sensors

Special Issue Reprint

Spectral Detection Technology, Sensors and Instruments, 2nd Edition

Edited by
Qing Yu, Ran Tu, Ting Liu and Lina Li

mdpi.com/journal/sensors

MDPI

Spectral Detection Technology, Sensors and Instruments, 2nd Edition

Spectral Detection Technology, Sensors and Instruments, 2nd Edition

Guest Editors

Qing Yu
Ran Tu
Ting Liu
Lina Li

MDPI

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

Guest Editors

Qing Yu
Fujian Key Laboratory of Green Intelligent Drive and Transmission for Mobile Machinery
Huaqiao University
Xiamen
China

Ran Tu
College of Mechanical Engineering and Automation
Huaqiao University
Xiamen
China

Ting Liu
College of Mechanical Engineering and Automation
Huaqiao University
Xiamen
China

Lina Li
College of Mechanical Engineering and Automation
Huaqiao University
Xiamen
China

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Sensors* (ISSN 1424-8220), freely accessible at: https://www.mdpi.com/journal/sensors/special_issues/8ULU3140KW.

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 978-3-7258-5953-5 (Hbk)
ISBN 978-3-7258-5954-2 (PDF)
https://doi.org/10.3390/books978-3-7258-5954-2

© 2025 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The book 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**

Yirong Lin, Shunlan Liu, Zhonghua Liu, Yuling Fan, Peizhong Liu and Xu Guo
FBStrNet: Automatic Fetal Brain Structure Detection in Early Pregnancy Ultrasound Images
Reprinted from: *Sensors* **2025**, *25*, 5034, https://doi.org/10.3390/s25165034 . . . . . **1**

Guglielmo Ferranti, Chiara Failla, Paolo Finocchiaro, Alessandro Pluchino, Andrea Rapisarda, Salvatore Tudisco and Gianfranco Vecchio
A Multi-Parameter Persistence Algorithm for the Automatic Energy Calibration of Scintillating Radiation Sensors
Reprinted from: *Sensors* **2025**, *25*, 4579, https://doi.org/10.3390/s25154579 . . . . . **17**

Junyi Chen, Qipeng Cai, Xinhai Hu, Qihuai Chen, Tianliang Lin and Haoling Ren
Detection and Tracking of Environmental Sensing System for Construction Machinery Autonomous Operation Application
Reprinted from: *Sensors* **2025**, *25*, 4214, https://doi.org/10.3390/s25134214 . . . . . **33**

Yang Zhao, Lei Zhang, Zhihui Tian, Xiuqing Zhang, Jiandong Bai and Wangbao Yin
Propagation Regimes and Signal Enhancement Mechanisms of Collinear Double-Pulse Plasma with Varying Inter-Pulse Delays
Reprinted from: *Sensors* **2025**, *25*, 3409, https://doi.org/10.3390/s25113409 . . . . . **55**

Ziyu Zheng, Ye Liu, Longxiang Feng, Peizhong Liu, Haisheng Song, Lin Wang and Fang Huang
Multimodal MRI Image Fusion for Early Automatic Staging of Endometrial Cancer
Reprinted from: *Sensors* **2025**, *25*, 2932, https://doi.org/10.3390/s25092932 . . . . . **69**

Tao Wu, Xiao Zhang, Xiao Chen, Wangwang Liu, Yan Han, Yubin Zhong, et al.
Simulation Research and Analysis of Wavelength Modulation Off-Axis Integrated Cavity Output Spectrum Measurement System
Reprinted from: *Sensors* **2025**, *25*, 2478, https://doi.org/10.3390/s25082478 . . . . . **85**

Xueqing Chen, Dacheng Li and Anjing Wang
Research on the Transformer Failure Diagnosis Method Based on Fluorescence Spectroscopy Analysis and SBOA Optimized BPNN
Reprinted from: *Sensors* **2025**, *25*, 2296, https://doi.org/10.3390/s25072296 . . . . . **105**

Yiwei Huang, Xiaoying Zheng, Yongxin Zhu, Tom Trigano, Dima Bykhovsky and Zikang Chen
Deep Learning Based Pile-Up Correction Algorithm for Spectrometric Data Under High-Count-Rate Measurements
Reprinted from: *Sensors* **2025**, *25*, 1464, https://doi.org/10.3390/s25051464 . . . . . **126**

Zirak Khan, Seung-Chul Yoon and Suchendra M. Bhandarkar
Deep Learning Model Compression and Hardware Acceleration for High-Performance Foreign Material Detection on Poultry Meat Using NIR Hyperspectral Imaging
Reprinted from: *Sensors* **2025**, *25*, 970, https://doi.org/10.3390/s25030970 . . . . . **145**

Zhonghua Liu, Jin Wang, Guorong Lyu, Haisheng Song, Weifeng Yu, Peizhong Liu, et al.
Automatic Measurement of Frontomaxillary Facial Angle in Fetal Ultrasound Images Using Deep Learning
Reprinted from: *Sensors* **2025**, *25*, 633, https://doi.org/10.3390/s25030633 . . . . . **167**

Imran Said, Vasit Sagan, Kyle T. Peterson, Haireti Alifu, Abuduwanli Maiwulanjiang, Abby Stylianou, et al.
Seed Protein Content Estimation with Bench-Top Hyperspectral Imaging and Attentive Convolutional Neural Network Models
Reprinted from: *Sensors* **2025**, *25*, 303, https://doi.org/10.3390/s25020303 . . . . . **184**

Rafał Stojek, Anna Pastuszczak, Piotr Wróbel, Magdalena Cwojdzińska, Kacper Sobczak and Rafał Kotyński
High-Resolution Single-Pixel Imaging of Spatially Sparse Objects: Real-Time Imaging in the Near-Infrared and Visible Wavelength Ranges Enhanced with Iterative Processing or Deep Learning
Reprinted from: *Sensors* **2024**, *24*, 8139, https://doi.org/10.3390/s24248139 . . . . . **213**

Mohammad Rahman, Shyh Wei Teng, Manzur Murshed, Manoranjan Paul and David Brennan
BSDR: A Data-Efficient Deep Learning-Based Hyperspectral Band Selection Algorithm Using Discrete Relaxation
Reprinted from: *Sensors* **2024**, *24*, 7771, https://doi.org/10.3390/s24237771 . . . . . **235**

Renan Falcioni, Roney Berti de Oliveira, Marcelo Luiz Chicati, Werner Camargos Antunes, José Alexandre M. Demattê and Marcos Rafael Nanni
Fluorescence and Hyperspectral Sensors for Nondestructive Analysis and Prediction of Biophysical Compounds in the Green and Purple Leaves of *Tradescantia* Plants
Reprinted from: *Sensors* **2024**, *24*, 6490, https://doi.org/10.3390/s24196490 . . . . . **257**

Eduardo Viciana, Juan Antonio Martínez Lao, Emilio López-Lao, Ignacio Fernández and Francisco Manuel Arrabal Campos
Development of an Earth-Field Nuclear Magnetic Resonance Spectrometer: Paving the Way for AI-Enhanced Low-Field Nuclear Magnetic Resonance Technology
Reprinted from: *Sensors* **2024**, *24*, 5537, https://doi.org/10.3390/s24175537 . . . . . **279**

William Yamada, Jerry Cherney, Debbie Cherney, Troy Runge and Matthew Digman
Handheld Near-Infrared Spectroscopy for Undried Forage Quality Estimation
Reprinted from: *Sensors* **2024**, *24*, 5136, https://doi.org/10.3390/s24165136 . . . . . **302**

About the Editors

Qing Yu

Qing Yu, College of Mechanical Engineering and Automation, Huaqiao University. His interests include multispectral detection, machine vision, and precision measurement.

He worked as PI in projects such as: National Natural Science Foundation of China (Young Scientists Program), National Natural Science Foundation of China (General Program), and Fujian Provincial Science and Technology Plan Project, etc.

He has published more than 50 scientific articles, authorized more than 20 patents, and received 4 provincial and ministerial-level science and technology progress awards.

Ran Tu

Ran Tu, College of Mechanical Engineering and Automation, Huaqiao University. His interests include thermal safety and signal detection, energy conversion and monitoring methods in EVs, combustion science and detection technology, and intelligent detection technology.

He worked as PI in projects such as: National Natural Science Foundation of China (Young Scientists Program), National Natural Science Foundation of China (General Program), and Open Research Topics of the State Key Laboratory of SKLFS (USTC), etc.

He has published more than 90 scientific articles, and 1 bestselling popularized science reading, authorized more than 20 patents, received 6 awards from international and domestic authoritative academic organizations.

Ting Liu

Ting Liu, College of Mechanical Engineering and Automation, Huaqiao University. Her research focuses on optical fiber spectroscopic sensing technology, intelligent detection technology, and precision photoelectric detection technology. As Principal Investigator, she has led multiple research projects including the Young Scientists Program of the National Natural Science Foundation of China (NSFC), the NSFC General Program, the Foreign Cooperation Project of Fujian Provincial Department of Science and Technology, and the Open Fund of the State Key Laboratory of Precision Measurement Technology and Instruments (Tsinghua University). To date, she has published over 50 scientific papers and holds 6 authorized patents.

Lina Li

Lina Li, College of Mechanical Engineering and Automation, Huaqiao University. Her research interests lie in spectroscopy, Visible Near Infrared spectral analysis technology, and the development of photoelectric detection instruments. She has been awarded research grants as the Principal Investigator from various sources, including the Young Scientists Program of the National Natural Science Foundation of China and the Fujian Provincial Science and Technology Plan Guiding Project for High-Tech and Industrial Technology. She has authored 15 SCI-indexed international journal articles and holds 5 patents.

sensors

MDPI

Article

FBStrNet: Automatic Fetal Brain Structure Detection in Early Pregnancy Ultrasound Images

Yirong Lin [1,†], Shunlan Liu [2,†], Zhonghua Liu [3,†], Yuling Fan [4], Peizhong Liu [4,*] and Xu Guo [3,*]

[1] School of Medicine, Huaqiao University, Quanzhou 362021, China; linyirong2000@126.com
[2] Department of Ultrasound, The Second Affiliated Hospital of Fujian Medical University, Quanzhou 362000, China; 85443785@fjmu.edu.cn
[3] Department of Ultrasound, Quanzhou First Hospital Affiliated to Fujian Medical University, Quanzhou 362000, China; liuzhonghua2005@126.com
[4] College of Engineering, Huaqiao University, Quanzhou 362021, China; yling_fan@126.com
* Correspondence: pzliu@hqu.edu.cn (P.L.); guoxu_2003@126.com (X.G.)
† These authors have contributed equally to this work.

Abstract

Ultrasound imaging is widely used in early pregnancy to screen for fetal brain anomalies. However, the accuracy of diagnosis can be influenced by various factors, including the sonographer's experience and environmental conditions. To address these limitations, advanced methods are needed to enhance the efficiency and reliability of fetal anomaly screening. In this study, we propose a novel approach based on a Fetal Brain Structures Detection Network (FBStrNet) for identifying key anatomical structures in fetal brain ultrasound images. Specifically, FBStrNet builds on the YOLOv5 baseline model, incorporating a lightweight backbone to reduce model parameters, replacing the loss function, and utilizing a decoupled detection header to improve accuracy. Additionally, our method integrates prior clinical knowledge to minimize false detection rates. Experimental results demonstrate that FBStrNet outperforms state-of-the-art methods, achieving real-time detection of fetal brain anatomical structures with an inference time of just 11.5 ms. This capability enables sonographers to efficiently visualize critical anatomical features, thereby improving diagnostic precision and streamlining clinical workflows.

Keywords: deep learning; early pregnancy; fetal brain; ultrasound imaging; anatomical structure detection

1. Introduction

The first-trimester fetal ultrasound scan is a critical component of obstetric care, aimed at assessing fetal viability and evaluating fetal anatomy for the early detection of structural anomalies [1,2]. Anomalies that should always be detected in a first-trimester fetal ultrasound scan include body stalk anomaly, anencephaly, alobar holoprosencephaly, exomphalos, gastroschisis, and megacystis [3]. The International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) clinical protocol for the first trimester scan defines these key tasks and measurements, carried out between 11^{+0} and 14^{+0} weeks of gestation. According to the ISUOG guidelines, examination of the fetal head and central nervous system is best achieved using a combination of axial and midsagittal planes [2]. According to ISUOG guidelines, we selected four fetal brain ultrasound (FBUS) images to assess fetal brain anatomy, such as the trans-lateral ventricular axial plane (TLVAP), trans-posterior

fossa axial plane (TPFAP), trans-thalamic axial plane (TTAP), and mid-sagittal plane (MSP). The four planes make it possible to clearly observe the calcification of the skull. The fetal brain's various anatomical structures, such as the cranial bone (CB), brain midline (BM), choroid plexus (CP), aqueduct of Sylvius (AS), thalamus and cerebral peduncles (T&P), cisterna magna (CM), brainstem (B), fourth ventricle (FV), nasal bone (NB), mandible (M), third ventricle (TV), hard palate (HP), and other important areas, can be seen. This enables the doctor to conduct a thorough anatomical and physiological evaluation of the fetal brain.

Ultrasound imaging is a widely utilized tool for screening fetal brain anomalies during early pregnancy [4]. However, several unavoidable factors can influence the accuracy of a sonographer's diagnosis. First, technical limitations such as the inherent inaccuracy of ultrasound imaging, low signal-to-noise ratios, and the presence of speckle noise often compromise the clarity of the images. Second, the diagnostic outcome is heavily dependent on the operator's skill, including their proficiency in ultrasound operation, adherence to standardized screening protocols, and ability to accurately identify fetal brain anatomical structures [5]. Additionally, external factors such as the equipment model, the small size of fetal organs in early pregnancy, and the fetal position during the examination can further impact the completeness and accuracy of the anatomical visualization [6]. Given these challenges, there is a pressing need to develop more effective methods to enhance the screening and diagnostic efficiency of sonographers, thereby improving the detection of fetal malformations and contributing to the goals of eugenics.

Deep learning developments in recent years have yielded innovative solutions for FBUS analysis. Although convolutional neural networks (CNNs), such as YOLOv3 [7] and YOLOv5 [8], have demonstrated increased accuracy in adult ultrasound tasks, ultrasound-specific noise and the small size of anatomical targets during early gestation continue to restrict their performance in FBUS. Transformer-based architectures employ self-attention mechanisms to introduce global context modeling, but their quadratic computational complexity renders them difficult to apply in real-world scenarios. Lightweight designs of domain-specific adaptations like MobileNetV2 [9] and EfficientNetV2 [10] help to alleviate these problems to some degree, but they still struggle to strike a balance between bounding box precision and classification accuracy when facing FBUS constraints. Our proposal is the Fetal Brain Structures Detection Network (FBStrNet), a specialized framework for accurate, real-time key anatomical structures (KASs) localization in early pregnancy FBUS to overcome this technical bottleneck. In summary, the key contributions of this paper are as follows:

(1) This study addresses the challenge of automatic detection of KASs in FBUS images through FBStrNet, an innovative multi-task deep learning framework.
(2) By integrating YOLOv5 with a feature fusion strategy, FBStrNet achieves unprecedented precision in detecting 12 critical anatomical structures, significantly improving both detection accuracy and computational efficiency.
(3) Experimental results demonstrate the model's exceptional robustness under low signal-to-noise ratio conditions, making it a valuable assistive tool for sonographers to identify key structures in real-time, thereby enhancing the efficiency and accuracy of fetal brain development assessments.

2. Related Work

With the progress of deep learning, the application of these methods for medical image analysis has increased. Several clinical issues have been solved, improving the accuracy of clinical diagnoses and reducing the burden on sonographers. Among the most advanced and wide applications is the recognition of medical images.

Yaqub et al. [5] employed deep learning techniques to achieve precise fetal brain localization, enabling detection of the cavum septi pellucidi (CSP) and accurate identification of the trans-ventricular axial plane (TVAP). Their framework facilitates comprehensive fetal brain symmetry assessment, providing clinically actionable insights for neurodevelopmental evaluation. Lin et al. [11] designed a Mask Focal R-CNN (MF R-CNN) to detect six anatomical structures: lateral sulcus, thalamus, CP, CSP, TV, and BM. By integrating image magnification with clinical protocols, their method enhances TTAP, standardization, and ultrasound image quality assessment. Zhang et al. [12] proposed a multi-task model for automated quality assessment of fetal head, abdomen, and cardiac ultrasound images. By detecting anatomical structures to determine compliance with clinical standards, their approach achieves state-of-the-art performance in both segmentation accuracy and diagnostic reliability. Pu et al. [13] developed a hybrid CNN-RNN architecture for fetal standard plane recognition in industrial IoT environments. The CNN identifies anatomical structures and classifies four planes (TTAP, Trans-cerebellar Axial Plane (TCAP), abdomen, and lumbosacral vertebrae plane), while the RNN captures temporal dependencies for organ tracking across video frames. Coronado-Gutierrez et al. [14] implemented a deep learning system to identify TTAP, TVAP, and TCAP planes. Their framework automatically segments nine anatomical structures, including CSP, anterior horn superior/inferior, thalamus superior/inferior, Sylvian fissure, interhemispheric fissure, and cerebellum, while performing automated biometric measurements. Guo et al. [15] introduced a coarse-to-fine detection framework combined with multi-task learning to optimize ultrasound standard plane acquisition. Their feature fusion strategy significantly improves detection efficiency and accuracy across multi-scale images. Han et al. [16] created CUPID, an automated software for standard plane recognition (TCAP, TTAP, TVAP) and fetal biometric measurement. This tool enhances clinical workflows through efficient detection of anatomical structures and quantitative assessment of fetal brain development.

Automatic recognition and detection of FBUS images have been the focus of extensive research in recent years, with most studies concentrating on fetal ultrasound images from 18 to 40 weeks of gestation. In contrast, first-trimester fetal ultrasound images have received limited attention. Existing research in this area has primarily focused on identifying fetal images with normal nuchal translucency [17], evaluating fetal health [18], and automating the segmentation of video clips containing anatomical structures [1]. However, there is a significant gap in research on the automatic recognition of key early-pregnancy ultrasound planes, including the TLVAP, TPFAP, TTAP, and MSP, as well as the critical anatomical structures visible in these planes. The model developed in this study aims to address this gap by assisting diagnostic sonographers in rapidly acquiring FBUS images for use as a teaching tool and for ultrasound didactic training. Additionally, it will aid in the identification of important anatomical structures within FBUS images.

3. Methodology

3.1. Model Framework

In this study, we established YOLOv5 as the baseline architecture and developed FBStrNet, a novel deep learning framework specifically designed for detecting twelve critical anatomical structures in FBUS images. As depicted in Figure 1, FBStrNet is structured around three synergistic components: (1) a lightweight backbone network optimized for efficient feature extraction, (2) a hierarchical interaction module designed to integrate multi-scale contextual information, and (3) a decoupled prediction head tailored for precise localization and classification of anatomical structures. Owing to their critical roles in model efficacy, we now provide an in-depth analysis of the backbone network and the decoupled prediction head, elucidating their architectural innovations and functional contributions.

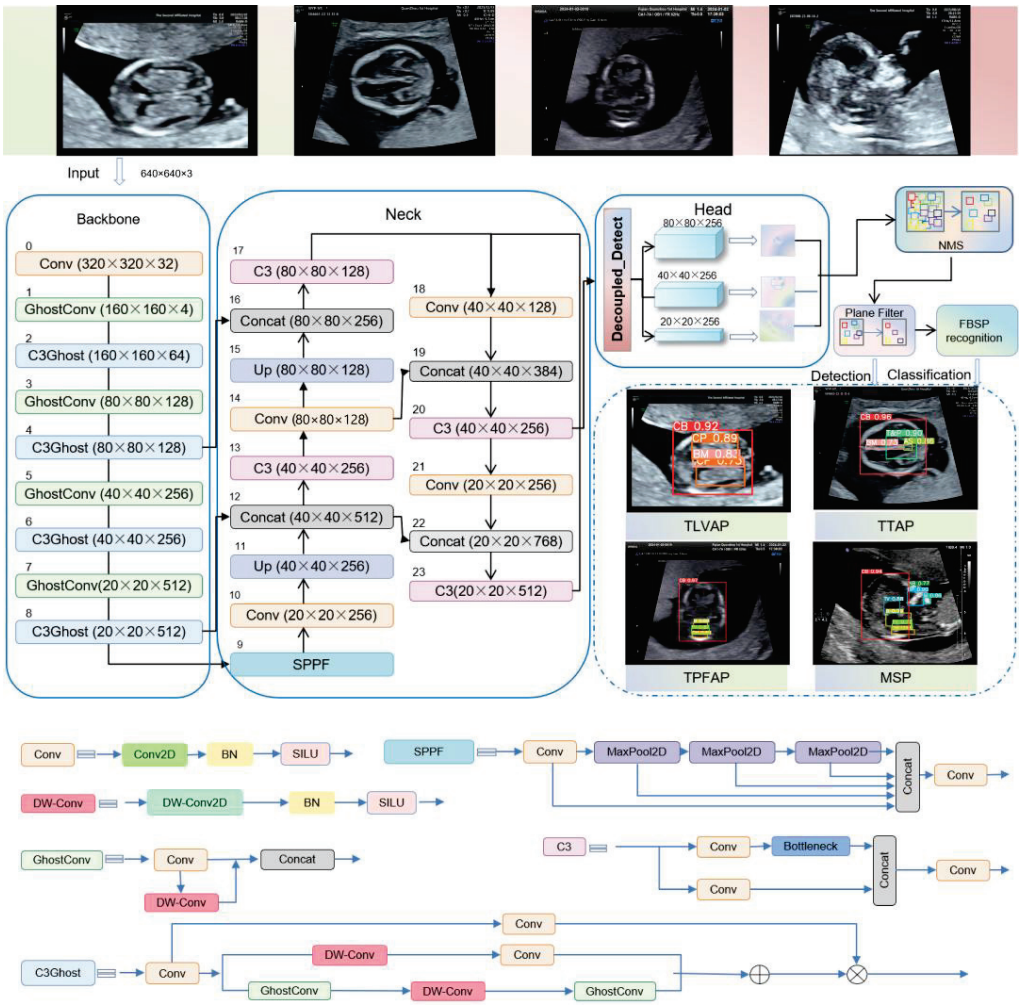


Figure 1. The structure of the FBStrNet model.

To balance model efficiency and detection accuracy, we focused on optimizing the backbone architecture. Specifically, we integrated the Ghost module [19] into its backbone architecture. In the Ghost module, the input (I) is processed through a 1×1 convolution to produce many intrinsic features (X) and reduce the feature channel to half its original number. Then, each intrinsic feature is convolved by a deep convolution of the Cheap Linear Operations (CL) to produce a ghost feature (Y). The concatenation (Cat) of intrinsic and ghost features yields the final feature map, which is the output (O). Figure 2 demonstrates the Ghost module's mechanism of function. Through a decrease in convolution kernel count and an increase in linear operations, the Ghost module effectively decreases network complexity. We optimized the Conv and C3 modules as the GhostConv and C3Ghost modules, respectively, and introduced the Ghost module to YOLOv5. The deep convolution is employed in the GhostConv module to create extra features through low-cost linear operations. Figure 1 describes the structure of the GhostConv module and the C3Ghost module in FBStrNet. Replaced the C3 module in Layers 1, 3, 5, and 7 with the C3Ghost module, and the Conv module in Layers 2, 4, 6, and 8 with the GhostConv module; the stride size of Conv2D is 2.

