated in Figure 12 and Table 8, ADBS-Net achieves seven best results and one second-best result, demonstrating overall superior performance. While RDBSMaxSR also performs well, AUC_{BS} indicates that ADBS-Net exhibits stronger resistance to background interference. These results demonstrate that, on the Pavia dataset, ADBS-Net achieves a good balance between anomaly enhancement and background suppression. The specific band subset contents are shown in Table 9.

Table 8. AUC results of RX detector on Pavia.

Method	AUC (P_D, P_F) ↑	AUC (P_D, τ) ↑	AUC (P_F, τ) ↓	AUC_{TD} ↑	AUC_{BS} ↑	AUC_{ODP} ↑	AUC_{TD-BS} ↑	AUC_{SNPR} ↑
Full bands	0.9887	0.1533	0.0259	1.1420	0.9628	1.1161	0.1274	5.9197
ONR	0.9875	0.1721	0.0114	1.1596	0.9761	1.1482	0.1607	15.1053
RDBSMaxSR	0.9811	0.1834	0.0114	1.1645	0.9697	1.1531	0.1720	16.0877
MBTBR_BS	0.9755	0.1816	0.0161	1.1571	0.9594	1.1410	0.1655	11.2863
EHBS	0.9871	0.1049	0.0073	1.0920	0.9798	1.0847	0.0976	14.3699
HADUBS	0.9895	0.1446	0.0107	1.1341	0.9788	1.1234	0.1339	13.5150
SSBS	0.9771	0.1686	0.0224	1.1457	0.9547	1.1233	0.1462	7.5313
UBS	0.9874	0.1628	0.0110	1.1502	0.9764	1.1392	0.1518	14.7982
ADBS-Net	0.9913	0.1923	0.0105	1.1836	0.9808	1.1731	0.1818	18.3143

As shown in Figure 13, the higher AUC(P_F, τ) and lower AUC_{BS} of MBTBR_BS and SSBS result in increased false alarms in road surface areas. In contrast, the ADBS-Net method performs well in both highlighting anomalies and suppressing background interference.

Table 9. The 20 selected band subsets on Pavia.

Methods	Index of the Selected Bands
Full bands	1:1:102
ONR	5, 18, 31, 44, 55, 58, 61, 64, 67, 70, 73, 76, 79, 82, 85, 88, 91, 94, 97, 100
RDBSMaxSR	76, 77, 78, 79, 80, 81, 82, 83, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 101, 102
MBTBR_BS	53, 55, 57, 60, 63, 66, 68, 70, 73, 75, 78, 81, 84, 86, 88, 90, 93, 96, 98, 101
EHBS	5, 7, 9, 11, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 95, 97, 99
HADUBS	1, 6, 14, 19, 29, 31, 33, 34, 37, 38, 40, 49, 51, 71, 72, 74, 85, 89, 93, 95
SSBS	71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90
UBS	1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 71, 76, 81, 86, 91, 102
ADBS-Net	14, 18, 23, 29, 39, 49, 56, 59, 63, 68, 71, 72, 76, 81, 83, 85, 90, 91, 95, 102

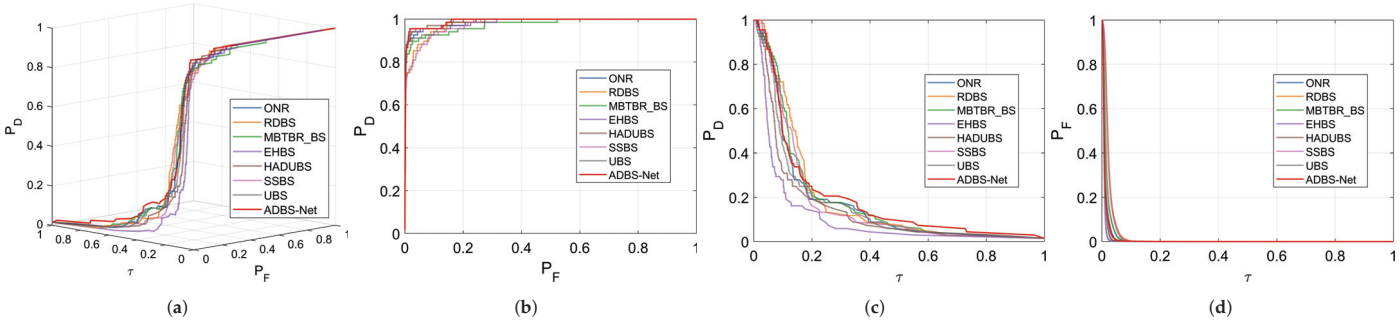


Figure 12. ROC curves of different methods with 20 bands in Pavia. (**a**) 3D ROC curves (P_D, P_F, τ). (**b**) Corresponding 2D ROC curves (P_D, P_F). (**c**) Corresponding 2D ROC curves (P_D, τ). (**d**) Corresponding 2D ROC curves (P_F, τ).