After optimizing the backbone architecture, we turned our attention to the detection head, which plays a vital role in accurately identifying and locating the target anatomical structures. To enhance the detection performance, we introduced Decoupled_Detect. It serves as a decoupling head [20], extracting the target location and category informa-

tion, learning these aspects separately through distinct network branches, and ultimately fusing them. Specifically, each layer of FPN features initially employs a 1×1 convolutional layer to diminish the number of feature channels to 256. Subsequently, it integrates two parallel branches, each of which consists of two stacked 3×3 convolutional layers and is dedicated to classification and regression tasks, respectively. An IoU branch is appended to the regression branch. In the regression branch, four parameters are predicted to define the detection box: the x-coordinate (x) and y-coordinate (y) of the center point of the detection box, as well as its width (w) and height (h). The decoupling head mechanism operates as shown in Figure 3.

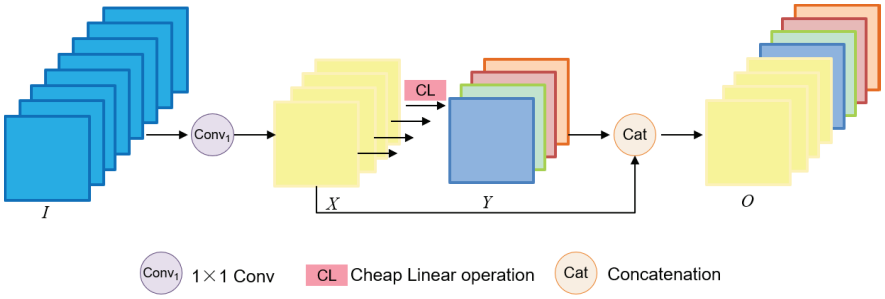


Figure 2. Ghost module.

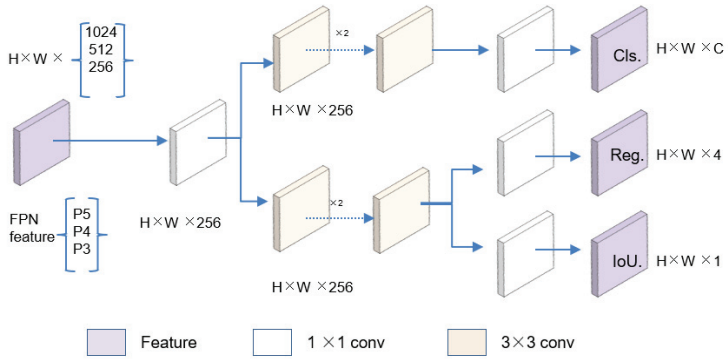


Figure 3. The module of the Decoupled_Detect.

3.2. Loss Function

A critical factor in evaluating the model's performance is its loss function. While YOLOv5 employs the CIoU loss function, CIoU fails to address the balance between positive and negative samples [21]. To overcome this limitation, Gevorgyan proposed the SIoU loss function, which redefines the penalty metric by incorporating the vector angles between adaptive weights and target regressions [22]. By leveraging the vector angle between regressions and adaptive weights, SIoU improves the accuracy of prediction boxes and enhances the overall detection performance of the system. This approach reduces the risk of overfitting or underfitting by effectively balancing positive and negative samples. The SIoU loss function primarily consists of four components: angular loss, shape loss, distance loss, and IoU loss, and is defined as follows:

$$L_{SIoU} = 1 - IoU + (\Delta + \Omega)/2 \tag{1}$$

where the shape loss and the distance loss are represented by Ω and Δ, respectively, SIoU incorporates the consideration of angular loss into the distance loss, as evidenced by the distance loss formula in Equation (2). The precise formulas for ρ_x, ρ_y, and γ are presented in Equation (3).

$$\Delta = \sum_{t=x,y} \left(1 - e^{-\gamma \rho_t}\right) \tag{2}$$

$$\rho_x = \left((b_{c_x}^{gt} - b_{c_x})/c_w \right)^2, \rho_y = \left((b_{c_y}^{gt} - b_{c_y})/c_h \right)^2, \gamma = 2 - \Lambda \tag{3}$$

where c_w and c_h are the width and height of the tiniest external rectangles that enclose both the ground truth and predicted boxes, respectively. Λ represents the angle loss, which is calculated as shown in Equation (4).

$$\Lambda = 1 - 2 * \sin^2(\arcsin(c_h/\sigma) - \pi/4) \tag{4}$$

$$\sigma = \sqrt{\left(b_{c_x}^{gt} - b_{c_x} \right)^2 + \left(b_{c_y}^{gt} - b_{c_y} \right)^2} \tag{5}$$

$$c_h = \max\left(b_{c_y}^{gt}, b_{c_y} \right) - \min\left(b_{c_y}^{gt}, b_{c_y} \right) \tag{6}$$

where σ symbolizes the gap, as determined by the given Equation (5), between the central point of the authentic ground box and the anticipated box. The height difference between the predicted box's center and the ground truth box's center is denoted by c_h, calculated as shown in Equation (6). The coordinates of the ground truth box's center point (x, y) are $b_{c_x}^{gt}$ and $b_{c_y}^{gt}$. The coordinates of the center point of the prediction box's center point (x, y) are represented by the variables b_{c_x} and b_{c_y}. The shape loss is calculated using Equation (7). The precise formulas for ω_w and ω_h are presented in Equations (8) and (9), respectively.

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta \tag{7}$$

$$\omega_w = |w - w^{gt}|/\max(w, w^{gt}) \tag{8}$$

$$\omega_h = |h - h^{gt}|/\max(h, h^{gt}) \tag{9}$$

where the prediction box's height and width are indicated by h and w, respectively. The ground truth box's dimensions are represented by w^{gt} and h^{gt}. The degree of concern for form shape loss is represented by θ.

In the SIoU loss function, the mismatch penalty metric is evaluated between the predicted and ground truth bounding boxes, a task traditionally handled by conventional loss functions. While traditional approaches primarily focus on shape, IoU, and distance, SIoU introduces an additional consideration: the alignment between the ground truth box and the predicted box's orientation. By incorporating direction as a key factor, SIoU reduces the overall degrees of freedom during optimization, enabling the predicted box to converge more rapidly toward the ground truth. This enhancement significantly improves both the inference speed and accuracy of the detection model in identifying anatomical structures of the fetal brain.

3.3. Target Filter

In contrast to natural image processing, a significant amount of prior medical information can be utilized in processing medical images [8]. This is especially true for object detection of anatomical structures in medical imaging, which requires analyzing each KAS in the image. In our fetal brain detection task, there are four standard planes and twelve anatomical structures, and these four planes are TLVAP, TTAP, TPFAP, and MSP. There are four anatomical structures in the TLVAP, which are CB, BM, and tow CPs; the TTAP contains four anatomical structures: CB, BM, AS, and T&P. There are four anatomical structures in the TPFAP: CB, CM, B, and FV. The MSP contains eight anatomical structures: CB, CM, B, FV, NB, M, TV, and HP. Except for the standard plane of the MSP, all other standard planes have four anatomical structures.

Algorithm 1 describes the filtering and sorting of anatomical structure detection results, removing duplicate and erroneous detection results and retaining the results with high

confidence for each class of anatomical structures. Res_I = [Obj_1, Obj_2, Obj_3. . ., Obj_n], Obj_n = (x1, y1, x2, y2, conf, cls); the upper-left coordinate points of the object bounding boxes are indicated by x1, y1, and the lower-right coordinate points by x2, y2; conf refers to the object's confidence in being detected, and cls stands for the object's category. P_Objs is a dictionary that contains the target categories that are expected to be present in each standard plane. P_Objs = {TLVAP: [CB, BM, CP], TTAP: [CB, BM, AS, T&P], TPFAP: [CB, CM, B, FV], MSP: [CB, CM, B, FV, NB, M, TV, HP]}. Filtered_Res is derived from P_Objs to eliminate erroneous detection results, with an initial value of None. Best_Conf_Res retains the results from Filtered_Res that have a confidence level greater than 0.5 and selects the best result for each class, with an initial value of None. The prediction result for cp is referred to as Res_cp, and its initial value is None. Res_max represents the result with the highest confidence for each class, with an initial value of None. Res_now is the concatenation of Best_Conf_Res and Top_cp, with duplicates removed based on all columns, and its initial value is None. Top_cp is the best result selected from the top two records of Res_cp.

The model utilizes non-maximum suppression (NMS) to further refine object detection within the YOLO series. However, due to the complexity and variability of features in FBUS images, certain regions of the ultrasound image may exhibit characteristics resembling anatomical structures. This often leads to misidentification and the detection of redundant objects. To address this issue, we introduce a post-processing step based on medical prior knowledge, as outlined in Algorithm 1. This step involves filtering, reorganizing, and sorting the detection results to eliminate false positives and duplicate detections, ultimately retaining only the results with the highest confidence scores.

Algorithm 1. An algorithm for filtering and sorting object detection results.

Require: Res, P_Objs
Ensure: Filtered_Res ← None, Best_Conf_Res ← None, Res_cp ← None, Top_cp ← None, Res_now ← None.
Step 1: Filter_results based on P_Objs
for result in Res **do**
  if result matches criteria defined by P_Objs **then**
    Filtered_Res ← Filtered_Res ∪ {result}
  end if
end for
Step 2: Retain results with conf > 0.5 and best per class
for each result in Filtered_Res **do**
  if result.conf > 0.5 **then**
    Best_Class_Res[result.class] ← Update with result ▷ Keep the best result per class
  end if
end for
Step 3: Retain top 2 results with the highest confidence for TLVAP with 2 CPs
if "CP" in Res **then**
  Res_cp ← FilterResultsByNameCP(Res)
  Top_cp ← GetTopTwoByConfidence(Res_cp)
end if
Step 4: Sort, screen and rank the results
Res now ← ConcatenateAndRemoveDuplicates(Best_Class_Ress, Top_cp)
Res_now ← SortByAscending(Res_now, 'class')
return Res_now

4. Experiments and Results

4.1. Experimental Setup

Four standard ultrasound planes of the fetal brain were systematically collected during early pregnancy, with anatomical structures annotated for each plane as demonstrated in Figure 4 The dataset was developed in collaboration with clinical ultrasound specialists at two tertiary medical centers: the Quanzhou First Hospital and the Second Affiliated Hospital of Fujian Medical University. Following established clinical guidelines, ultrasound examinations were performed using high-frequency linear array probes (4–7 MHz) across eight advanced Doppler systems (IU22, IUELET, EPIQ5, EPIQ7 (Philips Healthcare, Best, North Brabant, The Netherlands); UGEO WS80A (Samsung, Medison Co., Ltd., Seoul, Republic of Korea); VolusonE6, VolusonE8, and VolusonE10 (General Electric Healthcare, Chicago, IL, USA). A total of 4086 FBUS images were acquired, capturing 1–3 distinct images per fetal plane to ensure anatomical variation representation. Among these images, the KASs were annotated in 2531 images by a professional sonographer, which were used for model training. The remaining 1573 unannotated images were set aside for testing the model's performance, thereby enabling an evaluation of its generalization ability when encountering unseen data. The final dataset composition for fetal brain structure analysis is detailed in Table 1, showing the distribution between training and test sets.

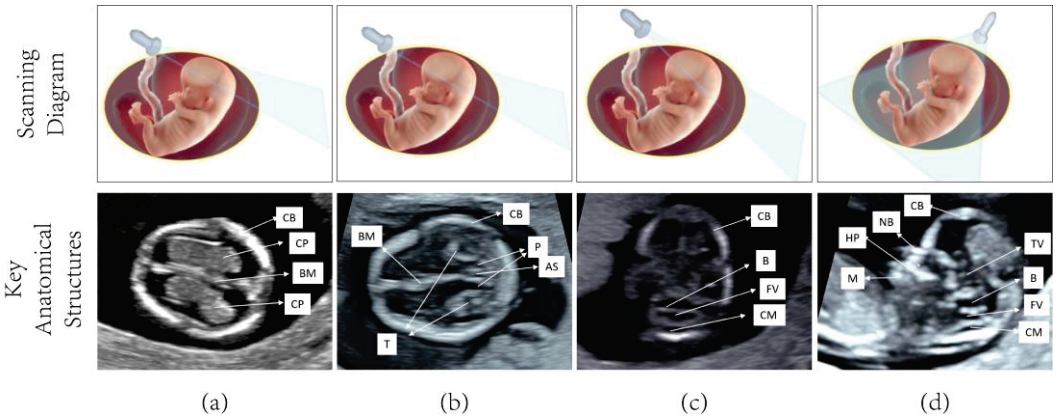


Figure 4. In these four subplots, the first row displays the clinical scanning diagram of the plane, while the second row showcases the ultrasound image of the plane along with its anatomical structures. (**a**) TLVAP; (**b**) TTAP; (**c**) TPFAP; (**d**) MSP.

Table 1. The number of anatomical structures contained in the training set and the test set.

	Train	Test
CB	2518	1866
BM	1251	1010
CP	1266	938
CM	1267	856
B	1267	856
FV	1267	856
AS	618	539
T&P	618	539
NB	649	317
M	649	317
TV	649	347
HP	649	347

The experiment was conducted using the following hardware configuration: an NVIDIA GeForce RTX 3050 GPU equipped with 16 GB of RAM (NVIDIA Corporation, Santa Clara, CA, USA), a 12th Generation Intel® Core™ i5-12400F CPU (Intel Corporation, Santa Clara, CA, USA). The operating system was Windows 10. The programming environment utilized was Python 3.8.18, and the deep learning framework employed for the experiment was PyTorch 2.1. The optimizer, learning rate, and other test parameters were configured based on YOLOv5. Specifically, the learning rate was set to 0.001, and the weight decay was also set to 0.001. The Intersection over Union threshold was configured at 0.41, while the anchor threshold was set to 5.0. The model was trained for 300 epochs, with an input image size of 640 × 640 pixels. The Stochastic Gradient Descent optimizer was used for training. These parameters were carefully selected to optimize the model's performance.

4.2. Evaluation Metrics

The detection performance was evaluated using Precision, Recall, mAP@0.5, and mAP@0.5:0.9, as specified in the Formulas (10), (11), (12), and (13), respectively. In Equations (10) and (11): TP: The model correctly detects fetal brain anatomical structures. FP: The model falsely detects non-existent structures. TN: The model correctly excludes non-target regions. FN: The model misses actual anatomical structures. These symbols are used to compute the metrics mentioned above. The area under the P–R curve is called the Average Precision, or AP for short. A higher AP value indicates better model performance. The mean Average Precision, or mAP for short, is the average value of AP for each category. One of the most crucial evaluation metrics for object detection tasks is the mAP, whose values range from 0 to 1. The higher the mAP, the better the model's performance. When the IoU is set to 0.5, mAP@0.5 is used to calculate mAP; when the IoU threshold is between 0.5 and 0.95 (step size of 0.05), mAP@0.5:0.95 is used to calculate mAP.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{10}$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{11}$$

$$\text{AP} = \int_0^1 \text{P}(\text{r})\text{dr} \tag{12}$$

$$\text{mAP} = \sum_{\text{i}=1}^{\text{n}} \text{AP}_\text{i}/\text{n} \tag{13}$$

For evaluating our model's complexity, we employed the number of parameters as a typical metric used in deep learning models. Nevertheless, given the typically vast number of parameters in such models, we typically resort to megabytes (MB) as the unit of quantification. Turning our attention to the inference speed, we aimed to minimize the potential for accidental errors by selecting a subset of 600 FBUS images from the test set and permitting the model to perform inference on the GPU. Following this, we determined the inference speed for each image by averaging the time taken by the model to process each of the 600 images. Additionally, to ensure that our results more closely reflect the model's performance in real-world scenarios, we configured the batch size to 1 when utilizing the target detection models for inference on FBUS images.

4.3. Experiment Results

4.3.1. Detection Results

During the object detection task, we compare our model with the currently better-performing object detection models, such as YOLOv3 [23], YOLOv5 [24], YOLOv7 [25],

YOLOv9 [26], Cascade R-CNN [27], QueryInst [28], and Faster R-CNN [29]. The performance metrics presented in Table 2 demonstrate that our FBStrNet model achieves superior performance across all metrics for the FBUS anatomical structure detection task. Among the compared models, YOLOv9 stands out as the top performer for this task. Our model exceeds YOLOv9 by a marginal 0.3% improvement in precision, while simultaneously exhibiting significantly fewer parameters and delivering faster inference speeds compared to YOLOv9. In Figure 5, it presents the results obtained by several well-performing models for FBUS image detection. The portrayed results clearly illustrate that FBStrNet excels at accurately and effectively identifying each KAS within an FBUS image. To further assess the recognition capabilities of FBStrNet across various anatomical structures, we have plotted the P–R curves on the test set, as illustrated in Figure 6. Furthermore, it is worth noting that our proposed approach enables rapid identification and detection of FBUS images within a mere 11.5 ms, making it highly suitable for real-time clinical scanning requirements during fetal brain examinations, where sonographers can conveniently visualize anatomical structures in ultrasound images.

Table 2. The results of each model's identification of the FBUS's anatomical structure.

Model	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Parameters (M)	Times (ms)
YOLOv3	98.6%	98.6%	98.3%	66.4%	61.5	30.9
YOLOv5-n	98.7%	98.5%	98.3%	65.8%	1.8	8.4
YOLOv5-s	98.2%	97.4%	98.3%	64.2%	7.0	11.2
YOLOv7	94.8%	95.1%	96.4%	56.5%	36.5	65.4
YOLOv7-x	95.5%	96.6%	96.8%	58.3%	70.9	53.1
YOLOv9-c	98.3%	98.5%	98.8%	69.0%	60.8	58.5
YOLOv9-e	98.2%	98.3%	98.5%	69.3%	69.5	62.4
Faster R-CNN	-	-	91.4%	57.8%	41.4	56.1
Cascade R-CNN	-	-	92.0%	59.8%	69.4	74.5
QueryInst	-	-	96.1%	65.9%	172.5	121.0
FBStrNet (Ours)	98.6% (+0.3%)	98.5%	98.7%	64.3%	6.9	11.5

4.3.2. Model Visualization Results

Grad-CAM (Gradient-weighted Class Activation Mapping) [30] has emerged as a robust visualization technique for interpreting deep learning models in classification tasks. This method generates class activation maps by computing gradients from the final convolutional layer, enabling the identification of critical image regions that significantly influence the model's classification outcomes. By highlighting class-specific discriminative features, Grad-CAM provides valuable insights into the model's decision-making process. In the current study, we implemented Grad-CAM to precisely localize and visualize regions of interest within fetal ultrasound images, as demonstrated in Figure 7.

4.4. Ablation Experiments

To evaluate the effectiveness of our proposed model in identifying fetal brains, we performed ablation studies targeting Ghost, the SIoU loss function, and Decoupled_Detect. The hyperparameter settings were kept constant during the ablation experiments performed. The results of these experiments are summarized in Table 3, which highlights the benefits of various model-structure optimization techniques. Notably, FBStrNet achieved a 0.4% boost in mAP@0.5 while maintaining the same number of parameters compared to the baseline YOLOv5 model.

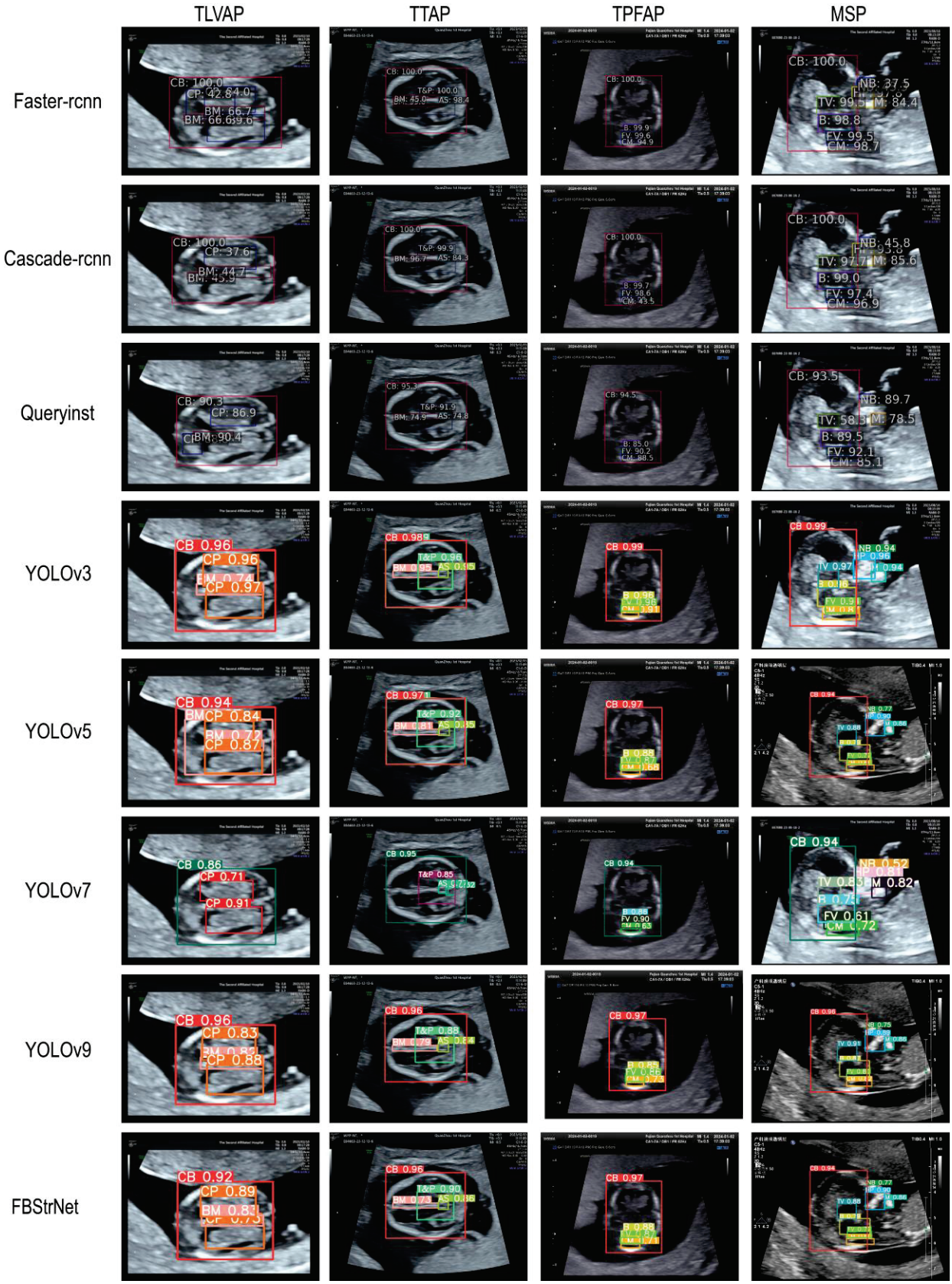


Figure 5. The visualization of several models that perform well in the detection task of FBUS images, as well as the detection results of our model FBStrNet.

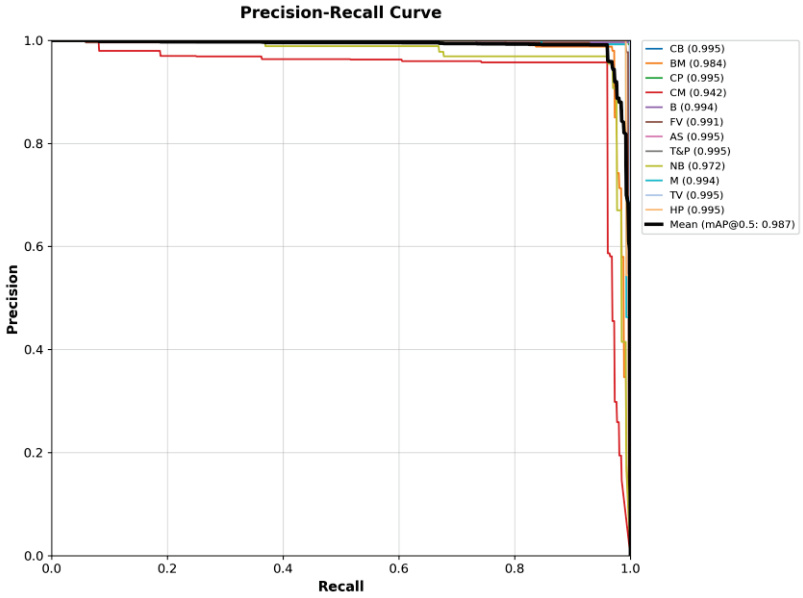


Figure 6. FBStrNet's P–R curve.

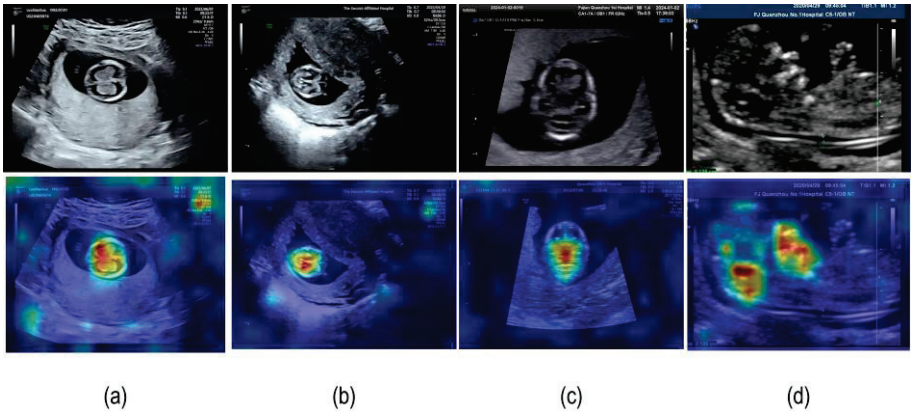


Figure 7. The Grad-CAM visualization results of FBStrNet. (**a**) TLVAP; (**b**) TTAP; (**c**) TPFAP; (**d**) MSP.

Table 3. The results of ablation tests conducted using Ghost, SIoU, and Decoupled_Detect in the FBStrNet.

Ghost	SIoU	Decoupled_Detect	mAP@0.5	mAP@0.5:0.95	Parameters (M)	Time (ms)
Baseline(YOLOv5-s, CIoU)	×	×	98.3%	64.2%	7.0	9.8
√	×	×	98.4%	64.6%	5.1	10.0
×	√	×	98.9%	65.6%	7.0	9.8
×	×	√	98.8%	66.4%	8.8	11.6
√	√	×	98.5%	63.1%	5.1	10.0
√	×	√	98.2%	64.4%	6.9	11.3
√	√	√	98.7%	64.3%	6.8	11.5

First, we test the performance of each module on a baseline basis. The Ghost module reduces the number of parameters to 5.1 M, slightly increases the inference time to 10.0 ms, and marginally improves mAP@0.5 and mAP@0.5:0.95 to 98.4% and 64.6%, respectively. The SIoU loss function significantly enhances mAP @0.5 and mAP@0.5:0.95 to 98.9% and 65.6% while maintaining the same number of parameters and inference time. Although the Decoupled_Detect module increases the number of parameters to 8.8 M and extends the inference time to 11.6 ms, it also significantly improves mAP@0.5 and mAP@0.5:0.95 to

98.8% and 66.4%, respectively. Next, building upon the baseline model, we introduced the Ghost and SIoU modules, which reduced the parameter count to 5.1 M while maintaining an inference time of 10.0 ms. While this configuration resulted in a slight improvement in mAP@0.5, it led to a decline in mAP@0.5:0.95. To further enhance performance, we integrated the Ghost, SIoU, and Decoupled_Detect modules into the baseline model. This combination increased the parameter count to 6.8 M and extended the inference time to 11.5 ms, but it achieved notable improvements in both mAP@0.5 and mAP@0.5:0.95 metrics.

The SIoU and Decoupled modules demonstrate remarkable effects in enhancing accuracy, particularly for the mAP@0.5:0.95 metric. The Ghost module effectively reduces the number of parameters with minimal impact on inference time. The combination of Ghost, SIoU, and Decoupled achieves a good balance between accuracy and parameter count.

To evaluate the impact of different loss functions on the performance of FBStrNet, we conducted experiments using both CIoU and SIoU as the loss functions. The results are presented in Table 3. When compared with the CIoU loss function, the SIoU loss function brings about a significant improvement in mAP@0.5, with an increase of 0.5%. However, there is a minor decrease of 0.1% in mAP@0.5:0.95. In terms of model complexity, the number of parameters reduces by 0.1 M when using the SIoU loss function. On the other hand, the inference time increases by 0.3 ms.

5. Discussion

This study proposes the FBStrNet model, which can effectively detect anatomical structures in fetal brain ultrasound images during early pregnancy. Experimental results demonstrate that it significantly outperforms existing methods in detection accuracy. FBStrNet builds upon YOLOv5 as the baseline model, incorporating a series of improvements tailored for detecting key anatomies in fetal brain ultrasound images. The selection of YOLOv5 as the baseline model is justified by several advantages: (1) YOLOv5 has a smaller number of parameters compared to other models (as shown in Table 2), making it easier to deploy and run on resource-constrained devices, such as hospital equipment; (2) YOLOv5 offers faster inference speeds, which is critical for real-time clinical scanning; and (3) YOLOv5 provides high flexibility, allowing us to adapt it to our specific data characteristics and requirements.

The filtering logic of Algorithm 1 strictly relies on the predefined target categories in P_Objs, which effectively reduces false positives. Still, it simultaneously limits its ability to detect new or undefined categories, leading to false negatives, especially when encountering unlisted anatomical structures such as rare pathological variants. Additionally, the accuracy of P_Objs directly impacts the algorithm's performance; if the definitions are incomplete, the filtering effectiveness is significantly compromised. The use of a fixed confidence threshold (e.g., 0.5) to filter results lacks flexibility, making it difficult to adapt to differences in categories and image quality. This approach can easily result in the erroneous filtering of true targets in low-quality images, while in high-quality images, artifacts or noise may cause false positives. Therefore, it is necessary to introduce a dynamic category expansion mechanism, integrating unsupervised learning or anomaly detection to identify new categories, and adopt a dynamic threshold mechanism that automatically adjusts the confidence threshold based on image quality and category characteristics to enhance the algorithm's adaptability.

Previous studies have primarily focused on the research of fetal brain ultrasound images obtained during the second or third trimester of pregnancy [11–16,31–36]. In these ultrasound images, anatomical structures form a distinct contrast with the surrounding tissues, facilitating clear observation and detection. In contrast, this study shifts its focus

to ultrasound images of the fetal brain in the first trimester, particularly TLVAP, TTAP, TPFAP, and MSP. As shown in Figure 6, compared with other structures, FBStrNet has a lower mAP@0.5 in identifying CM and NB. The reason is that these two anatomical structures have lower contrast with their surrounding tissues, which makes identification difficult. Research on the fetal brain in the first trimester has mainly concentrated on measuring the crown-rump length and nuchal translucency within the MSP plane, which are macroscopic indicators [1,14,17,18]. However, FBStrNet exhibits unique potential, as it can comprehensively detect anatomical structures across multiple planes. To fully leverage this capability, future work should prioritize the adoption of image processing techniques that can enhance the clarity of ultrasound images, such as noise reduction and contrast adjustment, thereby improving the model's ability to distinguish minute structures.

6. Conclusions

This study introduces FBStrNet, a model for real-time, accurate detection of key structures in fetal brain ultrasound images, aiding clinicians in rapid screening for cranial abnormalities. Results show FBStrNet effectively locates critical structures, improving detection accuracy and efficiency. It serves as a valuable tool for both clinical diagnosis and medical education. However, its performance is limited by dataset size and diversity. Future work will involve collaborating with hospitals to expand datasets and explore image quality assessment methods to enhance robustness. In summary, FBStrNet offers a practical solution for fetal brain analysis, with significant clinical potential and a foundation for further research.

Author Contributions: Conceptualization, X.G.; data curation, X.G.; formal analysis, P.L.; methodology, Y.L.; resources, S.L., Z.L. and P.L.; software, Y.L.; supervision, S.L., Z.L. and P.L.; validation, X.G.; writing—original draft, Y.L.; writing—review and editing, Y.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Joint Funds for the Innovation of Science and Technology, Fujian Province, under Grant Numbers 2024Y9392, 2024Y9386, and 2024Y9435. Additionally, it received support from the Quanzhou Scientific and Technological Planning Projects under Grant Number 2023NS055.

Data Availability Statement: The data are not accessible to the public owing to patient privacy concerns. If you have any research needs, you can contact the corresponding author for assistance.

Conflicts of Interest: We declare that we have no conflicts of interest.

References

1. Yasrab, R.; Fu, Z.; Zhao, H.; Lee, L.H.; Sharma, H.; Drukker, L.; Papageorgiou, A.T.; Noble, J.A. A Machine Learning Method for Automated Description and Workflow Analysis of First Trimester Ultrasound Scans. *IEEE Trans. Med. Imaging* **2023**, *42*, 1301–1313. [CrossRef]
2. Bilardo, C.M.; Chaoui, R.; Hyett, J.A.; Kagan, K.O.; Karim, J.N.; Papageorghiou, A.T.; Poon, L.C.; Salomon, L.J.; Syngelaki, A.; Nicolaides, K.H. ISUOG Practice Guidelines (updated): Performance of 11–14-week ultrasound scan. *Ultrasound Obstet. Gynecol.* **2023**, *61*, 127–143. [CrossRef]
3. Esteves, K.M.; Tugarinov, N.; Lechmann, G.; Abi Habib, P.; Cagliyan, E.; Goetzinger, K.R.; Turan, O.M.; Turan, S. The value of detailed first-trimester ultrasound in the era of noninvasive prenatal testing. *Am. J. Obstet. Gynecol.* **2023**, *229*, 326.e1–326.e6. [CrossRef]
4. Liao, M.; Wang, L.; Shang, N.; Hu, X.; He, B.; Liu, X.; Xiang, G.; Zhong, W. Ultrasound measurements of fetal facial profile markers and their associations with congenital malformations during early pregnancy. *BMC Pregnancy Childb.* **2023**, *23*, 772. [CrossRef]
5. Yaqub, M.; Kelly, B.; Papageorghiou, A.T.; Noble, J.A. A Deep Learning Solution for Automatic Fetal Neurosonographic Diagnostic Plane Verification Using Clinical Standard Constraints. *Ultrasound Med. Biol.* **2017**, *43*, 2925–2933. [CrossRef]

6. Xie, B.; Lei, T.; Wang, N.; Cai, H.; Xian, J.; He, M.; Zhang, L.; Xie, H. Computer-aided diagnosis for fetal brain ultrasound images using deep convolutional neural networks. *Int. J. Comput. Assist. Radiol.* **2020**, *15*, 1303–1312. [CrossRef]
7. Yang, T.; Yuan, L.; Li, P.; Liu, P. Real-Time Automatic Assisted Detection of Uterine Fibroid in Ultrasound Images Using a Deep Learning Detector. *Ultrasound Med. Biol.* **2023**, *49*, 1616–1626. [CrossRef]
8. Zeng, P.; Liu, S.; He, S.; Zheng, Q.; Wu, J.; Liu, Y.; Lyu, G.; Liu, P. TUSPM-NET: A multi-task model for thyroid ultrasound standard plane recognition and detection of key anatomical structures of the thyroid. *Comput. Biol. Med.* **2023**, *163*, 107069. [CrossRef]
9. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *arXiv* **2018**, arXiv:1801.04381v4.
10. Tan, M.; Le, Q.V. EfficientNetV2: Smaller Models and Faster Training. *arXiv* **2021**, arXiv:2104.00298.
11. Lin, Z.; Li, S.; Ni, D.; Liao, Y.; Wen, H.; Du, J.; Chen, S.; Wang, T.; Lei, B. Multi-task learning for quality assessment of fetal head ultrasound images. *Med. Image Anal.* **2019**, *58*, 101548. [CrossRef]
12. Zhang, B.; Liu, H.; Luo, H.; Li, K. Automatic quality assessment for 2D fetal sonographic standard plane based on multitask learning. *Medicine* **2021**, *100*, e24427. [CrossRef]
13. Pu, B.; Li, K.; Li, S.; Zhu, N. Automatic Fetal Ultrasound Standard Plane Recognition Based on Deep Learning and IIoT. *IEEE Trans. Ind. Inform.* **2021**, *17*, 7771–7780. [CrossRef]
14. Gofer, S.; Haik, O.; Bardin, R.; Gilboa, Y.; Perlman, S. Machine Learning Algorithms for Classification of First-Trimester Fetal Brain Ultrasound Images. *J. Ultras. Med.* **2022**, *41*, 1773–1779. [CrossRef]
15. Guo, J.; Tan, G.; Wu, F.; Wen, H.; Li, K. Fetal Ultrasound Standard Plane Detection with Coarse-to-Fine Multi-Task Learning. *IEEE J. Biomed. Health* **2023**, *27*, 5023–5031. [CrossRef]
16. Han, X.; Yu, J.; Yang, X.; Chen, C.; Zhou, H.; Qiu, C.; Cao, Y.; Zhang, T.; Peng, M.; Zhu, G.; et al. Artificial intelligence assistance for fetal development: Evaluation of an automated software for biometry measurements in the mid-trimester. *BMC Pregnancy Childbirth* **2024**, *24*, 158. [CrossRef]
17. Singh, R.; Mahmud, M.; Yovera, L. Classification of First Trimester Ultrasound Images Using Deep Convolutional Neural Network. In *Communications in Computer and Information Science*; Joaquim, F., Ashish, G., Prates, R.O., Zhou, L., Eds.; Springer International Publishing: Cham, Switzerland, 2021; Volume 1435, pp. 92–105.
18. Ryou, H.; Yaqub, M.; Cavallaro, A.; Papageorghiou, A.T.; Noble, J.A. Automated 3-D Ultrasound Image Analysis for First Trimester Assessment of Fetal Health. *Phys. Med. Biol.* **2019**, *64*, 185010. [CrossRef]
19. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. GhostNet: More Features from Cheap Operations. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 1577–1586.
20. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430.
21. Li, J.; Li, J.; Zhao, X.; Su, X.; Wu, W. Lightweight detection networks for tea bud on complex agricultural environment via improved YOLO v4. *Comput. Electron. Agr.* **2023**, *211*, 107955. [CrossRef]
22. Gevorgyan, Z. SIoU Loss: More Powerful Learning for Bounding Box Regression. *arXiv* **2022**, arXiv:2205.12740.
23. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
24. Chen, W.; Li, C.; Guo, H. A lightweight face-assisted object detection model for welding helmet use. *Expert. Syst. Appl.* **2023**, *221*, 119764. [CrossRef]
25. Wang, C.; Bochkovskiy, A.; Liao, H.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
26. Wang, C.; Yeh, I.; Liao, H. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv* **2024**, arXiv:2402.13616.
27. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving into High Quality Object Detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
28. Fang, Y.; Yang, S.; Wang, X.; Li, Y.; Fang, C.; Shan, Y.; Feng, B.; Liu, W. Instances as Queries. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; pp. 6890–6899.
29. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal.* **2016**, *39*, 1137–1149. [CrossRef]
30. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626.
31. Xie, H.N.; Wang, N.; He, M.; Zhang, L.H.; Cai, H.M.; Xian, J.B.; Lin, M.F.; Zheng, J.; Yang, Y.Z. Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal. *Ultrasound Obstet. Gynecol.* **2020**, *56*, 579–587. [CrossRef]

32. Qu, R.; Xu, G.; Ding, C.; Jia, W.; Sun, M. Standard Plane Identification in Fetal Brain Ultrasound Scans Using a Differential Convolutional Neural Network. *IEEE Access* **2020**, *8*, 83821–83830. [CrossRef]
33. Krishna, T.B.; Kokil, P. Standard fetal ultrasound plane classification based on stacked ensemble of deep learning models. *Expert. Syst. Appl.* **2024**, *238*, 122153. [CrossRef]
34. Płotka, S.; Włodarczyk, T.; Klasa, A.; Lipa, M.; Sitek, A.; Trzciński, T. FetalNet: Multi-task deep learning framework for fetal ultrasound biometric measurements. *arXiv* **2021**, arXiv:2107.06943.
35. Coronado-Gutiérrez, D.; Eixarch, E.; Monterde, E.; Matas, I.; Traversi, P.; Gratacós, E.; Bonet-Carne, E.; Burgos-Artizzu, X.P. Automatic Deep Learning-Based Pipeline for Automatic Delineation and Measurement of Fetal Brain Structures in Routine Mid-Trimester Ultrasound Images. *Fetal Diagn. Ther.* **2023**, *50*, 480–490. [CrossRef]
36. Qu, R.; Xu, G.; Ding, C.; Jia, W.; Sun, M. Deep Learning-Based Methodology for Recognition of Fetal Brain Standard Scan Planes in 2D Ultrasound Images. *IEEE Access* **2020**, *8*, 44443–44451. [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.

sensors

MDPI

Article

A Multi-Parameter Persistence Algorithm for the Automatic Energy Calibration of Scintillating Radiation Sensors

Guglielmo Ferranti [1,2], Chiara Rita Failla [1,3], Paolo Finocchiaro [3,*], Alessandro Pluchino [1,2], Andrea Rapisarda [1,2,4], Salvatore Tudisco [3] and Gianfranco Vecchio [3]

[1] Department of Physics and Astronomy, University of Catania, Via S. Sofia 64, 95123 Catania, Italy; guglielmo.ferranti@phd.unict.it (G.F.); failla@lns.infn.it (C.R.F.); alessandro.pluchino@ct.infn.it (A.P.); andrea.rapisarda@ct.infn.it (A.R.)
[2] Istituto Nazionale di Fisica Nucleare, Sezione di Catania, Via S. Sofia 64, 95123 Catania, Italy
[3] Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali del Sud, Via S. Sofia 62, 95123 Catania, Italy; tudisco@lns.infn.it (S.T.); vecchio@lns.infn.it (G.V.)
[4] Complexity Science Hub, Metternichgasse 8, 1030 Wien, Austria
* Correspondence: finocchiaro@lns.infn.it

Abstract

Peak detection is a fundamental task in spectral and time-series data analysis across diverse scientific and engineering disciplines, yet traditional approaches are highly sensitive to the choice of algorithm parameters, complicating reliable and consistent interpretation. Triggered by the requirement for the energy calibration for the 128 detectors of the PI3SO gamma ray scanner, we introduce a versatile methodology inspired by concepts from persistent homology, extending the traditional notion of persistence to a multi-parameter setting. Our approach systematically explores the space defined by multiple detection parameters and quantifies peak robustness through the hyper-volume in the parameter space where each peak is consistently identified. This volumetric multi-parameter persistence (VM-PP) measure enables robust peak ranking and significantly reduces the sensitivity of detection outcomes to individual parameter selection, demonstrating utility across simulated and experimental spectral datasets. Extensive validation reveals that this method reliably differentiates genuine peaks from noise-induced fluctuations under diverse noise conditions, proving effective in practical spectroscopic calibration scenarios. This framework, general by design, can be readily adapted to diverse signal-processing applications, enhancing interpretability and reliability in complex feature-detection tasks.

Keywords: peak detection; automatic energy calibration; topological data analysis; multi-parameter persistence; persistent homology

1. Introduction

Peak detection is a fundamental task in signal processing, underpinning applications in spectroscopy, mass spectrometry, astrophysics, and biomedical signal analysis [1,2]. This study was specifically motivated by the challenge of performing energy calibration for the 128 scintillation detectors of the PI3SO gamma-ray scanner [3]. The Proximity Imaging System for Sort and Segregate Operations (PI3SO) is an advanced gamma-ray spectroscopic imaging system developed to enhance the characterization and sorting of radioactive waste. It is specifically designed to identify radioactive hotspots within waste drums, facilitating the management of intermediate-level waste (ILW), low-level waste (LLW), and very low level

waste (VLLW) [4,5]. The PI3SO system comprises two arrays of 64 thallium-doped cesium iodide (CsI(Tl)) scintillators, each measuring $1 \times 1 \times 1$ cm^3, optically coupled to 6×6 mm^2 silicon photomultipliers (SiPMs) via optical grease with intermediate refractive index, enabling high-sensitivity gamma-ray detection. This configuration allows for high-spatial-resolution imaging and spectroscopic analysis of gamma-ray emissions. By automating the identification and localization of radioactive sources, PI3SO significantly reduces human intervention and exposure, while improving the accuracy and speed of the waste characterization process.

In this work, PI3SO's detector array serves as the basis for developing and validating a novel peak detection and energy calibration methodology, aimed at automating the processing of gamma-ray spectra collected during radioactive waste sorting operations.