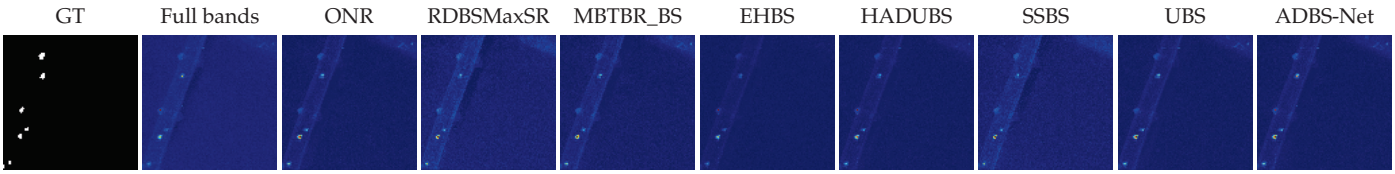


Figure 13. Detection results of GRX using different band selection methods on Pavia.

4.5.4. Gulfport

As shown in Table 10, ADBS-Net achieves the best results in six metrics. However, it does not achieve the highest values in AUC(P_D, τ) and AUC$_{BS}$, indicating that its background suppression capability is not optimal. Nevertheless, the highest levels in AUC(P_D, P_F), AUC(P_D, τ), and AUC$_{TD}$ demonstrate that ADBS-Net provides the most effective enhancement of anomalies. Further examination of AUC$_{ODP}$, AUC$_{TD-BS}$, and AUC$_{SNPR}$ shows that although ADBS-Net does not possess the strongest background suppression, it offers the best overall performance in balancing anomaly detection accuracy and background interference resistance. In contrast, while ONR achieves the best background suppression, it excessively suppresses the background, resulting in the loss of some anomalies as reflected by the lower AUC$_{ODP}$ and AUC$_{TD-BS}$ values. The quantitative analysis above is consistent with the ROC curves presented in Figure 14. The specific band subset contents are shown in Table 11.

As depicted in Figure 15, ADBS-Net significantly enhances the detector's response to anomalies. It is also observed that the background suppression capability of ONR and HADBUS excessively compromises anomaly enhancement. While MBTBR_BS yields quantitative results close to those of ADBS-Net, the runway false alarms observed in the figure indicate that ADBS-Net possesses superior background interference resistance.

When aggregating the results of all band selection methods, they produce false alarms of varying degrees on the runway in Gulfport. According to [55], this may be because the runway occupies a small proportion of the background while exhibiting high anomaly magnitude, despite not being included in the ground-truth annotation.

Table 10. AUC results of RX detector on Gulfport.

Method	AUC (P_D, P_F) ↑	AUC (P_D, τ)↑	AUC (P_F, τ) ↓	AUC_{TD} ↑	AUC_{BS} ↑	AUC_{ODP} ↑	AUC_{TD-BS} ↑	AUC_{SNPR} ↑
Full bands	0.9526	0.0736	0.0248	1.0262	0.9278	1.0014	0.0488	2.9677
ONR	0.9670	0.0747	0.0126	1.0417	0.9544	1.0291	0.0621	5.9294
RDBSMaxSR	0.9292	0.2329	0.0447	1.1621	0.8845	1.1174	0.1882	5.2107
MBTBR_BS	0.9847	0.3339	0.0487	1.3186	0.9360	1.2699	0.2852	6.8556
EHBS	0.9234	0.1604	0.0480	1.0838	0.8754	1.0358	0.1124	3.3417
HADUBS	0.9706	0.1470	0.0233	1.1176	0.9473	1.0943	0.1237	6.3167
SSBS	0.9643	0.2445	0.0795	1.2088	0.8848	1.1293	0.1650	3.0761
UBS	0.9834	0.1673	0.0287	1.1507	0.9547	1.1220	0.1386	5.8317
ADBS-Net	0.9894	0.3659	0.0388	1.3553	0.9506	1.3164	0.3271	9.4330

Table 11. The 20 selected band subsets on Gulfport.