In this context, spectra are acquired from standardized gamma-ray sources—typically ^{137}Cs, ^{60}Co, and ^{22}Na—and a linear calibration is then applied to convert the known peak positions from channel units to energy units. However, identifying the peaks and determining their centroids is still largely a manual or semi-manual process: even when automated routines are used, close supervision and frequent corrections are often required. Traditional algorithms, including those based on wavelet transforms [6] or built-in functions from mathematical software packages, depend heavily on user-defined hyperparameters such as peak width and inter-peak distance. These methods are highly sensitive to small changes in parameter values, often resulting in inconsistent outcomes: a slight adjustment can lead to missing genuine peaks or, conversely, detecting spurious peaks caused by noise [7,8]. Recent advances have sought to mitigate these issues by either training deep neural networks on raw liquid chromatography–mass spectrometry (LC–MS) [7], or employing statistical modeling to reduce reliance on arbitrary thresholds, validating a Bayesian framework for peak detection on challenging data coming from two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (GC×GC-TOF MS) [2]. Similarly, Kilgour et al. [9] introduced an unsupervised threshold optimization approach tailored to mass spectrometry, significantly enhancing detection reliability across instruments. While these methods mitigate noise and offer improvements in robustness, they either remain highly sensitive to individual hyperparameter settings, or often obscure why a peak is retained or rejected, leaving room for approaches that more transparently quantify feature stability. Motivated by these challenges, our work draws inspiration from the field of topological data analysis (TDA) and the concept of persistent homology, which has been celebrated for its ability to capture the underlying shape of data through so-called barcode plots, that display the interval for which a feature is detected when varying a single filtration parameter; by plotting the ranges in which all features (e.g., peaks in a noisy signal) appear and vanish (the birth–death intervals), barcode plots highlight long-lived, and therefore significant, features [10]. Along this line, Scholkmann and colleagues [1] have presented the Automatic Multiscale Peak Detection (AMPD) algorithm, which considers a peak robust if it remains a local maximum at many scales on periodic data. Although effective in single-parameter scenarios, rigorously extending these ideas to multiple parameters encounters fundamental challenges as barcode plots fully characterize data shape for single-parameter filtrations only while, as shown by Carlsson and Zomorodian [11], no equally simple and complete representation exists for multi-parameter settings. To address this, we introduce a practical framework that aggregates peak detection outcomes across a multi-dimensional hyperparameter space Θ, deliberately simplifying the representation by disregarding detailed shape information. We quantify peak robustness as a scalar measure, defined as the accumulated hyper-volume within the hyperparameter space Θ over which each peak is consistently identified. Concretely, each detected peak corresponds to a specific sub-region of Θ, as shown in Figure 1; panels (b) and (c) in particular show that peaks which appear visually more significant in the raw spectra naturally occupy larger associated

hyper-volumes, providing an intuitive basis for ranking stability using their persistence. While exploring an m^n hyperparameter space—where m represents the number of discretized values per parameter and n the number of parameters—can potentially lead to significant computational costs, we effectively mitigate this by employing computationally efficient peak detection routines and extensive parallelization techniques, keeping the runtime of the energy calibration conducted in Section 3.2 at around 40 spectra (each consisting of 2048 channels) per second running the code locally on a laptop. Additionally, our framework is well suited for future enhancements such as adaptive sampling strategies on the parameter space Θ, which can significantly reduce computational overhead by concentrating computational resources in regions of Θ where peak detection outcomes exhibit greater variability.

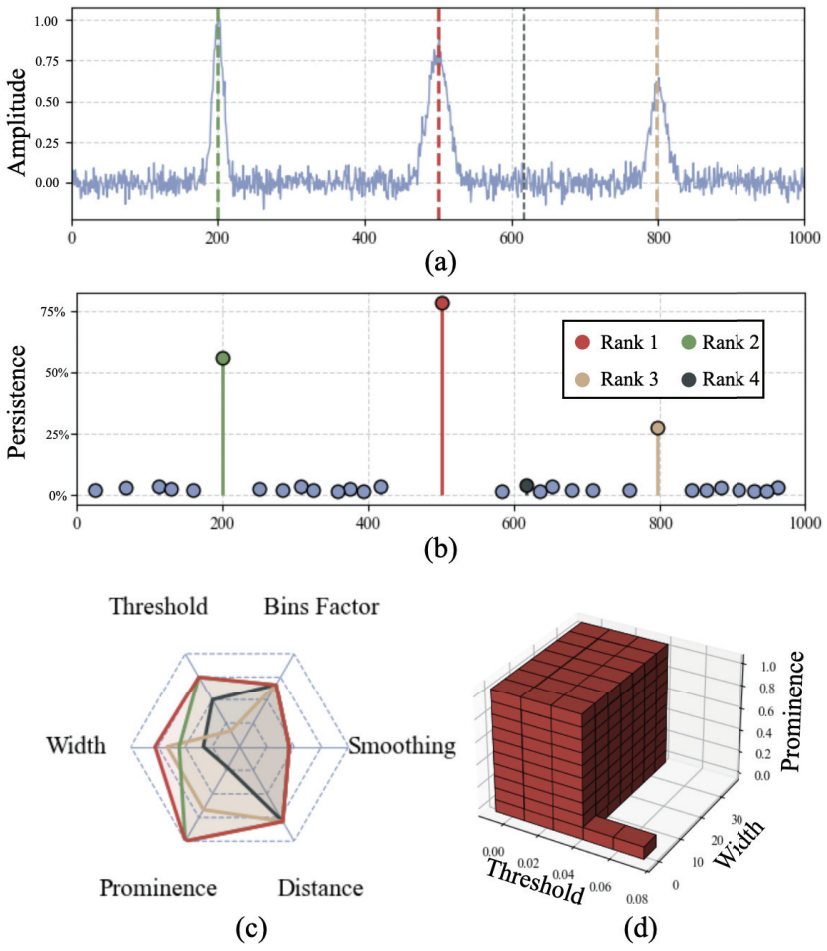


Figure 1. (**a**) Simulated signal, obtained by adding a bounded-range noise ($\sigma = 0.05$) to a clean spectrum consisting of 3 Gaussian peaks, overlaid with the top 4 most persistent detected peaks; because only 3 true peaks exist, rank 4 necessarily corresponds to noise. The colors of lines in a, b and c indicate the corresponding rank. (**b**) True peaks (ranks 1–3) exhibit markedly higher persistence than any noise-induced candidate; the rank 4 point (dark circle) lies well below the true peaks, demonstrating easy separability. (**c**) Radar chart projecting the six-dimensional persistence hyper-volumes to 2D. This is useful for tuning the parameter ranges to explore but, because of linear projection, the axis magnitudes do not preserve the true volume ratios seen in panel b. (**d**) A 3D slice of the threshold–width–prominence subspace, with the region in which the rank-1 peak is consistently detected shaded in red.

The remainder of the paper is organized as follows. Section 2 describes our methodology in detail, Section 3.1 presents validation on simulated data, and Section 3.2 demonstrates the application of our method to the real case of the PI3SO spectroscopic data.

2. Materials and Methods

The identification of peaks in spectral or time-series data typically requires careful tuning of various hyperparameters. Traditional methods evaluate these parameters individually or in limited combinations, resulting in peak detection outcomes that are highly sensitive to specific settings. Our methodology, inspired by the field of topological data analysis and persistent homology techniques [10–13], addresses this by introducing the concept of volumetric multi-parameter persistence or VM-PP. Here, the robustness of a detected peak is quantified by systematically exploring and aggregating detection results across a multi-dimensional parameter space.

Formally, we define the hyperparameter space Θ as a discrete grid spanned by the following parameters, each defined as follows:

- Smoothing window size: Number of neighboring data points averaged to smooth short-term fluctuations.
- Bin aggregation factor: Number of adjacent spectral channels combined into one, reducing spectral resolution and noise.
- Prominence: The minimum vertical distance between the peak and the lowest point to which one must descend to reach a higher peak. It quantifies how much a peak stands out from its surrounding landscape—not just its immediate neighbors—in the context of the whole signal.
- Threshold: The minimum vertical difference between the peak and its *immediate neighbors*. Unlike prominence, this is a strictly local criterion: it filters out small fluctuations or noise spikes that do not sufficiently rise above their direct surroundings.
- Width: Required minimum width of detected peaks at half-prominence.
- Peak-to-peak distance: Minimum separation allowed between neighboring peaks.

The grid cells in this parameter space correspond to unique parameter combinations, each having equal volume ΔV. While cell volume could be defined proportionally to parameter step sizes, such an approach may disproportionately emphasize parameters explored at a more coarse scale. Thus, for simplicity and fairness, we assign each cell unitary volume, making the persistence of a peak simply equal to the count of parameter cells where it is detected:

$$P(f) = \sum_{k \in K(f)} \Delta V_k \tag{1}$$

where $K(f)$ denotes the set of cells in the parameter space where feature f is identified. In other words, under this simplification the method simply measures the fraction of parameter combinations that result in a peak being detected, using this value as a metric proportional to the peak significance. While this may seem a trivial approach, our tests (exemplified visually by Figures 1 and 2) show the remarkable ability of the method to robustly separate genuine peaks from noise-induced fluctuations, minimizing spurious detections when evaluated on simulated signals (Figure 3). Even though in the following we will always apply this simplification, the method remains flexible for future tasks that may benefit from setting specific weights to each explored dimension in order to tune each cell's volume based on the impact of individual parameters on peak detection. Peak detection within each cell utilizes Python's `scipy.signal.find_peaks` function. To enhance computational efficiency, the parameter exploration is structured in two nested grids: an outer grid iterates over preprocessing parameters (smoothing and bin aggregation), while an inner grid extensively covers detection parameters (prominence, width, threshold, and distance). For each outer-grid configuration, data is preprocessed once, and peak detection is executed in parallel across all inner-grid configurations. Detected peaks in closely adjacent positions (within a predefined

spatial merging tolerance) are merged and mapped back to their original spectral channels when bin aggregation is applied. Each peak thus acquires a persistence value quantifying its detection robustness across the parameter space. Peaks are ranked according to their persistence scores, which correlate directly with physical significance and detection stability.

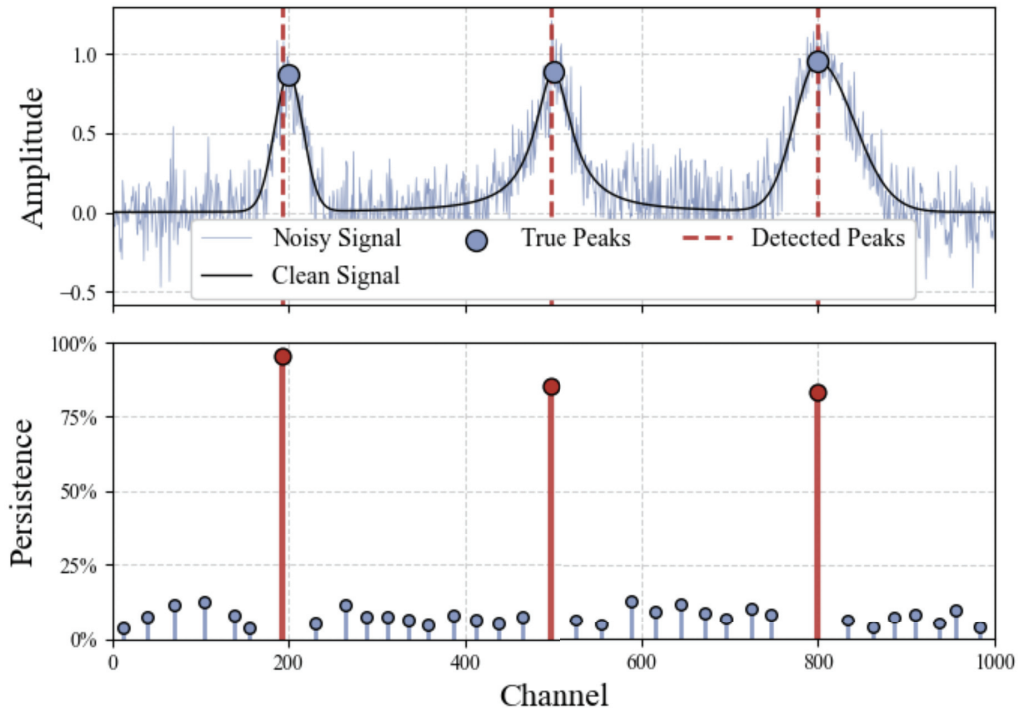


Figure 2. Top panel: Simulated test spectra composed of Gaussian, Lorentzian, and asymmetric peaks under intense Gaussian noise ($\sigma = 0.15$). Vertical lines indicate detected peak channels. **Bottom panel**: VM-PP of peaks expressed as percentage of the total parameter space, allowing peaks that fall below the 25% persistence threshold to be discarded.

To ensure reproducibility, the following parameter ranges were fixed for all applications presented in the paper:

- Smoothing window size: $\{1, 3, 5\}$ (1 = no smoothing).
- Bin aggregation factor: $\{1, 2\}$ (1 = no aggregation).
- Prominence: 10 values uniformly spaced from 0.0 to 1.0.
- Threshold: 5 values uniformly spaced from 0.0 to 0.08.
- Width: 7 values uniformly spaced from 1 to 40.
- Distance: 5 values uniformly spaced from 1 to 30.

These choices were the result of a short trial-and-error selection and offer a practical balance between detailed parameter exploration and computational feasibility. Future implementations may employ adaptive selection of the parameter ranges that automatically maximize the difference in persistence between genuine peaks and noise-induced detections, running a preliminary calibration run on a small subset of hand-labeled signals.

For illustrative clarity, Figure 1 panels (c) and (d) visualize this concept explicitly, showing the hyper-volumes occupied by the top-ranked peaks in the parameter space. Specifically, panel (c) shows a linear projection of all 6 explored dimensions that is useful for tuning the ranges of parameters to explore, while panel (d) shows the 3D width–threshold–prominence parameter subspace, with the region occupied by the most persistent peak of the signal in panel (a) colored in red. This approach is able to effectively separate genuine peaks from noise-induced fluctuations even in intense-noise conditions, as shown

in Figure 2, where the amplitude of fluctuations has a comparable scale to the clean signal. When the number k of expected peaks in a spectrum is known a priori, as is the case for the spectra employed for energy calibration detailed in Section 3.2, one can simply rank all detected peaks by their persistence and take the top k most persistent peaks, discarding all others without the need for additional hyperparameters. For more agnostic exploratory tasks where the number of expected peaks is not known, we can select a persistence threshold below which peaks are discarded; the bottom panel of Figure 2 shows that the method effectively separates genuine peaks from noise-induced detections, making the appropriate range for this threshold quite wide, even for very noisy spectra.

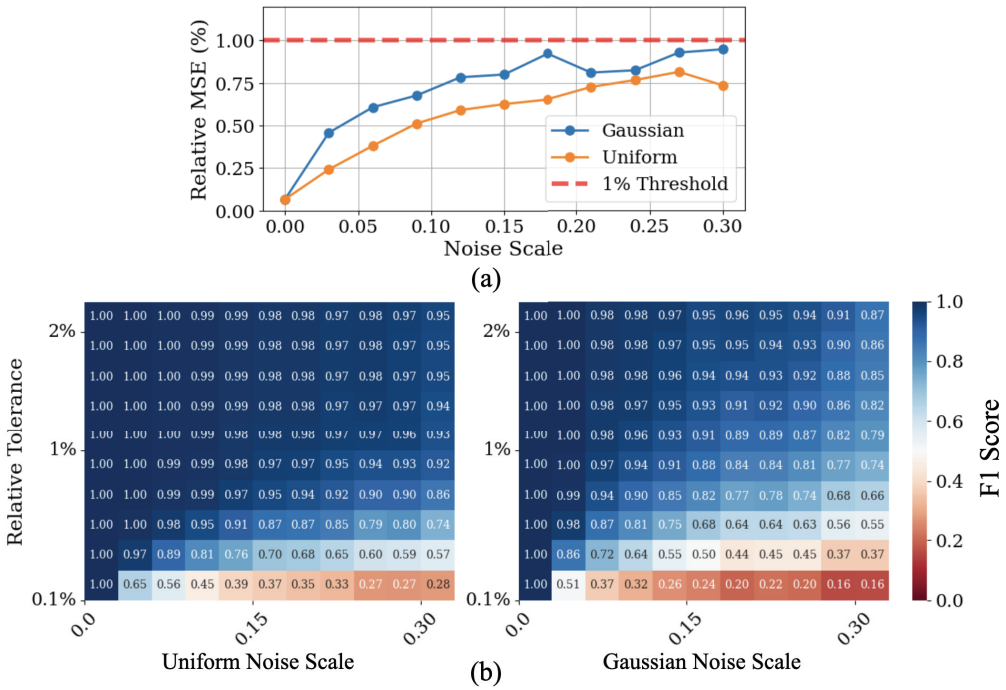


Figure 3. Performance of the multi-parameter persistence approach under varying noise conditions. Panel (**a**) plots the relative mean square error (%) against noise scale for Gaussian (blue) and uniform bounded-range (orange) noise, with a horizontal red line marking the 1% threshold. Panel (**b**) displays maps of detection rate as a function of noise scale (x-axis) and relative tolerance (y-axis) for uniform (left) and Gaussian (right) noise types. Cooler colors in the heatmaps indicate higher detection rates. It is worth underscoring that the ranges explored represent quite heavy noise conditions, as shown in Figure 2, which reports an example spectrum under Gaussian noise with $\sigma = 0.15$.

To contextualize computational performance, our implementation processes approximately 30 to 50 spectra per second (depending on the resolution and size of the parameter space) using parallel computation on a standard laptop (Python version 3.11.x, tested primarily on macOS Sequoia 15.4 but portable to any platform supporting Python), for spectra consisting of 2048 channels. Despite the exponential scaling inherent to a full m^n exploration of the parameter space (with m steps per n parameters), computational costs can be mitigated; specifically, the use of `scipy.signal.find_peaks` ensures minimal per-cell runtime, while independent parameter grid cells are evaluated in parallel using multiprocessing or distributed frameworks. Moreover, because the detection function is piecewise continuous (Appendix A.1) over the parameter space, adaptive sampling strategies can be employed in future implementations to reduce the number of necessary evaluations. The methodology described has been fully implemented in Python, leveraging parallelization libraries to ensure efficiency. To foster reproducibility and support the scientific community, our implementation, along with an interactive demonstration, is publicly accessible via GitHub [14].

3. Results

The following section is divided into two parts: in the first, we benchmark our method on synthetic data and evaluate its performance when varying noise conditions; in the second, we showcase an application of our method to real spectroscopy data.

3.1. Synthetic Data Benchmark

To evaluate the robustness of our multi-parameter persistence method under controlled conditions, we generated a synthetic dataset consisting of 5000 test spectra. Each test spectrum contained three distinct peaks: one Gaussian, one Lorentzian, and one asymmetric, carefully positioned to avoid overlap. The amplitudes of these peaks were normalized so that each clean (noise-free) spectrum remained within the range $(0, 1)$, ensuring that no individual peak exceeded an amplitude of 1.

To simulate realistic measurement conditions, we introduced two types of noise separately into each test spectrum:

- Uniform noise, where each data point was randomly perturbed by an amount between $-a$ and $+a$, with the noise amplitude a varied from 0 to 0.3.
- Gaussian noise (additive white Gaussian noise; AWGN), where each data point was perturbed by a value drawn from a Gaussian distribution with mean 0 and standard deviation σ varied from 0 to 0.3.

An example synthetic spectrum with moderate Gaussian noise ($\sigma = 0.15$) is shown in Figure 2. Detected peaks are marked by vertical dashed lines. The bottom panel illustrates the peak persistence scores, clearly showing how the three true peaks stand out significantly from random noise fluctuations.

The performance of our detection method across varying noise levels is summarized in Figure 3. To quantify detection accuracy clearly, we defined two separate metrics:

- Relative mean detection error: For each detected peak, we measured the absolute difference between the predicted channel (detected peak position) and the actual known peak position. This error was normalized relative to the full range of the spectrum to produce a percentage error. We then averaged this error over all three peaks in each spectrum and over all 5000 test spectra for each noise level. Panel (a) displays this averaged error metric, showing how detection accuracy degrades with increasing noise.
- Detection rate (F1 score): To assess how reliably peaks were found, we used a common classification metric known as the F1 score. This metric combines two important aspects into a single value:
  1. Recall: The fraction of actual peaks successfully detected.
  2. Precision: The fraction of detected peaks that corresponded to actual (rather than noise-induced) peaks.

  The F1 score is defined as the harmonic mean of recall and precision and reaches its maximum value of 1.0 (100%) when all true peaks are detected correctly without false positives. For instance, an F1 score of 0.98 indicates that overall about 2% of detections were incorrect—either missed true peaks or included spurious noise peaks.
  Panel (b) of Figure 3 shows the F1 score for varying noise levels and different tolerance values. The relative tolerance refers to how close a detected peak must be to the true peak position to be considered a correct detection, again expressed as a percentage of the full spectral range. Darker blue areas in the heatmaps indicate excellent detection performance (close to 1), whereas lighter (warmer) colors indicate lower reliability.

The heatmaps demonstrate that our approach maintains strong performance under substantial noise. The uniform noise scenario (panel b) consistently showed excellent perfor-

mance, while the Gaussian noise scenario (panel c) exhibited slightly lower but still robust detection rates, reflecting the more challenging nature of unbounded Gaussian fluctuations.

In summary, the synthetic tests clearly demonstrate that our multi-parameter persistence approach reliably detects and accurately locates peaks even under high-noise conditions. The results underscore both the method's ability to consistently identify all peaks (high recall) and avoid false detections (high precision), qualities that directly translate into practical advantages when applied to experimental spectroscopic data.

3.2. Validation on Spectroscopy Data

To validate the effectiveness of our peak detection and classification methodology, we utilized experimental gamma-ray spectroscopy data obtained using the Proximity Imaging System for Sort and Segregate Operations (PI3SO) [3]. This advanced spectroscopic system automates the identification of gamma-ray-emitting sources within radioactive waste, specifically targeting intermediate-level waste (ILW), low-level waste (LLW), and very low level waste (VLLW) [4,5]. The PI3SO system can improve radwaste (re)conditioning and management by rapidly identifying radioactive hotspots with minimum human intervention.

3.2.1. PI3SO Spectroscopic System

The PI3SO project primarily addresses two interconnected tasks: Hotspot search and gamma spectrometry. Hotspot search involves identifying areas containing radioactive objects (see Figure 4a). Once identified, these hotspots become regions of interest (ROIs) for subsequent detailed gamma spectroscopic analyses, as illustrated in Figure 4b.

The PI3SO instrumentation consists of a robust table coupled with a mechanical sliding bridge. Two linear arrays of 64 gamma-ray detectors, are arranged above and beneath the table. Each array contains four modules of 16 detectors, each detector coupling a cubic (1 cm^3) CsI(Tl) scintillator with a silicon photomultiplier (SiPM). The detectors are encapsulated in reflective casings, enhancing photon collection efficiency. The structural layout and module configuration are detailed in Figure 5.

CsI(Tl) crystals were chosen for their optimal properties, including high density (4.51 g/cm^3), superior energy resolution, and significant light yield (60,000 photons/MeV), with a peak emission around 550 nm [15,16]. Each scintillator is coupled to an MPPC (multi-pixel photon counter) SiPM from Hamamatsu Photonics K.K. (325-6, Sunayama-cho, Chuo-ku, Hamamatsu City, Shizuoka Pref., 430-8587, Japan), featuring an active area of 6×6 mm^2 and 14,331 cells [17]. Previous tests confirmed an energy resolution (FWHM) around 5–6% for the entire detection system [3]. Signals from each detector are digitized using two VX2745 digitizers [18], produced by CAEN SpA (Via Vetraia 11, 55049 Viareggio, Italy) each one capable of simultaneously sampling signals from 64 SiPM channels at 125 MSamples/s with a 16-bit resolution. Data acquisition is synchronized via timestamps, allowing precise temporal alignment of detected events. The acquisition system, interfaced via Ethernet, supports real-time data analysis, spectral generation, and comprehensive event logging. During typical calibration or operational measurements, nearly all 128 detectors are actively engaged, generating extensive datasets. Each dataset typically comprises multiple gamma-ray spectra requiring energy calibration and isotope identification. Given the complexity and volume of the data, manual analysis is impractical. Consequently, an automated, robust peak identification system is essential.

Our validation procedure leveraged detector calibration tasks involving three standardized radioactive sources: ^{137}Cs (1.4 MBq, 0.662 MeV), ^{60}Co (0.056 MBq, 1.173 MeV, and 1.331 MeV), and ^{22}Na (0.015 MBq, 0.511 MeV , and 1.274 MeV). Each of the 128 detec-

tors was exposed to these sources individually to validate automated peak detection and classification without manual intervention.

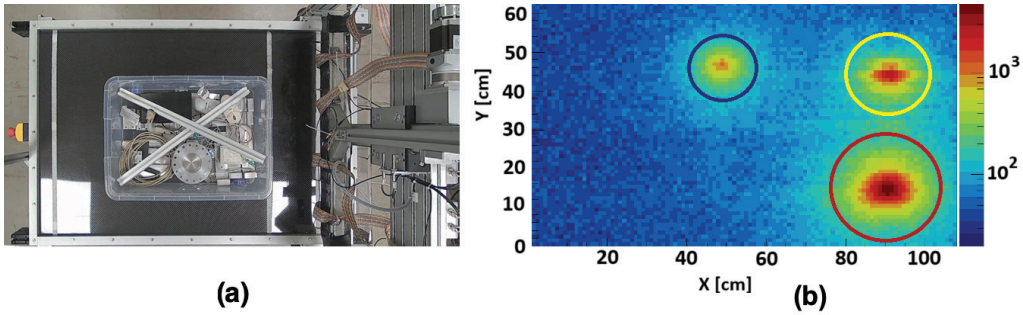


Figure 4. (**a**) System operation diagram: Radioactive waste is distributed over the scanning table to identify radioactive hotspots. (**b**) Visualization of the result of a scan of three different radioactive point-like sources; color scale on the right reports the activity in cps. Red circle: source of ^{137}Cs, with an activity of 1.4000 MBq; yellow circle: source of ^{60}Co, with an activity of 0.0560 MBq; blue circle: source of ^{22}Na, with an activity of 0.0154 MBq.

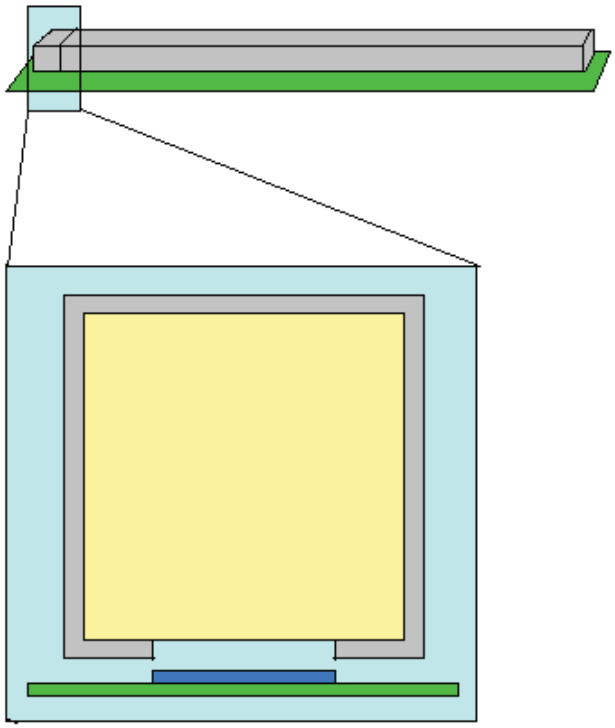


Figure 5. **Top**: Sketch of a single component of one of the two linear arrays, which consists of four modules like the one shown, placed next to each other in a row; each module houses a total of 16 CsI(Tl) crystals embedded in a reflective resin case and coupled to as many SiPMs that are housed on an electronic board. **Bottom**: Sketch of a single detector: In gray is the reflective mask in which the $1 \times 1 \times 1$ cm^3 crystal of CsI(Tl) (in yellow) is embedded; the latter is optically coupled through its free face to a 6×6 mm^2 SiPM (blue), using optical grease with intermediate refractive index to minimize reflection losses at the interface. The SiPM is housed in the electronic board (green).

3.2.2. Peak Ranking and Automatic Selection

Figure 6 illustrates the application of our VM-PP method to typical emission spectra from sodium, cesium, and cobalt radioactive sources, for which the noise scale was estimated to be between $\sigma = 0.05$ and $\sigma = 0.10$ (lower noise scales could be achieved at the cost of increasing the exposure time), highlighting the robustness of the peak-ranking process across spectra with varying complexities. Using our multi-parameter persistence ranking,

we select from each detector the top k most persistent peaks. The choice of k depends on the source's spectral shape, but requires almost no additional filtering.

- For sodium, we set $k = 2$ to catch both the main emission peak at 0.511 MeV and a secondary peak at lower energy, attributed to backscattered photons. This peak originates from gamma rays that escape the detector without interacting and are scattered backward (at angles close to 180°) by surrounding materials, and then re-enter the detector with reduced energy (note that this is not the Compton edge, which corresponds to the maximum energy transferred to an electron during a single Compton scattering event).
- For cesium, $k = 2$ again recovers the main peak (0.662 MeV) plus a secondary region at lower energy.
- For cobalt, $k = 5$ covers the overall complexity of the spectrum well; the two highest energy peaks of those five correspond to the known 1.17 and 1.33 MeV lines.

Initially, three detectors were excluded from the analysis due to severe hardware damage, clearly identified by significant deviations in their calibration fits (Figure 7b,c). Among the remaining 125 detectors, the correct calibration peaks were identified directly within the top k sets for 123 cases without any additional processing. However, in two detectors, the peaks of interest were ranked just outside the top k sets, appearing as the $(k+1)$-th most persistent peaks. We found that minimal additional filtering—such as applying a broad energy/channel cut to eliminate clear backscattering peaks—was sufficient to recover these peaks accurately. Ultimately, using the persistence-based ranking approach combined with this minor post-processing step in the two challenging instances, we successfully identified all four targeted emission peaks across the full set of 125 operational detectors.

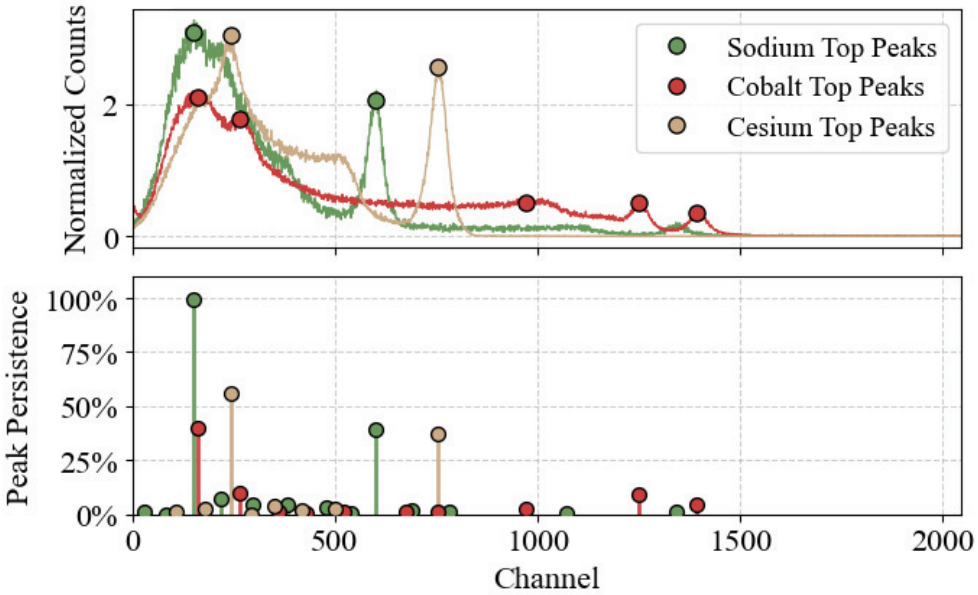


Figure 6. VM-PP applied to peak detection on 3 sample sources: **Top panel** shows the emission spectra of sodium, cobalt, and cesium sources, normalized by area under curve and annotated with the top k most persistent peaks for each source; **bottom panel** shows the persistence of detected peaks, reported as the percentage of volume in parameter space where the peak is detected. For all 3 spectra, despite significant differences in peak shapes, the same portion of parameter space is explored, showcasing the versatility and robustness of our method.

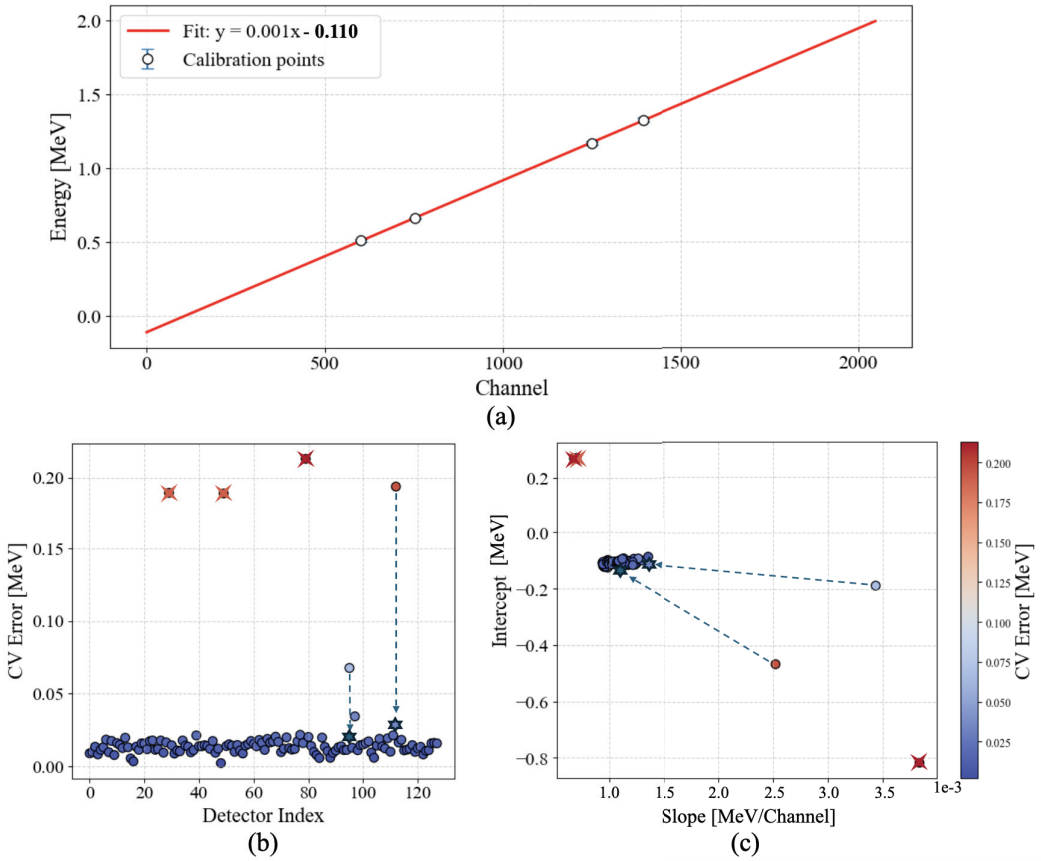


Figure 7. Calibration results: (**a**) An example linear best fit on a single detector's data, using 4 emission peaks of known energy from sodium (0.511 MeV), cesium (0.662 MeV), and cobalt (1.17 and 1.33 MeV). (**b**) The leave-one-out (LOO) cross-validation (CV) error of each detector's calibration, highlighting 3 outlier points resulting from faulty detectors (crossed circles). The 2 remaining points with poor performance (circles) are easily recovered with minimal post-processing, the results of which are shown as stars. (**c**) Distribution of best-fit slope and intercept parameters across all detectors, colored by the LOO-CV error. Negative intercept values arise from well-known non-linearities near the extremes of the detectors' ranges.

3.2.3. Calibration Outcomes

Once the correct peak positions are identified in channel space, a simple linear fit converts channel number to energy. The fit uses the known energies of each peak as the x values and the measured channel positions as the y values. In Figure 7, we illustrate the overall calibration performance.

Panel (a) of Figure 7 shows a typical example of a linear energy calibration fit for a single detector, using the four known gamma-ray emission peaks. To evaluate the reliability of these calibrations across all detectors, we employed a standard technique known as leave-one-out cross-validation (LOO-CV). In this procedure, the calibration is repeated four times per detector, each time leaving out one of the four reference peaks, fitting a line to the remaining three, and then computing the prediction error on the omitted point. Panels (b) and (c) summarize the results of this analysis across the full detector array. For each detector, the individual prediction errors from the four LOO-CV iterations are averaged to yield a single scalar error value. These average errors are reported in panel (b) as a measure of calibration reliability and are also used to color-code the corresponding slope and intercept values in panel (c). Lower average error indicates that all four calibration peaks lie close to a consistent linear fit, while higher error suggests issues such as misidentified peaks or

detector faults. In particular, the three marked detectors with extremely high errors were identified as irreparably damaged and removed from subsequent analyses. The remaining two elevated-error points arise from the two detectors in which the primary cobalt peaks initially ranked as $(k+1)$-th. Because the correct peaks are nonetheless detected (just slightly lower in the persistence order), a mild filter on backscattering peaks fixes the issue.

Panel (c) in Figure 7 summarizes how the fit parameters vary across the detectors. Most calibrations cluster around consistent slope/intercept values, with small variations attributed to manufacturing differences, gain settings, or mild hardware non-linearities. Negative intercepts persist in all detectors, reflecting known edge effects and electronic offsets but still yielding an accurate mapping over the region of interest.

3.2.4. Channel-to-Energy Translations

To visualize the overall outcome of this multi-detector calibration, we compare raw (channel-space) spectra to their energy-space equivalents. Figure 8 shows, for a cobalt source, how the large spread of peak positions among the 128 detectors (left) collapses into well-aligned peaks once the linear transformations are applied and the 3 faulty detectors removed (right).

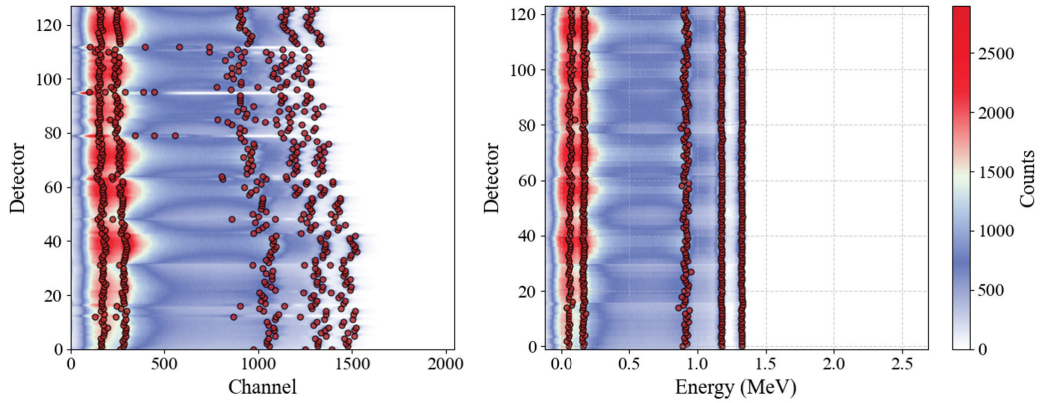


Figure 8. Example cobalt emission spectra from all detectors, annotated with the top 5 most persistent peaks. (**Left**): raw data in channel space reveal large variability in apparent peak locations. (**Right**): the same spectra after linear calibration and removal of 3 faulty detectors, translating channel to energy. The highest two (energy-wise) of these 5 persistent peaks correspond precisely to the known 1.17 MeV and 1.33 MeV emissions. The triangular-shaped shadows visible in the region around 0.3 MeV with a periodicity of 16 detectors depend on the different counting rates due to different center-to-side solid angles when placing the point-like source in front of each 16-unit detector module.

By ranking peaks via multi-parameter persistence, each detector automatically yields the relevant peaks. Even small deviations in detection (as for the two problematic detectors mentioned above) are trivially corrected by discarding the low-energy backscattering regions. Hence, without ever manually labeling any peak, we successfully identify all calibration peaks for 125 detectors, requiring minimal filtering to recover even the two initially mislabeled peaks. This amounts to detecting (and correctly matching) $4 \times 125 = 500$ emission peaks across all spectra, confirming the robustness of the proposed persistence-based approach for automating detector calibration.

4. Discussion

The proposed multi-parameter persistence framework demonstrates substantial advantages in robustness, interpretability, and practical utility for peak detection across both simulated and real-world datasets. The methodology generalizes traditional single-

parameter persistence concepts, quantifying the stability of detected features by integrating across multiple influential detection parameters such as prominence, distance, width, and threshold. This comprehensive approach significantly mitigates the sensitivity to arbitrary parameter selection, an inherent limitation of conventional peak detection algorithms. Validation on test datasets (Section 3.1) underscores the robustness of this approach. Our experiments revealed consistently low detection errors—under 1%—across a wide range of noise intensities, including challenging scenarios with both Gaussian and uniform noise distributions. Heatmap analyses further confirmed the method's resilience, highlighting expansive parameter regions where near-perfect detection rates were consistently achieved, thus establishing its reliability under various practical conditions. When applied to real-world spectroscopic data for detector calibration (Section 3.2), the method's effectiveness was clearly demonstrated. Using automated, persistence-based peak ranking without manual labeling, all critical calibration peaks across a large detector array were successfully identified. The framework proved not only capable of accurately calibrating functional detectors but also adept at diagnosing and isolating hardware faults, streamlining the instrument maintenance process. In cases where the automated method initially ranked correct peaks slightly lower in persistence, minimal post-processing sufficed for full recovery, highlighting the practicality and flexibility of the approach.

An important strength of this method lies in its intuitive interpretability. By assigning each detected feature a clear scalar metric—the accumulated hyper-volume across the parameter space—practitioners gain a straightforward and quantitative measure of peak robustness. This metric inherently simplifies the feature-ranking process, distinguishing meaningful peaks from noise-induced artifacts effectively and transparently. From a theoretical standpoint, this work bridges traditional signal-processing techniques with concepts from multi-parameter persistent homology (MPH). While formal MPH can offer meaningful insights by analyzing data through intricate topological invariants, our simpler volumetric persistence measure provides an accessible yet powerful alternative that is both computationally efficient and easy to interpret. Future research could further explore adaptive or probabilistic sampling strategies to optimize computational efficiency or integrate formal MPH invariants to uncover additional structural insights in complex multi-dimensional datasets.

In conclusion, the multi-parameter persistence method introduced here represents a robust, versatile, and intuitive solution to feature detection tasks across a variety of applications. Its successful application in simulated and real-world spectroscopic scenarios demonstrates its broad potential, offering enhanced reliability, reduced parameter sensitivity, and clear interpretability that are essential for modern scientific and engineering analyses.

Author Contributions: Conceptualization, G.F.; methodology, G.F.; software, G.F. and G.V.; validation, G.F.; formal analysis, G.F.; investigation, G.F.; resources, G.F. and C.R.F.; data curation, C.R.F., G.F., and P.F.; writing—original draft preparation, G.F. and C.R.F.; writing—review and editing, G.F., C.R.F., and P.F.; visualization, G.F. and C.R.F.; supervision, A.R., A.P., P.F., and S.T.; project administration, A.R., A.P., P.F., and S.T.; funding acquisition, A.R. and P.F. All authors have read and agreed to the published version of the manuscript.

Funding: INFN-E committee PI3SO project

Data Availability Statement: The spectroscopy data analyzed in this study were provided under license by the INFN-Laboratori Nazionali del Sud and are not publicly available. The complete Python implementation of the multi-parameter persistence peak detection method introduced in this study is openly available at https://github.com/gullo97/Volumetric-MPP (accessed on 23

July 2025). An interactive online demonstration of the method, allowing users to test it on custom synthetic data, as well as user-provided signals, will also be publicly accessible at the same repository upon publication.

Acknowledgments: G.F., A.P., and A.R. acknowledge support by the University of Catania with the project "Linea di Intervento 1. Progetti di Ricerca Collaborativa" of Department of Physics and Astronomy "E.Majorana". C.R.F. acknowledges support to her PhD from Regione Sicilia in the framework of the PNRR action. The authors also thank Enrico Martello for the feedback regarding some images in this paper. During the preparation of this manuscript, the author(s) used OpenAI's *o3* LLM for the purposes of translating and streamlining text. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

LOOCV	Leave-one-out cross-validation
MPP	Multi-parameter persistence
LC–MS	Liquid chromatography–mass spectrometry
GC×GC-TOF MS	2D gas chromatography with time-of-flight mass spectrometry
VM-PP	Volumetric multi-parameter persistence

Appendix A

Appendix A.1. Piecewise Continuity of the Detection Function

In this section we establish that, under reasonable assumptions, the detection function used in our methodology is piecewise continuous. Let $\Theta \subset \mathbb{R}^n$ denote the hyperparameter space and consider the function

$$f : \Theta \to \{0, 1\},$$

where $f(\theta) = 1$ indicates that a particular feature (e.g., a peak) is detected for the hyperparameter setting θ, and $f(\theta) = 0$ otherwise.

We assume that the underlying data function $D(x)$ is continuous and that the detection of a feature at a candidate location x_0 is determined by inequalities of the form

$$D(x_0) \geq \tau(\theta) \quad \text{and} \quad D(x_0) - D(x) \geq \delta(\theta)$$

for x in a neighborhood of x_0, where $\tau : \Theta \to \mathbb{R}$ and $\delta : \Theta \to \mathbb{R}$ are continuous functions representing, respectively, a detection threshold and a margin requirement.

Suppose first that for some $\theta_0 \in \Theta$ the detection conditions are satisfied (i.e., $f(\theta_0) = 1$). By the continuity of τ and δ, there exists an open neighborhood $U \subset \Theta$ containing θ_0 such that the inequalities

$$D(x_0) \geq \tau(\theta) \quad \text{and} \quad D(x_0) - D(x) \geq \delta(\theta)$$

continue to hold for all $\theta \in U$. Hence, $f(\theta) = 1$ for all $\theta \in U$. Conversely, if $f(\theta_0) = 0$, then the failure of the detection condition persists in an open neighborhood $V \subset \Theta$ about θ_0, so that $f(\theta) = 0$ for all $\theta \in V$.

Thus, aside from the boundaries where the detection outcome changes, the function f is constant—and therefore continuous—on open subsets of Θ. The discontinuities, if any, are confined to the boundaries between these open regions, which are typically of lower dimension. In this sense, the detection function f is piecewise continuous.