Methods	Index of the Selected Bands
Full bands	1:1:191
ONR	6, 14, 21, 32, 39, 43, 48, 58, 65, 70, 76, 81, 90, 98, 115, 129, 136, 148, 158, 166
RDBSMaxSR	4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 30, 31, 32, 33, 34, 35, 36
MBTBR_BS	7, 12, 17, 21, 26, 31, 34, 65, 68, 79, 85, 91, 97, 100, 103, 170, 174, 179, 185, 190
EHBS	143, 144, 145, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 166
HADUBS	6, 7, 19, 28, 30, 31, 32, 33, 35, 42, 43, 49, 54, 65, 79, 88, 126, 168, 173, 176
SSBS	64, 65, 66, 67, 68, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 99, 100, 105
UBS	1, 11, 21, 31, 41, 51, 61, 71, 81, 91, 101, 111, 121, 131, 141, 151, 161, 171, 181, 191
ADBS-Net	4, 6, 9, 18, 23, 36, 54, 79, 86, 94, 111, 116, 140, 144, 147, 149, 153, 166, 173, 176

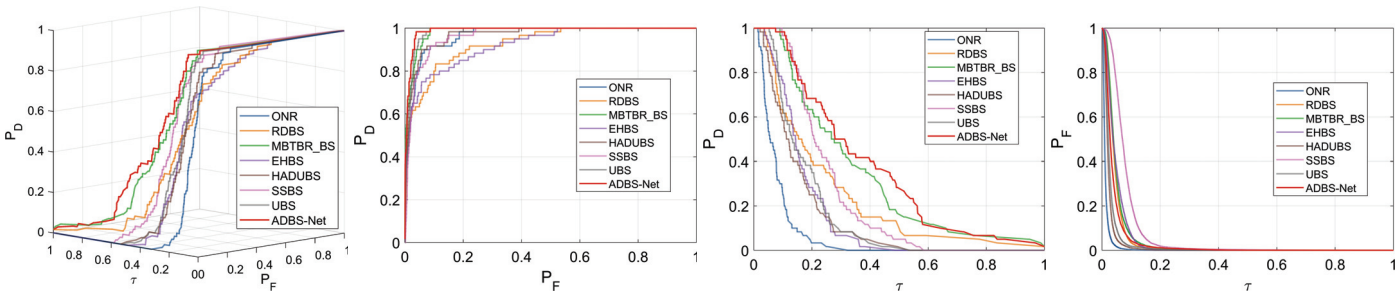


Figure 14. ROC curves of different methods with 20 bands in Gulfport: (**a**) 3D ROC curves (P_D, P_F, τ). (**b**) Corresponding 2D ROC curves (P_D, P_F). (**c**) Corresponding 2D ROC curves (P_D, τ). (**d**) Corresponding 2D ROC curves (P_F, τ).

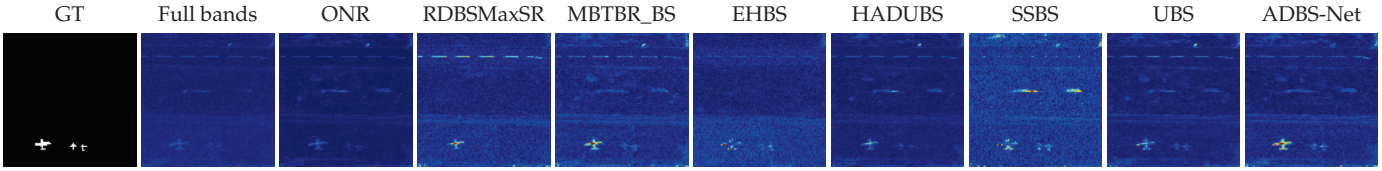


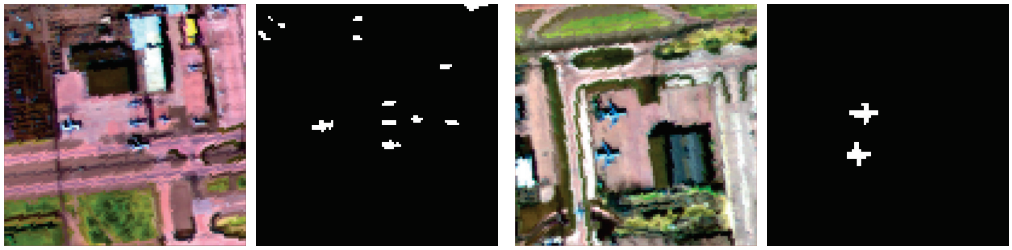
Figure 15. Detection results of GRX using different band selection methods on Gulfport.