This property is essential for justifying adaptive sampling strategies. In regions where f is continuous, a coarse sampling suffices to capture its behavior, while finer resolution can be concentrated near the transition boundaries, thereby mitigating the computational burden of a uniform grid search.

Appendix A.2. Connection to Multi-Parameter Persistent Homology

From a TDA perspective, detecting and tracking features across an n-dimensional parameter space is akin to multi-parameter persistence. Classical persistence homology often focuses on a single parameter (e.g., a filtration value), but Θ here involves multiple dimensions—leading to more complex invariants in multi-parameter persistent homology (MPH) [13]. In principle, one could model feature detections themselves (e.g., peaks) as 0D topological features that appear or vanish across the hyperparameter lattice.

Our extension of "persistence" to a higher-dimensional hyperparameter domain is conceptually aligned with MPH, yet we opt for a simpler, domain-specific aggregator: the accumulated hyper-volume in which a feature is detected. This single scalar score captures the same spirit of "robustness under parameter variation" without the overhead of constructing a full multi-parameter barcode. Its advantages are as follows:

- Straightforward to compute: A grid search (or analogous sampling) across the parameter space directly yields the measure of persistence for each feature.
- Intuitive to interpret: It quantifies how "hard" it is to destroy a feature by shifting the algorithm's hyperparameters.
- Avoids complexity: Formal MPH approaches may require the identification of complex topological invariants, which can lead to intricate representations of the data and high computational overhead.

In the TDA literature, advanced tools like bigraded Betti numbers or rank functions can become unwieldy in higher dimensions, because multi-parameter persistence is known to be "wild" from a representation-theoretic perspective [10–13]. By contrast, our aggregated volume measure is easy to compute and well suited to routine signal-processing tasks. Nevertheless, an interesting future direction is to compare our hyper-volume aggregator with more formal MPH tools to see whether further insights or additional topological structures might be revealed. We view this approach as a simple, application-driven extension of persistence ideas into robust feature detection while still connecting to the broader multi-parameter TDA framework [11].

References

1. Scholkmann, F.; Boss, J.; Wolf, M. An Efficient Algorithm for Automatic Peak Detection in Noisy Periodic and Quasi-Periodic Signals. *Algorithms* **2012**, *5*, 588–603. [CrossRef]
2. Kim, S.; Ouyang, M.; Jeong, J.; Shen, C.; Zhang, X. A New Method of Peak Detection for Analysis of Comprehensive Two-Dimensional GC×GC-TOF Mass Spectrometry Data. *Ann. Appl. Stat.* **2014**, *8*, 1209–1231. Available online: https://www.jstor.org/stable/24522093 (accessed on 20 July 2025). [CrossRef] [PubMed]
3. Poma, G. E.; Failla, C. R.; Amaducci, S.; Cosentino, L.; Longhitano, F.; Vecchio, G.; Finocchiaro, P. PI3SO: A Spectroscopic γ-Ray Scanner Table for Sort and Segregate Radwaste Analysis. *Inventions* **2024**, *9*, 85. [CrossRef]
4. Bernardini, G. Definizione Normativa e Classificazione dei Rifiuti Radioattivi. *Cammino Diritto*, 22 October 2022. Available online: https://rivista.camminodiritto.it/public/pdfarticoli/8838_10-2022.pdf (accessed 22 October 2022).
5. IAEA. *Classification of Radioactive Waste*; IAEA Safety Standards Series No. GSG-1; IAEA: Vienna, Austria, 2009. Available online: https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1419_web.pdf (accessed 22 October 2022).
6. Guo, T.; and Zhang, T.; Lim, E.; López-Benítez, M.; Ma, F.; Yu, L. A Review of Wavelet Analysis and Its Applications: Challenges and Opportunities. *IEEE Access* **2022**, *10*, 58869–58903. [CrossRef]
7. Melnikov, A.D.; Tsentalovich, Y.P.; Yanshole, V.V. Deep Learning for the Precise Peak Detection in High-Resolution LC–MS Data. *Anal. Chem.* **2020**, *92*, 588–592. [CrossRef] [PubMed]

8. Zhang, J.; Gonzalez, E.; Hestilow, T.; Haskins, W.; Huang, Y. Review of Peak Detection Algorithms in Liquid-Chromatography–Mass Spectrometry. *Curr. Genom.* **2009**, *10*, 388–401. [CrossRef] [PubMed]
9. Kilgour, D.P.A.; Hughes, S.; Kilgour, S.L.; Mackay, C.L.; Palmblad, M.; Tran, B.Q.; Goo, Y.A.; Ernst, R.K.; Clarke, D.J.; Goodlett, D.R. Autopiquer—A Robust and Reliable Peak Detection Algorithm for Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 253–262. [CrossRef] [PubMed]
10. Ghrist, R. Barcodes: The Persistent Topology of Data. *Bull. Am. Math. Soc.* **2008**, *45*, 61–75. [CrossRef]
11. Carlsson, G., Zomorodian, A. The Theory of Multidimensional Persistence. *Discret. Comput. Geom.* **2009**, *42*, 71–93. [CrossRef]
12. Lesnick, M.; Wright, M. Interactive Visualization of 2-D Persistence Modules. *arXiv* **2015**, arXiv:1512.00180. [CrossRef]
13. Botnan, M. B.; Lesnick, M. An Introduction to Multiparameter Persistence. *arXiv* **2022**, arXiv:2203.14289. [CrossRef]
14. Demo code for the Volumetric Multi-Parameter Persistence. Available online: https://github.com/gullo97/Volumetric-MPP (accessed on 20 July 2025).
15. Rossi, F.; Cosentino, L.; Longhitano, F.; Minutoli, S.; Musico, P.; Osipenko, M.; Poma, G. E.; Ripani, M.; Finocchiaro, P. The Gamma and Neutron Sensor System for Rapid Dose Rate Mapping in the CLEANDEM Project. *Sensors* **2023**, *23*, 4210. [CrossRef] [PubMed]
16. Longhitano, F.; Poma, G.E.; Cosentino, L.; Finocchiaro, P. A Scintillator Array Table with Spectroscopic Features. *Sensors* **2022**, *22*, 4754. [CrossRef] [PubMed]
17. Hamamatsu Photonics. MPPC S14160-6050HS. Available online: https://www.hamamatsu.com/eu/en/product/optical-sensors/mppc/mppc_mppc-array/S14160-6050HS.html (accessed on 20 July 2025).
18. CAEN Digitizer VX2745. Available online: https://www.caen.it/products/vx2745/ (accessed on 31 May 2025).

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.

sensors

MDPI

Article

Detection and Tracking of Environmental Sensing System for Construction Machinery Autonomous Operation Application

Junyi Chen [1,2], Qipeng Cai [1,2], Xinhai Hu [1], Qihuai Chen [1,2], Tianliang Lin [1,2,*] and Haoling Ren [1]

[1] College of Mechanical Engineering and Automation, Huaqiao University, Xiamen 361021, China; chjy@hqu.edu.cn (J.C.); 23014080001@stu.hqu.edu.cn (Q.C.); 21014080091@stu.hqu.edu.cn (X.H.); chen.qihuai@163.com (Q.C.); happyrhlly@126.com (H.R.)
[2] Fujian Key Laboratory of Green Intelligent Drive and Transmission for Mobile Machinery, Xiamen 361021, China
* Correspondence: ltlkxl@163.com

Abstract

There are a large number of unstructured scenes and special targets in the construction machinery application scene, which brings greater interference to the environment sensing system for Construction Machinery Autonomous Operation Application. The conventional mature sensing scheme in passenger cars is not fully applicable to construction machinery. By taking the environmental characteristics and operating conditions of construction machinery into consideration, a set of environmental sensing algorithms based on LiDAR for construction machinery scenarios is studied. Real-time target detection of the environment, trajectory tracking, and prediction for dynamic targets are achieved. Decision instructions are provided for upstream detection information for the subsequent behavioral decision-making, motion planning, and other modules. To test the effectiveness of the information exchange between the proposed algorithm and the overall machine interface, the early warning and emergency braking for autonomous operation is implemented. Experiments are carried out through an excavator test platform. The superiority of the optimized detection model is verified through real-time target detection tests at different speeds and under different states. Information exchange between the environmental sensing and the machine interface based on safety warning and braking is achieved.

Keywords: construction machinery; autonomous driving; object detection; object tracking

1. Introduction

Construction machinery is widely used in mining, municipalities, mountains and forests, and some other earthwork construction operations. The location of construction machinery operation is relatively fixed, and the process often requires repeating the same action extensively, which may cause operator fatigue. Additionally, the severe operational impacts pose risks to operator health. Currently, the intelligentization of construction machinery has become an important trend of development [1]. As an important part of the intelligentization of construction machinery, the environmental sensing system will directly lead to the decision-making of the whole machine and further affect the safety of operation. At present, considerable research has been carried out on the sensing systems for automatic vehicles. However, construction sites present unique challenges: a large number of unstructured scenes and special targets in construction machinery application scenarios [2], as well as the stringent

regulatory compliance requirements governing safety systems and sensing technology for construction machinery scenes needs further in-depth research.

LiDAR belongs to active vision sensors, with the advantages of far detection and high accuracy not affected by light conditions, and can work in all weather, as well as provide complete scene depth information of the surrounding targets, which has a wide range of applications in high-precision map localization, obstacle detection, target identification, and tracking. It is very suitable for the special application environment requirements of construction machinery. Currently, a large amount of research has been conducted on the environmental sensing based on LiDAR. Zhou and Tuzel proposed the pioneering work VoxelNet [3] for voxel processing based target detection methods, which divided the disordered 3D point cloud into regular voxel features to achieve classification detection and position regression of objects. Yan et al. proposed the SECOND [4] network based on the VoxelNet network, which solved the problem of invalid computation in the blank region and improved the computation speed to 26 fps. The PointPillars [5] network, proposed by Lang et al., for the first time adopted a special voxel division (Pillar), as a way to realize the conversion of three-dimensional features into two-dimensional features, which greatly accelerated the detection speed, with an average detection speed of 62 Hz, and the fastest speed can reach 105 Hz. The CenterPoint [6] network, proposed by Yin et al., introduces the concept of CenterNet [7], from 2D to 3D, for the first time, which enabled 3D target detection through center point detection.

While advanced sensing technologies, such as LiDAR, enable sophisticated environmental perception, their direct deployment within safety-critical control functions faces significant regulatory barriers. This study presents a semantic segmentation-based obstacle prediction framework as a research prototype for collision avoidance. This prototype focuses primarily on perceptual layer algorithms and operates outside the scope of industrial safety standards, such as ISO 13849 [8]Performance Level d or IEC 62061 SIL 2 [9], which define the stringent requirements for safety-related electronic control systems under regulations like the EU Machinery Directive (2006/42/CE) [10]. Certified safety components (e.g., laser scanners) necessitate comprehensive capabilities, including real-time fault diagnosis, guaranteed response times, and high levels of EMC/IP robustness—requirements that are inherently challenging for research-grade systems utilizing frameworks like ROS to guarantee fully fail-safe operation and to mitigate potentially hazardous false positives or negatives. Therefore, this work concentrates specifically on advancing perceptual algorithms for collision detection, acknowledging the distinct objectives and constraints between research prototypes and certified industrial safety systems. In point cloud target tracking, it is mainly categorized into point-based tracking, twin frame-based tracking, and detection-based tracking paradigms, the mainstream method in the field of autonomous driving is still the detection-based tracking paradigm. Luo et al. proposed the FaFNet [11] network, which combined the three modular tasks of detection, tracking, and prediction, and was able to run in real time and with good results. Weng et al. proposed the AB3DMOT [12] algorithm, which was a further extension of the 2D multi-target tracking algorithm to the 3D domain, achieving high computational efficiency with the simplest combination. The CenterPoint [6] network proposed by Yin et al. coupled detection with tracking, which in turn simplified the tracking task.

While advanced perception systems (e.g., LiDAR-based detection/tracking) provide environmental awareness, translating this awareness into safe operational decisions remains a critical challenge—particularly in construction machinery's unstructured environments. Here, reinforcement learning (RL) emerges as a pivotal decision-making paradigm, yet traditional RL suffers from safety risks and sample inefficiency in safety-critical applications [13]. The FHCPL [14] security reinforcement learning framework was proposed to

maximize the rewards while guaranteeing the absolute security of policies, providing reliable solutions for industrial high-risk scenarios. TAG [15] provides lightweight solutions for RL challenges such as sparse reward and high-dimensional continuous control through Gaussian process teacher model with confidence bootstrapping mechanism, and promotes the application of RL in data-scarce scenarios, such as robotics and autonomous driving. FISOR [16] utilizes the reachability analysis of security control theory to transform the hard security constraint (zero violation) in offline reinforcement learning into a feasible region identification problem, and achieves decoupled, efficient training through a novel energy-guided diffusion model. However, the process of identifying feasible regions itself may implicitly rely on or make assumptions about the distribution of rewards and behaviors, and the safety of AI has been controversial, with most of the issues of safety policy, safety complexity, and safety benchmarking [14] remaining to be resolved. Therefore, the starting point of this paper is a study of perceptual content and its combination, and safety-related content will not be introduced in the subsequent content.

Different from passenger cars, there are a large number of unstructured scenes and special targets in the construction machinery context, which brings greater interference to the environment sensing system; the original mature sensing scheme in passenger cars is not fully applicable to construction machinery. In this paper, the detection and tracking of dynamic targets of the environmental sensing of construction machinery scenarios is studied to obtain important environmental information for subsequent dynamic obstacle avoidance, multi-machine cooperation, automatic machine following, and automatic operation, among other tasks. In order to verify the output of the environmental sensing system, dynamic target trajectory prediction and safety warnings are implemented on the real construction machinery only as a perception work contribution, not a safety-certified solution; at the same time, a general attempt is made in automated construction machinery obstacle avoidance to verify the direction of the subsequent research.

2. LiDAR-Based Sensing System for Construction Machinery

2.1. Preliminary Consideration

The environment sensing system for construction machinery used in this paper is composed of LiDAR and the edge computing platform, as shown in Figure 1.

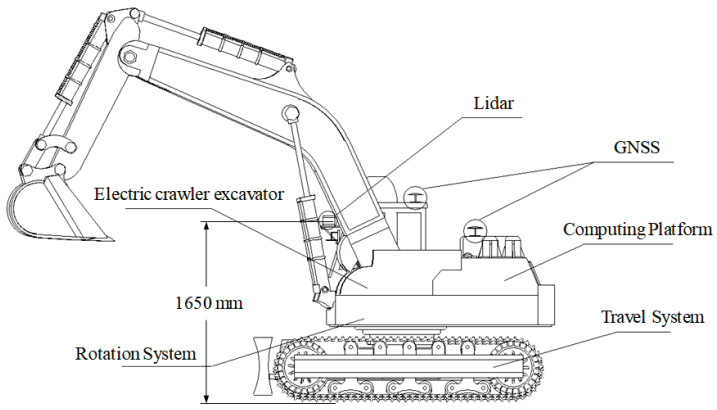


Figure 1. Schematic layout of the sensing system hardware.

2.2. ROS-Based Sensing and Vehicle Control

In order to realize the real-time communication between the environment sensing module and the planning and control modules, a distributed architecture Robot Operating System (ROS) is used [17]. The sensing system and the control strategy designed in this paper is

given in Figure 2, which includes the target detection module, the target tracking module, and the whole machine interaction interface (namely the early warning braking module in this paper). The communication of multiple modules is realized through the ROS. When performing the sensing task, the detection module inputs the point cloud information scanned by LiDAR in real time, and outputs the information of the surrounding targets of interest, such as category, location, size, orientation, and confidence level, which is sent to the visual interactive interface for viewing through the ROS. When each frame detection task is carried out, the output detection information is input to the tracking module in real time. The tracking module uses the detection information of the previous frame to initialize the tracker, and predicts the prediction information of the current frame through the Kalman filter. Then, the Hungarian matching algorithm is used to match the detection information of the current frame with the prediction information. The corrected tracking information is obtained to realize continuous trajectory tracking of multiple targets in consecutive point cloud frames. Finally, the predictive ability of the tracker is used to obtain the target position of the next frame, and the prediction information is transmitted to the safety module through the ROS. The safety module starts to judge the safety distance after receiving the prediction information of each target of the next frame, and different warning safety thresholds and braking safety thresholds have been pre-set according to the different categories in the safety module. When the safety warning conditions are met, the module will send a warning prompt message to the visual interactive interface. When the emergency braking conditions are met, the module will send a reset message to the whole machine electric proportional pressure reducing valve through CAN bus communication. The corresponding emergency braking will be applied.

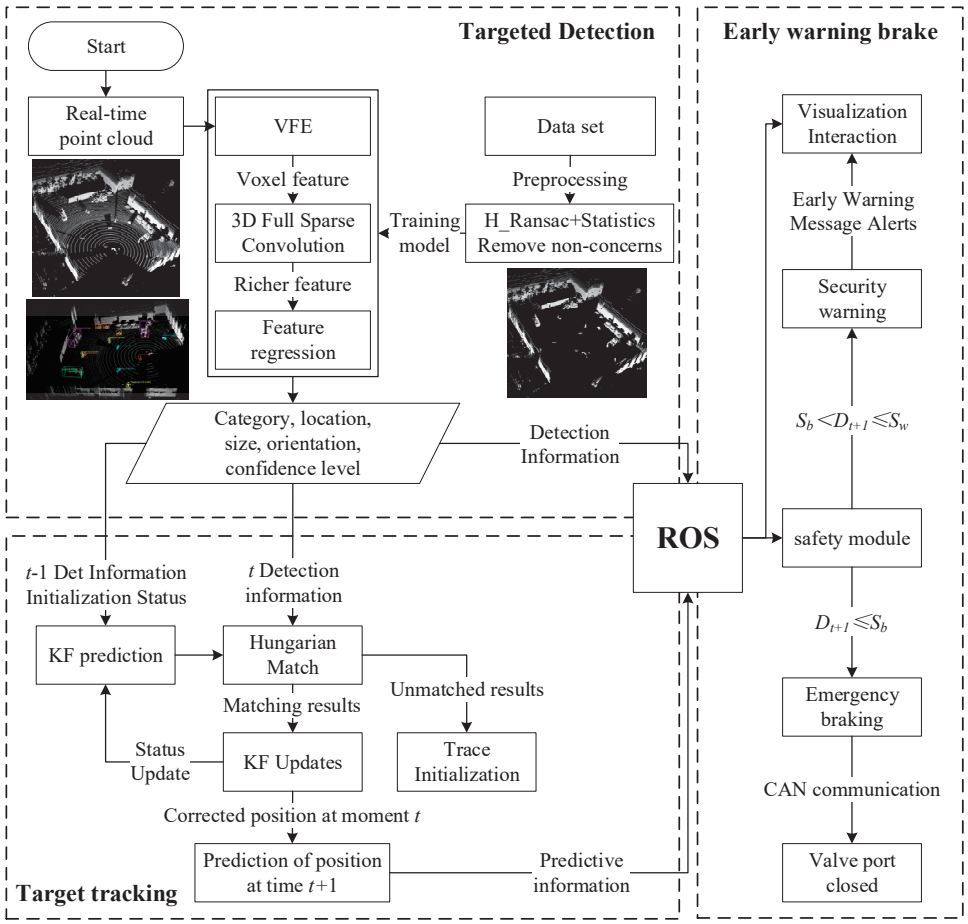


Figure 2. Sensing system software design program.

3. Detection Module

The environment sensing method based on deep learning is adopted in the detection module. Its performance is determined by large-scale high-quality datasets, excellent algorithm models, and high-performance hardware devices.

3.1. Dataset Construction and Processing

The detection effect of the model largely depends on a rich and high-quality dataset. Therefore, constructing a dataset specific to the construction machinery scenario is crucial for the study of target detection and target tracking tasks.

At present, there is no mature open-source dataset in the field of construction machinery. The passenger car dataset cannot be applied to the construction machinery scenario directly. As a result, constructing a dataset specific to the construction machinery scenario is a prerequisite for the research of target detection and the tracking task. In this paper, we refer to the Kitti open-source dataset format and construct the required dataset based on the real-vehicle platform, as shown in Figure 3, which is the data acquisition platform and the layout of the real vehicle.

Figure 3. Data Acquisition Platform and Vehicle Arrangement.

Two batches of point cloud data required for the training of deep learning algorithms are collected using the above data collection platform. In the first batch, point cloud data of seven scenes are collected, including point cloud data of different scenes of sunny daytime versus cloudy daytime and nighttime, as well as point cloud data of different walking modes, such as straight line round trip, driving in a circle, turning at right angles, and zigzag walking, among others. A total of 37,912 frames were obtained. Sensor failure problems caused by bad weather, such as rainy days, are not included in this study. The second batch of point cloud data is also collected for seven scenes, including common excavator behaviors such as traveling toward the truck, to the loading/unloading point, approaching pedestrians, and cockpit slewing. A total of 20,022 frames of point cloud data are collected. In order to be closer to the real operating scenes of construction machinery, and to facilitate repeated trials and tests, the data set chooses the courtyard where the construction machinery is parked as the simulation site. The environment in this data set is mainly simulated as a stockpile area and a narrow passageway.

Considering that the sampling frequency of LiDAR used is 10 Hz, 5500 frames of point cloud data are selected from two batches of data for data labeling by interval frame sampling every 10 frames. The labeling categories are shown in Figure 4. The 5500 frames of point cloud data are divided into 4400 frames for the training set and 1100 frames for the validation set, according to the ratio of 4:1.

Figure 4. Data set labeling.

As shown in Figure 5, which reflects the distribution of the number of 3D point cloud frames in each category of this dataset, it can be seen from the figure that the number of samples in the dataset prepared in this study increases or decreases according to the complexity of the state of the object; thus, it has a certain balance and reasonableness, and it can satisfy the subsequent research of the task of target detection and target tracking.

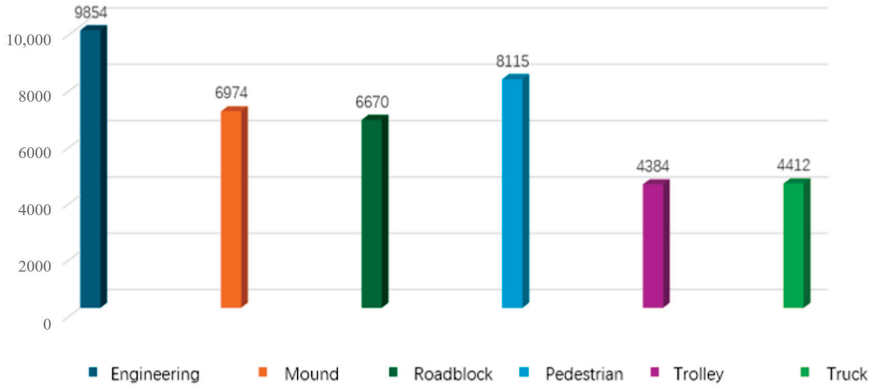


Figure 5. Distribution of data sets.

The quality of the dataset greatly affects the training effect of the algorithm model. In the application of construction machinery, it is more about collaborating with other work equipment and/or staff. The demand for environmental sensing is relatively low for ground points in unstructured scenes. Therefore, in order to improve the quality of the data, to reduce noise and redundancy, to improve data distribution, and to optimize feature extraction, a random sampling method is employed with restricted a height threshold, which preprocesses the existing construction machinery scene dataset based on the random sample consensus (RANSAC) algorithm [18] combined with the restricted height threshold and statistical filtering.

Since the RANSAC algorithm has randomness in initial point selection, it is prone to erroneous filtering when filtering ground points, as shown in Figure 6a. In order to select the correct initial points for plane fitting as much as possible, this paper filters the ground point cloud by restricting the height range of the selected points. To obtain the range interval where the ground points are located in the point cloud frames of the dataset, as shown in Figure 7, the height threshold of the selected points is obtained by counting the distribution of the points before and after the removal of the ground points of the point cloud in a single frame. The effect of the RANSAC algorithm combined with statistical filtering on the pre-processed point cloud data after limiting the height threshold is shown in Figure 6b. This approach not only accelerates the training speed of the algorithm model, but also facilitates the algorithm to better extract the feature information of the target points and improve the robustness and accuracy of the algorithm.

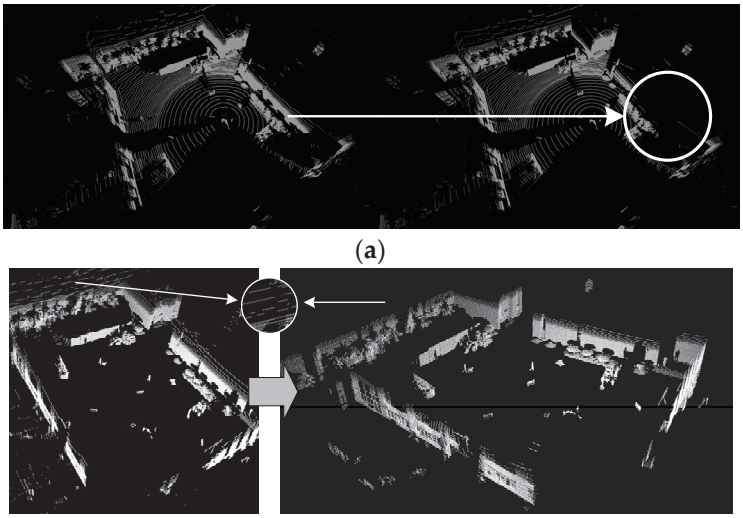


Figure 6. Point cloud preprocessing effect: (**a**) RANSAC error removes ground point cloud effect; (**b**) RANSAC combined with statistical filtering effects.

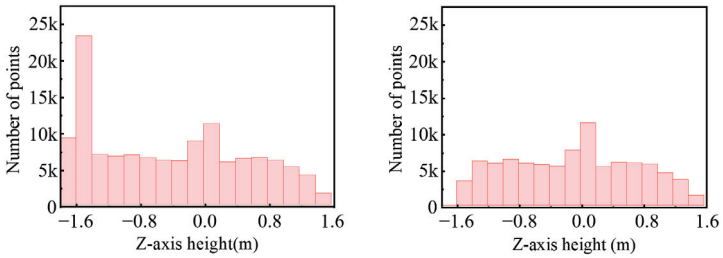


Figure 7. Statistics on the distribution of ground points before and after filtration: (**a**) pre-filter; (**b**) after filtration.

3.2. Detection Network and Accuracy Effects

As shown in Figure 8, it is the point cloud 3D target detection network framework adopted in this paper, which includes three modules of voxel feature encoding (VFE), 3D sparse convolution, and multi-target detection header. After the original point cloud is input into the network, the unordered original point cloud is first divided into voxel squares of the same size according to a specific size [0.1, 0.1, 0.1] in the specified point cloud range [−43.2, −49.6, −2, 43.2, 49.6, 2] by the VFE module for feature encoding. The point cloud space is divided into 864 × 992 × 40 voxel grids that facilitates subsequent feature extraction of point cloud voxel features by the network.

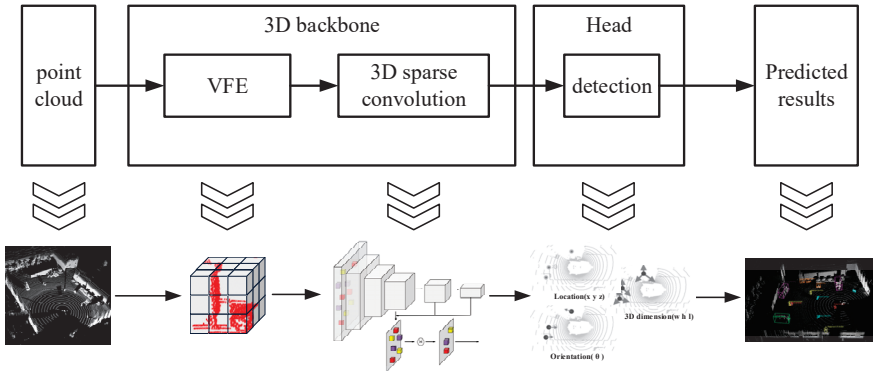


Figure 8. Point cloud 3D target detection network framework.

The 3D sparse convolution module uses a combination of sub-fluid convolution and spatial 3D sparse convolution to form a convolution layer for feature extraction of the VFE feature $N \times C$, which is converted to the standard voxel space of dimension $N \times 5$. Dimension 5 is [B,C,$grid_z$,$grid_y$,$grid_x$], respectively. The network sensory field is increased by means of multilayer convolutional layers, and the sparse features of the last two layers are combined into the fourth layer for summation at the same location by means of high compression, and transferred to the two-dimensional BEV sparse features.

Finally, the multi-target detection head module directly uses the 2D BEV sparse features output from 3D sparse convolutional compression to predict the confidence scores of the six types of voxels, the training is supervised by the Focal Loss function [19], the voxels closest to the center of the truth box are assigned as positive samples, and the sparse voxel features based on positive samples are supervised by the Loss function [20] to regress to obtain the bounding box size and location information. There is a multilayer perceptron (MLP) in each detection head, and the prediction information includes x, y, z, l, w, h, θ.

The 3D target detection network used in this paper is trained on 4400 frames of the self-made construction machinery training set during the training process. Due to the large scale of point cloud data, there is no need to image data like a large number of rounds of training before converging to the optimal. The number of training rounds is set to 50 epochs, and the batch size of the training is set to 8. It can be seen that a total of 27.5 K, as shown in Figure 9, is the change of the loss value during the training process.

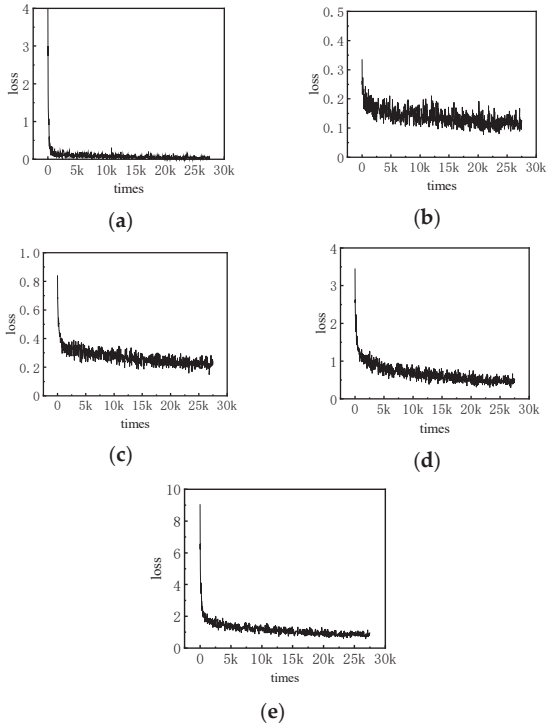


Figure 9. Loss change curve during training: (**a**) classification loss; (**b**) IOU loss; (**c**) DIOU loss; (**d**) localization loss; (**e**) total loss.

Figure 9a–d shows the classification loss, IOU loss, DIOU loss, and localization loss of the target detection network, which correspond to the classification, the degree of matching between the prediction frame and the truth frame, the distance between the prediction frame and the truth frame, and the localization task, respectively; Figure 9e is the total loss curve, which is the weighted sum of the first four types of loss values. It can be found that the four types of loss function curves and the total loss curve at the beginning of training very

rapidly decline, then began to enter the fine-tuning phase (tune-turning), and finally the trend is gradually smooth and finally stabilized at a smaller value near, indicating that the network has been converted to a more optimal state. The curve trend reflects the training process of the network as well as the number of training rounds of the setup of the rationality. The curve trend reflects the rationality of the training process and the setting of the number of training rounds.

In order to evaluate the effect of the trained model, AP and mAP, which are commonly used evaluation metrics for 3D target detection algorithms, are used to evaluate the model. The formulas are as follows [21]:

$$AP|_{R_{40}} = \frac{1}{40} \sum_{r \in R_{40}} \max_{\tilde{r} \geq r} \rho(\tilde{r}) \tag{1}$$

$$mAP = \frac{1}{n} \sum_{i=1}^{N} AP_i \tag{2}$$

where $\rho(\tilde{r})$ is the precision at the recall equipartition point and $AP\left|_{R_{40}}\right.$ is the mean value of the precision at the 40 recall equipartition points calculated.

Since the size and motion state of each category in the actual scene are different, in order to ensure the reasonableness of the evaluation, adhering to the principle of setting larger thresholds for easy-to-detect targets and smaller thresholds for difficult-to-detect targets, different evaluation thresholds are set for different categories. Table 1 shows the detection accuracy of this network training model in each category on the self-made construction machinery dataset.

Table 1. 3D Target Detection Network Detection Accuracy for different categories.

Categories	Threshold	AP
Engineering	0.7	76.35
Truck	0.7	84.17
Pedestrian	0.35	71.46
Trolley	0.5	71.03
Mound	0.5	78.18
Roadblock	0.5	80.38
mAP	-	76.93

Figure 10 shows the detection effect of this 3D target detection network training model visualized in the demo, in which different color detection bounding boxes are set according to different categories and target categories, and confidence, distance, and other information are visualized. All targets appearing in the scene are all detected.

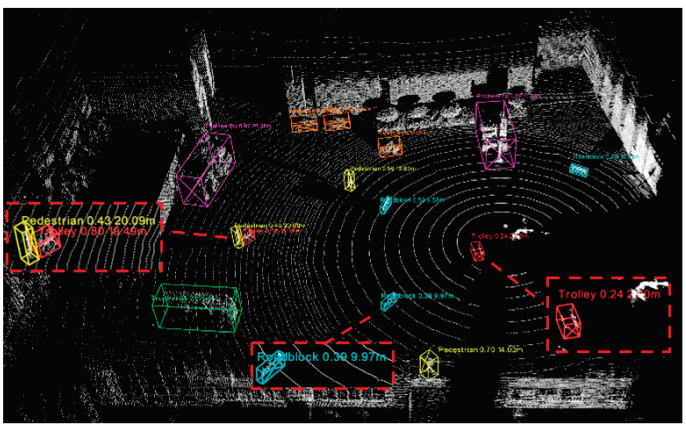


Figure 10. Visualization of target detection effects.

3.3. Detection Module Comparison and Ablation Test

(1) Comparison test

In order to verify the superiority of the performance of the 3D target detection network used in this paper, a comparison test between the network and the current mainstream four 3D target detection network detection effect are conducted, as shown in Table 2. The CenterPoint [6] algorithm is the feature encoding part of the two forms of Pillars and Voxel, therefore, the two participate in the algorithm at the same time to compare the detection accuracy. The methods involved in the comparison are PointPillars [5], SECOND [4], CenterPoint-V [6], CenterPoint-P [6], and VoxelNet [22], and the main comparison indexes are the mAP and the AP value of each category for each detection model trained for 50 rounds, and the network was trained by choosing adam as the optimizer, with the learning rate set to 0.003, weight decay to 0.01, momentum set to 0.9, and the learning rate tuning strategy to one cycle, setting each training batch size to 4.

Table 2. Comparison with other algorithms.

Meth	mAP	Bloc	Ped	Tro	Mou	Eng	Tru
PP	73.2	***82.7***	47.7	68.0	78.5	76.3	***85.8***
SEC	67.0	64.9	54.0	57.3	74.4	71.4	80.0
CPV	61.8	54.2	27.8	59.2	77.9	72.2	79.7
CPP	68.4	67.5	31.7	72.8	***80.8***	73.7	84.0
VoN	75.4	78.1	67.9	67.3	78.2	***77.7***	83.1
Our	***76.9***	80.4	***71.5***	***71.0***	78.2	76.4	84.2

In the models for comparison, the exact same dataset division was used and the same data-preprocessed point cloud data was used for training. As can be seen in Table 2, the 3D target detection algorithm used in the paper has an overall improvement in detection accuracy on the self-made construction machinery validation set compared to other algorithms. The detection algorithm in this paper improves 49.9% in detection accuracy for pedestrians, 4.4% for dump trucks, and 5.2% for mAP compared to the PointPillars algorithm. Compared to the VoxelNet [22] target detection network proposed by Chen et al. in 2023, our algorithm also improves 2.1% in overall detection accuracy for mAP. Compared to the CenterPoint [6] network based on CenterPoint detection, both in Voxel and Pillars form, it improves 24.4% and 12.5% in overall detection accuracy, respectively. From the comparison in the table, in the detection of the two categories of staff and dump truck, the detection accuracy of the network used in this paper is higher than that of other networks, which can be judged that the 3D target detection network used in this paper has better performance in the detection of the categories with more morphological changes and smaller size.

(2) Ablation test

In order to verify whether the operations used in this paper to optimize the performance of the 3D target detection network are effective and reasonable, ablation tests for different optimization operations on a self-made validation set of construction machinery are conducted.

There are three main measures to optimize the performance of the network, which are non-interest point removal for the dataset, increasing the IoU loss, and adjusting the weight coefficients. As shown in Table 3, the experiment is divided into four cases, the first case does not use any optimization measures and only relies on the original network trained on the homemade dataset without preprocessing, and the resulting detection model evaluates

a mAP of 60.63 in the validation set. The second case employs the original network trained on the homemade dataset that has undergone the above mentioned dataset preprocessing operations, and the mAP is 75.38, which is an improvement of 24.3%. The third scenario modifies the loss function part for this 3D target detection network by adding the IoU loss function to help the model improve localization, with a mAP of 76.22, an improvement of 1.1%. Finally, the weight coefficients of the loss function are further adjusted, and the default weight coefficients of classification loss, regression loss, and intersection and parallel ratio loss of this network are 1.0, 2.0, and 2.0, respectively, and the best results are obtained by testing and adjusting them to be 1.0, 1.0, and 1.0, with the mAP of 76.93, which is an improvement of 0.93%. After the ablation test analysis, it shows that the three performance optimization operations used in this network have different degrees of enhancement to the network, and have a certain degree of effectiveness.

Table 3. 3D Target Detection Network Detection Accuracy.

H_RANSAC Statistics Filter	IoU Loss	Weight Coefficient	mAP
			60.63
√			75.38
√	√		76.22
√	√	√	**76.93**

4. Tracking Modules and Early Warning Strategies

In practical operational scenarios, a single 3D target detection cannot meet the diverse sensing needs. Target detection is only limited to the environment sensing of single-frame data streams, but cannot support the connection between frames, so it is necessary to realize the sensing connection between multiple frames through 3D multi-target tracking.

4.1. Tracking Module

Detection-based tracking paradigm is still the dominant method in the field of autonomous driving, which is able to realize multi-target tracking well compared to point-based tracking and twin-frame-based tracking approaches without cumulative error, where detection and matching are being recalculated in every frame. The detection-based tracking paradigm is divided into three parts. including motion prediction, data association, and trajectory cycle management. Figure 11 shows the flow schematic of the tracking module in the sensing system of this paper.

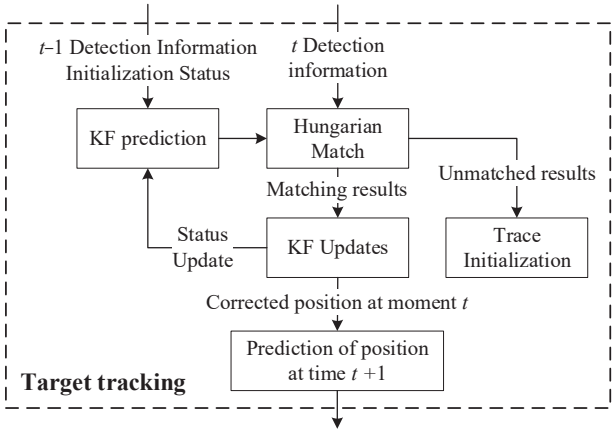


Figure 11. Target Tracking Algorithm Framework.

(1) Movement Forecasting

Given that construction machinery (e.g., excavators ≤ 10 km/h, loaders ≤ 8 km/h) and surrounding targets (e.g., trucks ≤ 50 km/h, pedestrians ≤ 6 km/h) operate at consistently low speeds with minimal acceleration variations (typically $|a| < 0.5$ m/s^2 during steady-state operations), we model their motion as a near-constant velocity. The Kalman filter algorithm is used for motion prediction.

The state prediction equation can be expressed as follows:

$$\hat{x}_t^- = F\hat{x}_{t-1}^- \tag{3}$$

$$P_t^- = FP_{t-1}F^T + Q \tag{4}$$

where $\hat{x}_t^-$ is the a priori estimate of the target state at moment t; $\hat{x}_{t-1}^-$ is the a posteriori estimate at moment $t-1$; F is the state transfer matrix; P_t^- is the a priori covariance matrix at moment t; P_t^- is the a posteriori covariance matrix at moment $t-1$; and Q is the process noise matrix.

The status update formula can be given as follows:

$$K_t = P_t^- H^T (HP_t^- H^T + R)^{-1} \tag{5}$$

$$\hat{x}_t = \hat{x}_t^- + K_t(Z_t - H\hat{x}_t^-) \tag{6}$$

$$P_t(I - K_t H)P_t^- \tag{7}$$

where K_t is the Kalman gain matrix representing the weights; R is the observation noise matrix; Z_t is the 3D detection information at moment t; H denotes the observation matrix of the system at a certain moment; I is the unit matrix; and P_t is the a posteriori covariance.

State prediction is preceded by initialization of the state quantity x for the detection information output by the target point cloud input. As such, x can be expressed as follows:

$$x = [x, y, z, l, w, h, \theta, v_x, v_y] \tag{8}$$

(2) Data Association

In order to make a connection between the previous and subsequent frames, it is necessary to correlate and match the historical frame prediction frame with the current frame detection frame. Firstly, the tracker is initialized with the previous frame detection information, and the current frame information is predicted by the Kalman filter. Then the similarity between the detected frame and the predicted frame is measured by the Mahalanobis distance, and the cost matrix is constructed. Then, the Hungarian algorithm is used to match the predicted frames with the detected frames in a cascade matching strategy, which prioritizes the detected frames to match with the predicted frames with a shorter vanishing time to avoid the frequent switching of tracking identifiers.

The Mahalanobis distance is not affected by the magnitude in the target detection information, and size, position, orientation, and other information magnitude are not the same. The Mahalanobis distance can be a good measure of the proximity between two points of information while ignoring the magnitude. The formula for the similarity measure of the Mahalanobis distance is shown as follows:

$$d^{(1)}(i,j) = (d_j - y_i)^T S_i^{-1} (d_j - y_i) \tag{9}$$

where d_j is the position of the jth detection frame, y_i is the position of the ith prediction frame, and S_i^{-1} is the covariance matrix between the detection and prediction frames.

(3) Trajectory cycle management

In order to avoid association matching error in multi-target tracking task, the prediction box is set to three states, including the tracking state, the pending tracking state, and the pending deletion state. When the detection frame is successfully matched with the prediction frame for the first time, it is in the pending tracking state, it starts the tracking counting, and turns to the tracking state if the number of matches reaches the threshold, and vice versa; when the prediction result fails to find a match for the detection frame, it enters into the pending deletion state, and deletes the target track if the number of losses reaches the threshold.

Different tracking thresholds and deletion thresholds are set accordingly for different dynamic and static target motion situations. The dynamic target motion situation is complex and variable, so larger thresholds are set, as shown in Table 4.

Table 4. Trajectory management thresholds.

Categories	Tracking Threshold	Deleting Thresholds
Engineering	3	4
Truck	3	4
Pedestrian	3	4
Trolley	1	2
Mound	1	2
Roadblock	1	2

4.2. Security Alert Strategy

The early control strategy based on the safety distance model typically consists of two levels: early warning and braking. The first level involves issuing a warning when the relative distance between the self-vehicle and the target vehicle, measured by the information acquisition module, reaches the critical threshold. This triggers an audible warning to prompt the driver to brake. The second level involves braking, where the system tracks the target's trajectory using sensory information, predicting its future position and relationship with the vehicle. By assessing whether the target will enter the warning or emergency braking zone, the system can issue an appropriate warning or braking signal. This ensures the safety of the electric crawler excavator during operations and multi-machine cooperation, preventing collisions and ensuring safe operation [23,24]. As a test of the perception module, we designed a simple warning strategy, as shown in Figure 12, which is a validation of our perception algorithm and confirms the feasibility of the perception algorithm.

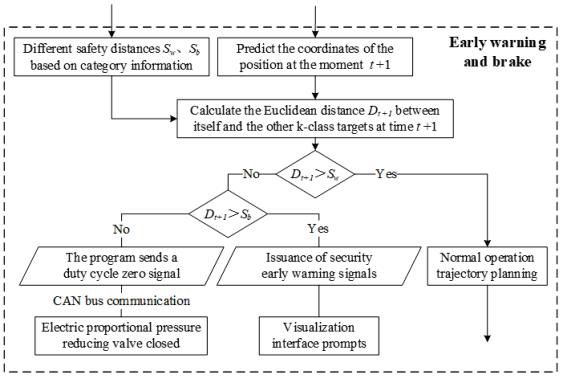


Figure 12. Early Warning Braking Strategy.

In the selection of safety distances, Gounis and Bassiliades [25] simulated a three-layer AEB control system that evaluates collision risks by comparing the relative distance from the vehicle ahead with an adaptive speed-based distance threshold; and Park et al. [26] calculated the minimum distance for collision avoidance during braking and steering, taking into account the relative motion of surrounding vehicles and the lane information obtained through vision sensors. As shown in the above conventional methods, the safety distance can be dynamically adjusted in conventional vehicle driving scenarios after sensing the content, but the complexity of the construction machinery operating scenarios is far more than that of conventional vehicle driving scenarios; at the same time, its characteristics, such as low speed and fixed dynamic objects, allow for the use of a more fixed and safe safety distance. Therefore, the safety distance in this paper is determined by actual excavator experiments, and a relatively safe threshold is set.