5. Discussion

In the last section, we analyze and discuss the key factors that must be considered in practical applications: portability analysis and time consumption analysis of ADBS-Net. We discuss the remaining limitations of the model and how it will function in real-world applications.

5.1. Portability Analysis

To further verify the transferability of the AD-BSNet's band-selection results, we adopt the band set selected on the Gulfport scene (Table 10) and transfer it to two same-scenario airport datasets, Airport-1 and Airport-2, for evaluation. As shown in Figure 16, both Airport-1 and Airport-2 feature airport backgrounds with aircraft targets and, like Gulfport, are captured by the ARIVIS sensor [56]. The resulting 8-AUC scores are reported in Tables 12 and 13. As evidenced by these results, ADBS-Net maintains the best overall performance in terms of the 8-AUC metric, validating the effectiveness of ADBS-Net for band pre-selection.



(**a**) (**b**) (**c**) (**d**)

Figure 16. Airport Data. (**a**) Pseudocolor image of Airport-1, (**b**) Ground truth of Airport-1, (**c**) Pseudocolor images of Airport-2, (**d**) Ground truth of Airport-2.

Table 12. AUC results of RX detector on Airport-1.

Method	AUC (P_D, P_F) ↑	AUC (P_D, τ) ↑	AUC (P_F, τ) ↓	AUC_{TD} ↑	AUC_{BS} ↑	AUC_{ODP} ↑	AUC_{TD-BS} ↑	AUC_{SNPR} ↑
full band	0.8404	0.1841	0.0516	1.0245	0.7888	0.9729	0.1325	3.5691
ONR	0.9347	0.1132	0.0212	1.0480	0.9136	1.0268	0.0921	5.3477
RDBSMaxSR	0.9110	0.1862	0.0461	1.0971	0.8649	1.0510	0.1401	4.0405
MBTBR_BS′	0.9206	0.1248	0.0199	1.0454	0.9007	1.0255	0.1050	6.2804
EHBS	0.8250	0.1295	0.0264	0.9545	0.7986	0.9281	0.1031	4.9112
HADUBS	0.9315	0.1282	0.0190	1.0597	0.9125	1.0407	0.1092	6.7446
SSBS	0.8782	0.0925	0.0331	0.9707	0.8451	0.9376	0.0594	2.7958
UBS	0.9290	0.1250	0.0203	1.0539	0.9087	1.0337	0.1047	6.1710
ADBS-Net	0.9382	0.1624	0.0136	1.1005	0.9246	1.0869	0.1488	11.9330

Table 13. AUC results of RX detector on Airport-2.

Method	AUC (P_D, P_F)↑	AUC (P_D, τ)↑	AUC (P_F, τ)↓	AUC_{TD}↑	AUC_{BS}↑	AUC_{ODP}↑	AUC_{TD-BS}↑	AUC_{SNPR}↑
full band	0.8221	0.0987	0.0424	0.9208	0.7797	0.8784	0.0563	2.3296
ONR	0.8823	0.0386	0.0121	0.9209	0.8702	0.9088	0.0265	3.2001
RDBSMaxSR	0.8657	0.0573	0.0435	0.9229	0.8222	0.8795	0.0138	1.3181
MBTBR_BS′	0.8718	0.0653	0.0233	0.9370	0.8484	0.9137	0.0419	2.7973
EHBS	0.7451	0.1651	0.0782	0.9102	0.6670	0.8321	0.0869	2.1118
HADUBS	0.8851	0.0840	0.0284	0.9691	0.8567	0.9407	0.0557	2.9634
SSBS	0.7745	0.0895	0.0526	0.8641	0.7219	0.8114	0.0369	1.7009
UBS	0.8678	0.0833	0.0302	0.9511	0.8376	0.9209	0.0531	2.7577
ADBS-Net	0.8832	0.1053	0.0176	0.9885	0.8656	0.9709	0.0877	5.9848

5.2. Convergence and Time Analysis

We conduct a comprehensive visualization of ADBS-Net's convergence behavior through 100-epoch training on four datasets. First, the loss convergence curve in Figure 17a–d reveals that the model achieves near-zero loss values after approximately 40 iterations, indicating rapid convergence and stable trainability.

Second, the evolutionary patterns of spectral attention weights are visualized in Figure 17e–h. The vertical axis is the number of training epochs, increasing downward from 1 to 100; the horizontal axis is the sequential band indices from left to right. It can be observed that within the first 40 epochs, the band contributions are randomly scattered

with no obvious heatmap peaks. As the model converges, the contributions of irrelevant spectral bands gradually decrease, while the more informative spectral bands accumulate stronger selection priorities, resulting in distinct heatmap peaks. This dynamic evolution further validates the effectiveness of the ADBS-Net modeling process.

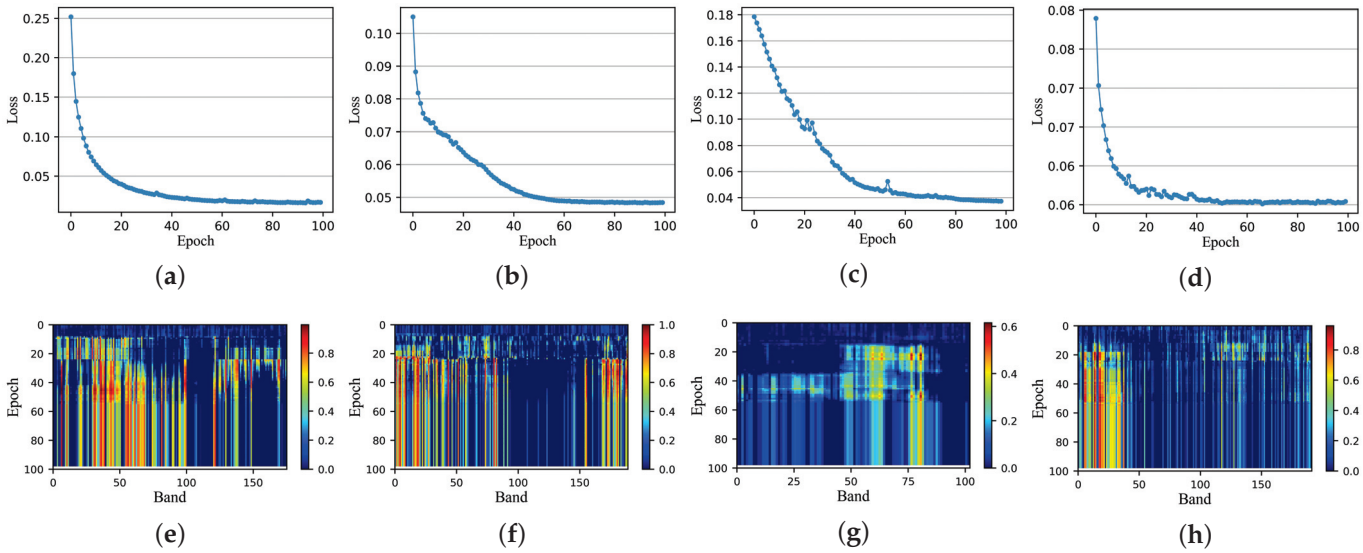


Figure 17. ADBS-Net loss and band weights output convergence diagram: (**a**–**d**) Training loss curve on four datasets. (**e**–**h**) Band weights output by the spectral attention module on four datasets. (**a**,**e**) HYDICE_urban; (**b**,**f**) Sandiego; (**c**,**g**) Pavia; (**d**,**h**) Gulfport.

To evaluate runtime, various band selection methods are executed on the same computer, and their running times are recorded under the condition for selecting 20 bands. Both SSBS and the proposed ADBS-Net are implemented in Python 3.8, while other baseline methods are implemented in MATLAB R2021a. The results are shown in Table 14. Deep learning-based approaches require extensive iterations for model training, resulting in significantly higher computational overhead compared to conventional methods. However, except for HADUBS and RDBS, other comparative methods are search based or clustering based, necessitating complete re-execution when the band subset size changes. ADBS-Net provides flexible band combinations for arbitrary subset sizes through a single execution. This means that after completing one training session, inference can be performed at any band size without the need for retraining. Although its runtime is longer, it remains within a practically acceptable range and is accompanied by optimal band selection performance.

Table 14. Comparison of runtime(s) on four dataset using 20 bands in seconds. (N\A means that the UBS method takes bands at equal intervals, so there is no runtime).

	ONR	RDBS	MBTBR	EHBS	HADUBS	UBS	SSBS	ADBS
HYDICE_urban	4.96	74.56	0.54	0.07	3.92	N\A	87.89	98.85
sandiego_plane	6.48	118.52	0.84	0.08	5.07	N\A	100.02	114.46
Pavia	0.97	61.16	0.33	0.06	2.88	N\A	83.46	89.86
Gulfport	6.34	113.02	0.83	0.07	5.07	N\A	108.97	119.32

Since deep learning models require multiple rounds of training, the overall process is time-consuming. Although ADBS-Net achieves significant performance improvements, the model remains computationally intensive. While this already meets the requirements for band pre-selection tasks, addressing this issue will be a focus of future work.

To our knowledge, semi-supervised band selection for anomaly detection is scarcely explored. ADBS-Net prioritizes background spectral modeling while being regularized by a small set of anomalous spectra. In practice, backgrounds are typically known and can be extensively sampled, whereas anomalies are scarce but their category space is often predefined. Under such prior knowledge, a semi-supervised band-selection module can be used offline to pre-select stable anomaly–background discriminative bands, thereby strengthening downstream detection and guiding deployment.

6. Conclusions

Existing band selection methods for anomaly detection often suffer from high subset redundancy, limited capability to enhance detection performance, and instability. This paper proposes a band selection model designed for specific anomaly detection tasks. First, we introduce an integrated network architecture combining convolutional and transformer-based autoencoders to thoroughly explore inter-spectral dependencies. Additionally, a designed anomaly–background separability constraint is used to highlight discriminative bands. Notably, we employ a band adaptive allocation algorithm to process the spectral attention module weights derived from the model, thereby improving the effectiveness of the band selection results. We have validated the proposed method across different datasets and various band subset sizes, and compared it with existing state-of-the-art methods. The results demonstrate that the proposed approach provides higher gain and more stable performance in anomaly detection. However, the band selection methods based on deep learning are quite time-consuming, which will be a key focus for subsequent optimization.

Author Contributions: Conceptualization, G.Z. and W.W.; Methodology, G.Z., W.W. and X.L.; Validation, G.Z.; Formal analysis, G.Z.; Resources, Y.H. and C.D.; Writing—original draft, G.Z.; Writing—review & editing, W.W. and X.L.; Project administration, Y.H. and C.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 62171040.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Qu, B.; Zheng, X.; Qian, X.; Lu, X. Research progress on hyperspectral anomaly detection. *Natl. Remote Sens. Bull.* **2024**, *28*, 42–54. [CrossRef]
2. Sawant, S.S.; Prabukumar, M. A survey of band selection techniques for hyperspectral image classification. *J. Spectr. Imaging* **2020**, *9*, a5. [CrossRef]
3. Xie, W.; Lei, J.; Liu, B.; Li, Y.; Jia, X. Spectral constraint adversarial autoencoders approach to feature representation in hyperspectral anomaly detection. *Neural Netw.* **2019**, *119*, 222–234. [CrossRef] [PubMed]
4. Chang, C.I.; Chiang, S.S. Anomaly detection and classification for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 1314–1325. [CrossRef]
5. Manolakis, D.; Shaw, G. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Process. Mag.* **2002**, *19*, 29–43. [CrossRef]
6. Raza Shah, N.; Maud, A.R.M.; Bhatti, F.A.; Ali, M.K.; Khurshid, K.; Maqsood, M.; Amin, M. Hyperspectral anomaly detection: A performance comparison of existing techniques. *Int. J. Digit. Earth* **2022**, *15*, 2078–2125. [CrossRef]
7. Shang, X.; Song, M.; Wang, Y.; Yu, C.; Yu, H.; Li, F.; Chang, C.I. Target-constrained interference-minimized band selection for hyperspectral target detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 6044–6064. [CrossRef]
8. Wang, L.; Chang, C.I.; Lee, L.C.; Wang, Y.; Xue, B.; Song, M.; Yu, C.; Li, S. Band subset selection for anomaly detection in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4887–4898. [CrossRef]

9. Wang, Q.; Zhang, F.; Li, X. Optimal clustering framework for hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5910–5922. [CrossRef]
10. Sun, X.; Zhang, H.; Xu, F.; Fu, X. Band selection for specific target detection of hyperspectral imagery. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 3404–3407.
11. Yang, H.; Du, Q.; Chen, G. Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 544–554. [CrossRef]
12. Qu, J.; Du, Q.; Li, Y.; Tian, L.; Xia, H. Anomaly detection in hyperspectral imagery based on Gaussian mixture model. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 9504–9517. [CrossRef]
13. Yu, C.; Lee, L.C.; Chang, C.I.; Xue, B.; Song, M.; Chen, J. Band-specified virtual dimensionality for band selection: An orthogonal subspace projection approach. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2822–2832. [CrossRef]
14. Chang, C.I.; Du, Q.; Sun, T.L.; Althouse, M.L. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 2631–2641. [CrossRef]
15. Xie, W.; Li, Y.; Lei, J.; Yang, J.; Chang, C.I.; Li, Z. Hyperspectral band selection for spectral–spatial anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 3426–3436. [CrossRef]
16. Chang, C.I.; Li, Y.; Hobbs, M.C.; Schultz, R.C.; Liu, W.M. Progressive band processing of anomaly detection in hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3558–3571. [CrossRef]
17. Ren, L.; Zhao, L.; Li, X. Anomaly detection-oriented band selection for hyperspectral image. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 3788–3791.
18. Yu, C.; Zhou, S.; Song, M.; Gong, B.; Zhao, E.; Chang, C.I. Unsupervised hyperspectral band selection via hybrid graph convolutional network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5530515. [CrossRef]
19. Li, X.; Liu, Y.; Hua, Z.; Chen, S. An unsupervised band selection method via contrastive learning for hyperspectral images. *Remote Sens.* **2023**, *15*, 5495. [CrossRef]
20. Chang, C.I.; Wang, S. Constrained band selection for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1575–1585. [CrossRef]
21. Ahmad, M.; Haq, D.I.U.; Mushtaq, Q.; Sohaib, M. A new statistical approach for band clustering and band selection using K-means clustering. *Int. J. Eng. Technol.* **2011**, *3*, 606–614.
22. Qian, Y.; Yao, F.; Jia, S. Band selection for hyperspectral imagery using affinity propagation. *IET Comput. Vis.* **2009**, *3*, 213–222. [CrossRef]
23. Wang, Q.; Zhang, F.; Li, X. Hyperspectral band selection via optimal neighborhood reconstruction. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8465–8476. [CrossRef]
24. Andika, F.; Rizkinia, M.; Okuda, M. A hyperspectral anomaly detection algorithm based on morphological profile and attribute filter with band selection and automatic determination of maximum area. *Remote Sens.* **2020**, *12*, 3387. [CrossRef]
25. Zhang, H.; Sun, H.; Sun, X.; Gao, H.; Gao, L.; Zhang, B. Spectral-spatial out-of-distribution-based unsupervised band selection method for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5538915. [CrossRef]
26. Yan, H.F.; Zhao, Y.Q.; Chan, J.C.W.; Kong, S.G. Rapid Hyperspectral Anomaly Detection Using Discriminative Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5533818. [CrossRef]
27. Shang, X.; Song, M.; Wang, Y.; Yu, H. Residual-Driven Band Selection for Hyperspectral Anomaly Detection. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6004805. [CrossRef]
28. Michel, A.; Gross, W.; Schenkel, F.; Middelmann, W. Hyperspectral band selection within a deep reinforcement learning framework. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 52–55.
29. Sun, W.; Du, Q. Hyperspectral band selection: A review. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 118–139. [CrossRef]
30. Liu, Y.; Li, X.; Xu, Z.; Hua, Z. BSFormer: Transformer-based reconstruction network for hyperspectral band selection. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 5507305. [CrossRef]
31. Zhao, J.; Hu, L.; Dong, Y.; Huang, L.; Weng, S.; Zhang, D. A combination method of stacked autoencoder and 3D deep residual network for hyperspectral image classification. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102459. [CrossRef]
32. Sun, H.; Ren, J.; Zhao, H.; Yuen, P.; Tschannerl, J. Novel gumbel-softmax trick enabled concrete autoencoder with entropy constraints for unsupervised hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5506413. [CrossRef]
33. Cai, H.; Chen, T. Multi-dimension CNN for hyperspectral image classificaton. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1275–1278.

34. Ranjan, P.; Kumar, R.; Girdhar, A. Recent CNN advancements for stratification of hyperspectral images. In Proceedings of the 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 3–4 March 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5.
35. Feng, S.; Zhao, Z.; Xi, B.; Zhao, C.; Li, W.; Tao, R.; Li, Y. DSNet: Dynamic Stitchable Neural Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5519214. [CrossRef]
36. Wang, X.; Wei, H.; Feng, S.; Su, N.; Zhao, C.; Li, W.; Tao, R. FCMMA: Fourier Conditional Mask based Mixed Attention Method for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5507012. [CrossRef]
37. Feng, S.; Wang, X.; Feng, R.; Xiong, F.; Zhao, C.; Li, W.; Tao, R. Transformer-based cross-domain few-shot learning for hyperspectral target detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *63*, 5501716. [CrossRef]
38. He, Y.; Tu, B.; Liu, B.; Chen, Y.; Li, J.; Plaza, A. Hybrid multi-scale spatial-spectral transformer for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5527918.
39. Dixit, A.; Gupta, A.K.; Gupta, P.; Srivastava, S.; Garg, A. UNFOLD: 3-D U-Net, 3-D CNN, and 3-D transformer-based hyperspectral image denoising. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5529710. [CrossRef]
40. Roy, S.K.; Das, S.; Song, T.; Chanda, B. DARecNet-BS: Unsupervised dual-attention reconstruction network for hyperspectral band selection. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 2152–2156. [CrossRef]
41. Cai, R.; Yuan, Y.; Lu, X. Hyperspectral band selection with convolutional neural network. In Proceedings of the Pattern Recognition and Computer Vision: First Chinese Conference, PRCV 2018, Guangzhou, China, 23–26 November 2018; Proceedings, Part IV 1; Springer: Berlin/Heidelberg, Germany, 2018; pp. 396–408.
42. Dou, Z.; Gao, K.; Zhang, X.; Wang, H.; Han, L. Band selection of hyperspectral images using attention-based autoencoders. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 147–151. [CrossRef]
43. Cai, Y.; Liu, X.; Cai, Z. BS-Nets: An end-to-end framework for band selection of hyperspectral image. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 1969–1984. [CrossRef]
44. Zhong, J.; Xie, W.; Li, Y.; Lei, J.; Du, Q. Characterization of background-anomaly separability with generative adversarial network for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 6017–6028. [CrossRef]
45. Jiang, T.; Li, Y.; Xie, W.; Du, Q. Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 4666–4679. [CrossRef]
46. Ye, Z.; Cai, W.; Liu, S.; Liu, K.; Wang, M.; Zhou, W. A band selection approach for hyperspectral image based on a modified hybrid rice optimization algorithm. *Symmetry* **2022**, *14*, 1293. [CrossRef]
47. Payra, S.; Sharma, A.; Verma, S. Application of remote sensing to study forest fires. In *Atmospheric Remote Sensing*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 239–260.
48. Liu, Y.; Jiang, K.; Xie, W.; Zhang, J.; Li, Y.; Fang, L. Hyperspectral anomaly detection with self-supervised anomaly prior. *Neural Netw.* **2025**, *187*, 107294. [CrossRef]
49. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
50. Wang, Q.; Li, Q.; Li, X. A fast neighborhood grouping method for hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5028–5039. [CrossRef]
51. Han, X.; Jiang, Z.; Liu, Y.; Zhao, J.; Sun, Q.; Li, Y. A spatial–spectral combination method for hyperspectral band selection. *Remote Sens.* **2022**, *14*, 3217. [CrossRef]
52. Wu, Z.; Paoletti, M.E.; Su, H.; Zheng, P.; Zhang, Z.; Haut, J.M.; Plaza, A. SMCNet: Sparse-inspired Masked Convolutional Network for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5535317. [CrossRef]
53. Liu, S.; Xue, B.; Song, M.; Bao, H.; Zhang, M. Multiobjective Optimization-Based Hyperspectral Unsupervised Band Selection for Anomaly Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *18*, 2813–2828. [CrossRef]
54. Reed, I.S.; Yu, X. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Trans. Acoust. Speech Signal Process.* **1990**, *38*, 1760–1770. [CrossRef]
55. Chang, C.I. An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5131–5153. [CrossRef]
56. Ren, L.; Gao, L.; Wang, M.; Sun, X.; Chanussot, J. HADGSM: A unified nonconvex framework for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5503415. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Remote Sensing Editorial Office
E-mail: remotesensing@mdpi.com
www.mdpi.com/journal/remotesensing



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editors. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editors and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.





Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-9404-8