Returning to the safety warning strategy, as shown in Figure 12, using the prediction information as an input, a warning braking strategy framework is constructed. First, according to the different categories of information to delineate the different warning safety distance S_w and braking safety distance S_b, set the warning safety distance and braking safety distance to take into account that the excavator itself also exists in the safety radius S. At the same time, using the Euclidean distance to calculate the position of the next moment of its own position and the other k categories of targets. distance D_{t+1}, then determine whether the predicted distance D_{t+1} is greater than the warning distance threshold S_w; if it is greater, then continue to maintain the use of real-time detection and tracking information for trajectory planning and other tasks, and vice versa. Then, enter into the next judgment logic, determine whether D_{t+1} is greater than S_b; if it is greater, then send a warning signal to the visualization interface for prompting, and vice versa. Then, send an emergency stop signal and send a zero message to the proportional pressure-reducing valve through CAN bus communication to close the valve to achieve safety. The zero-setting message is sent to the electric proportional pressure reducing valve through CAN bus communication to close the valve port and realize safe emergency braking.

For the above warning and braking strategy, it is necessary to set different safety thresholds in advance according to different categories, as shown in Table 5, and the basis for setting the thresholds mainly refers to the statistics of the maximum length and width dimensions of each category of targets in the labeled dataset. Combined with the actual motion state of each type of target, the corresponding warning safety distance S_w and braking safety distance S_b are set for each category.

Table 5. Maximum length and width statistics and thresholds for markup boxes.

Category	Maximum L (m)	Maximum W (m)	Warning S_w	Braking S_b
Eng	6.61	2.85	9	6
Tru	5.41	2.37	9	6
Ped	1.01	0.86	9	4
Trol	1.83	0.85	6	4
Mou	2.05	3.26	6	-
Roadb	0.41	1.56	6	4

Construction machinery and trucks belong to large moving targets, taking into account the excavator's own maximum digging radius, the two warning safety distance and braking safety distance are set to 9 m and 6 m, respectively The staff belong to small targets and the activity is larger, taking into account the redundancy of pedestrian safety, so the warning safety distance and the braking safety distance are set to 9 m and 4 m, respectively. The soil pile is an operational target, so only the warning safety distance is set to 6 m. The dump truck and the

roadblock belong to medium-sized targets and are generally stationary, so the warning safety distance and the braking safety distance of the two are set to 6 m and 4 m, respectively.

4.3. Track Matching

In order to verify the effectiveness of the detection-based tracking algorithm for dynamic targets, three dynamic targets, construction machinery, workers, and trucks, were selected to observe the position fitting degree between their detected and predicted positions. In a 35 s continuous point cloud frame, since the actual scanning frame rate of LiDAR is 90% of the theoretical one, a total of 319 frames are obtained, and the number of frames that can be matched between the predicted and the detected positions is 318 frames. As shown in Figure 13, in the global map of the predicted and detected positions, the three types of targets show a good degree of fit, and the overall tracking effect is remarkable, which indicates that the algorithm used in this module has a certain degree of effectiveness.

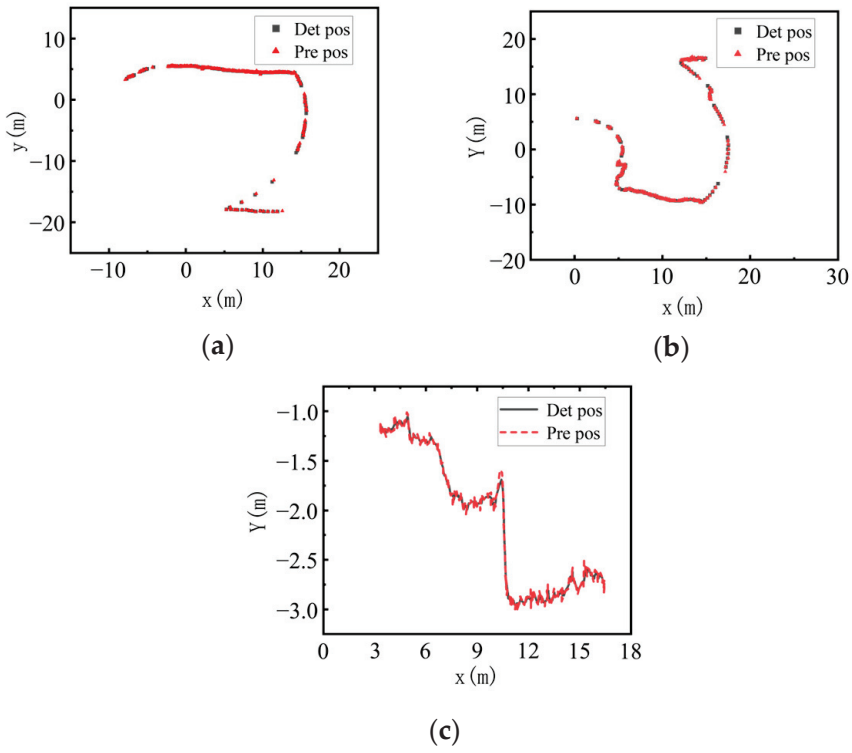


Figure 13. Global map of prediction and detection locations. (**a**) Engineering. (**b**) Pedestrian. (**c**) Truck.

5. Real-Vehicle Testing

In order to validate the effectiveness of the proposed sensing system, further real-vehicle tests are carried out for the accuracy, real-time performance, and dynamic safety warning of the detector, as shown in Figure 14 for the real-vehicle test platform. Table 6 shows the configuration parameters of the sensing system.

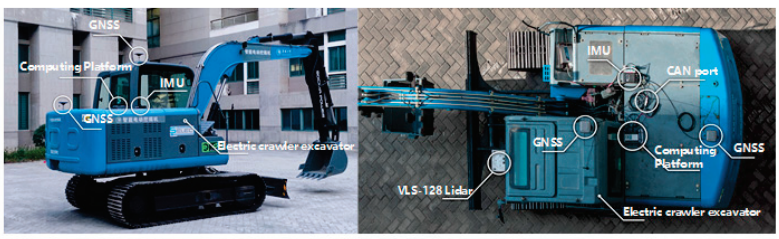


Figure 14. Vehicle test platforms.

Table 6. Specific configuration of the test environment.

Environment	Trial	Deployment
Processors	RTX 3090	Jetson AGV Orin
Ubuntu	18.04	20.04
ROS	Melodic	Noetic
CUDA	11.1	11.4
Python	3.8	3.8
PyTorch	1.8.1	1.11.2

5.1. Verification of Detection Accuracy

There are a variety of working conditions and motion states in the actual driving operation scenarios; in order to verify the accuracy of the detector under different circumstances, the following two sets of tests are designed.

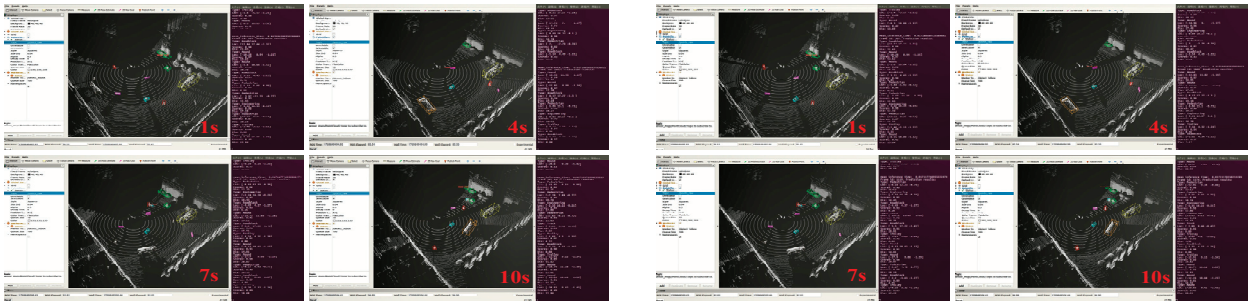
(1) Target detection test at different speeds

The electric crawler excavator used in this test has two motion modes, slow mode (turtle gear) and fast mode (rabbit gear). As shown in Table 7, the speed of the slow mode is 2.5 km/h, and the speed of the fast mode is 3.8 km/h, and the speed of the motion usually has a direct effect on the test.

Table 7. Excavator Motion Mode.

Mode	Velocity
Slow (turtle gear)	2.5 km/h
Fast (rabbit gear)	3.8 km/h

As shown in Figure 15, traveling at two speeds for 10 s respectively, the point cloud frames at 1 s, 4 s, 7 s, and 10 s are selected to show the real vehicle detection results, and 20 frames of point cloud statistics of 10 s consecutive frames are extracted to show the real vehicle detection effect, as shown in Table 8.



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

Figure 15. The effect of different speeds of real car testing: (**a**) the rapid mode real vehicle inspection effect; (**b**) the slow mode real car inspection effect.

The real-time target detection test conducted at different speeds shows that the average detection rate for different categories in the slow mode can reach 84.05%, and the same can have a high average detection rate of 82.03% in the fast mode.

(2) Target detection test under different motion states

Excavators in the actual working conditions usually have a variety of motion states, common motion states, including straight-line driving, turning, and cockpit rotation,

among others, in the excavation operation will be rotated through the cockpit to adjust the digging position, as well as dumping soil loading and unloading, for these three types of common motion states for the real-time target detection test.

Table 8. Comparative statistics of target detection at different speeds modes.

Mode	Category	GT/pc	Det/pc	Rate (%)	Average (%)
Fast	Eng	20	19	95.00	82.03
	Tru	20	15	75.00	
	Trol	21	15	71.43	
	Roadb	57	45	78.95	
	Ped	42	30	71.43	
	Mou	57	54	94.74	
Slow	Eng	40	33	82.50	84.05
	Tru	8	6	75.00	
	Trol	34	29	85.29	
	Roadb	55	43	78.18	
	Ped	60	49	81.67	
	Mou	60	56	93.33	

As shown in Figure 16, the point cloud frames at 1 s, 4 s, 7 s, and 10 s are selected to show the real vehicle detection results after traveling for 10 s in the three motion states, respectively, and 20 frames of the point cloud in the consecutive frames of 10 s are extracted to statistically measure the detection effect of the real vehicle, as shown in Table 9.

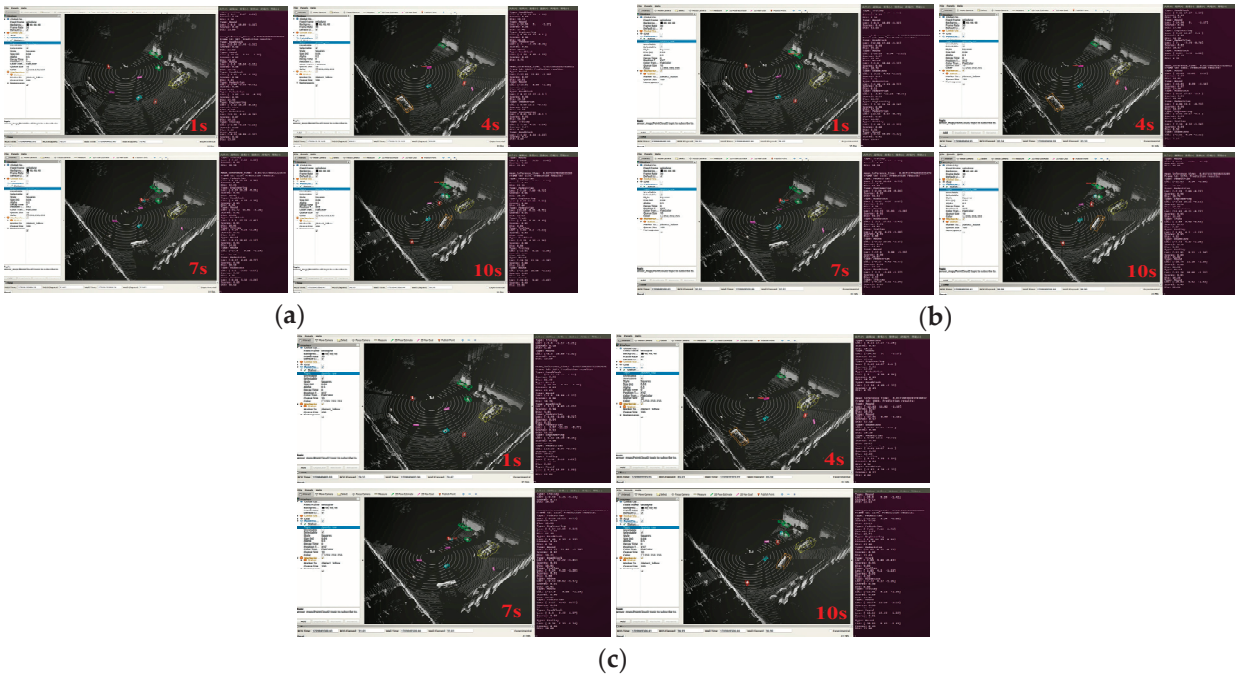


Figure 16. Real-time detection effect of different motion states: (**a**) real-time detection effect of straight-line driving.; (**b**) real-time detection effect of turn-by-turn driving; (**c**) cockpit rotation real-time detection effect.

The experimental results show that the target detection algorithm used in this study has more reliable detection performance in different speed modes and different motion states.

Table 9. Comparative statistics of target detection under different motion states.

Mode	Category	GT/pc	Det/pc	Rate (%)	Average (%)
Straight	Eng	40	31	77.50	87.80
	Tru	17	17	100.00	
	Trol	31	29	93.55	
	Roadb	53	42	79.25	
	Ped	53	46	86.80	
	Mou	60	58	96.67	
Turns	Eng	28	23	82.14	88.42
	Tru	20	16	80.00	
	Trol	32	29	90.63	
	Roadb	28	28	100.00	
	Ped	57	53	92.98	
	Mou	25	19	76.00	
Rotation	Eng	30	23	76.67	81.78
	Tru	20	16	80.00	
	Trol	40	30	75.00	
	Roadb	48	42	87.50	
	Ped	60	48	80.00	
	Mou	27	25	92.59	

5.2. Detection of Real-Time Validation

The theoretical scanning frame rate of LIDAR in the sensing system is 10 Hz. To further verify the real-time situation from the sensor input to the point cloud input, validation experiments are carried out for two separate cases.

(1) Increase point cloud input frame rate

In order to measure the limit detection frame rate of the detector, the frame rate of the point cloud input network is increased five-fold to measure the limit detection frame rate of the detector, as shown in Figure 17. The limit average detection frame rate of the detector in 20 s is 28.07 Hz, which is larger than the theoretical scanning frame rate of LiDAR.

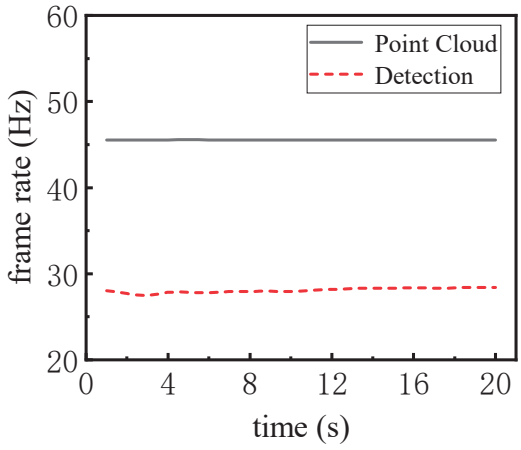


Figure 17. Comparison of point cloud input frame rate and detection frame rate.

(2) Real-time detection frame rate comparison

The theoretical scanning frame rate of LiDAR is 10 Hz, but the actual scanning to obtain the frame rate is only 90% of the theoretical value, so in order to verify the real-time situation of the actual operation of the detector, the real-vehicle point cloud input frame rate compared with the detection of the frame rate is shown in the Figure 18. It can be

seen that as the detector is just started, there is a certain delay but, after 5 s, it tends to be stabilized.

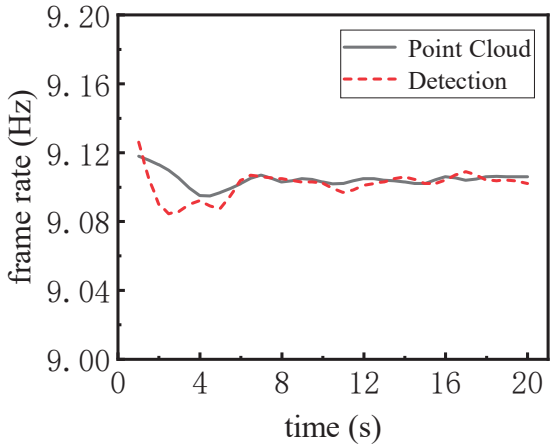


Figure 18. Comparison of real vehicle point cloud input frame rate and detection frame rate.

The detection performance of the network can be judged by the two ways of the above test to meet the real-time requirements of the motorized crawler excavator during the actual walking operation.

5.3. Security Alert Policy Validation

In order to verify the rationality and feasibility of the early warning braking strategy, the test is carried out. As shown in Figure 19, when the sensing system is running, a staff member is about to approach the excavator; according to the trajectory tracking, the system predicts that the target position of the next frame is in the range of the warning safety distance. At this time, the sensing system immediately sends out an early warning alert, suggesting that there are pedestrians in the non-safety zone, and sends the current position of the target and the predicted position of the next frame in real time to the visualization of the interactive interface.

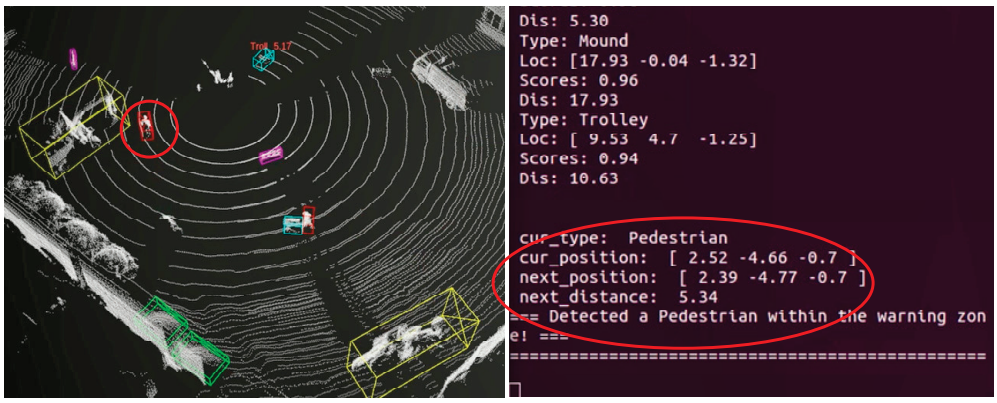


Figure 19. Pedestrian enters within warning threshold.

After the staff leaves the warning threshold range, the sensing system resumes normal operation. When the staff continues to drive into the warning threshold range of the roadblock, as shown in Figure 20, the sensing system sends the current position information of the roadblock as well as the relative predicted position of the next frame to the visualization interactive interface, and sends out a warning alert to inform that there is an obstacle in front of them.

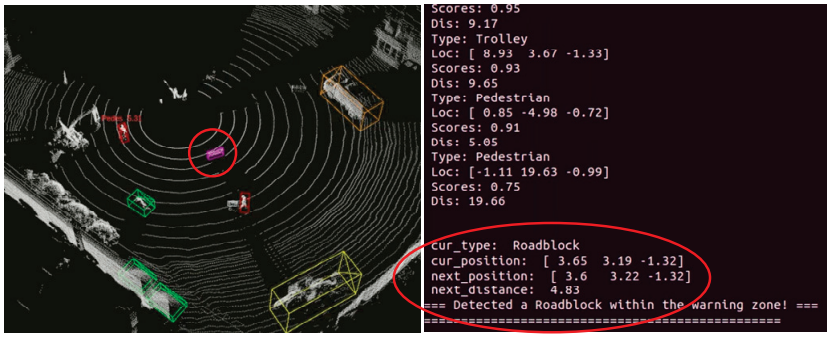


Figure 20. Entering within roadblock warning thresholds.

As shown in Figure 21, when the excavator enters the braking threshold range of the barricade, i.e., there is a possibility of collision at the next moment; at this time, the sensing system sends the zero duty cycle message to the proportional decompression valve through the way of CAN bus communication, so that the valve opening is closed, and the left and right track speeds are both set to zero, so as to achieve the purpose of emergency braking, and prevent the next moment collision from happening. The next moment collision occurs, to ensure the safety of the excavator itself and other objects in the vicinity.

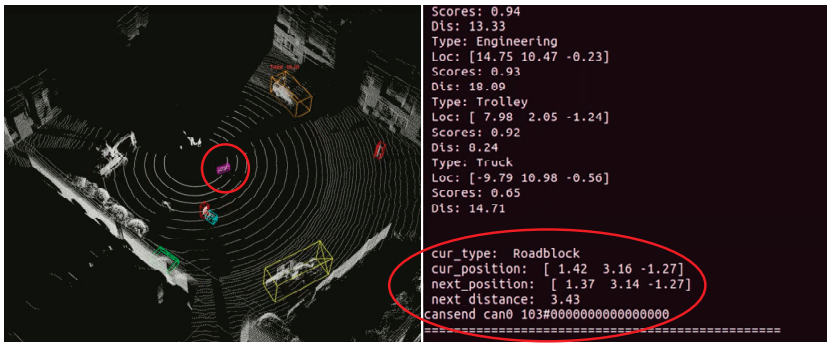


Figure 21. Entering the barricade braking threshold range.

6. Conclusions

Aiming at the unique challenges posed by unstructured scenes and special targets in construction machinery environments, and to address the scarcity of relevant open-source datasets, this paper constructed a dedicated data collection platform to develop a dataset customized for engineering scenarios. Focusing on the perception layer requirements for environmental sensing in these complex settings, we proposed a LiDAR-based system, which adopts the detection-based tracking paradigm. Based on the predicted trajectories of dynamic targets, a safety warning strategy was developed and demonstrated as a key application of the perception output. In order to verify the effectiveness of the proposed sensing system and dynamic target trajectory prediction, a test platform was built and experimental verifications were carried out. The experimental results demonstrated that the proposed system is better adapted to the specific needs of construction machinery applications, providing effective environmental perception capabilities in complex, unstructured environments. This enhanced perception supports the development of safer and more autonomous operations for construction machinery, by improving situational awareness for operators and enabling functionalities like dynamic obstacle avoidance verification. Most importantly, this work, implemented within the ROS framework, explicitly focuses on attempts to apply perceptual algorithms rather than normative system architecture

that fulfill standard safety requirements, and hopefully provide direction for subsequent research related to the automation of construction machinery.

Author Contributions: Conceptualization, J.C. and Q.C. (Qipeng Cai); methodology, X.H. and Q.C. (Qihuai Chen); writing—review and editing, T.L.; project administration, H.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data is available upon reasonable request.

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

References

1. Eraliev, O.M.U.; Lee, K.H.; Shin, D.Y.; Lee, C.H. Sensing, sensing, decision, planning and action of autonomous excavators. *Autom. Constr.* **2022**, *141*, 104428. [CrossRef]
2. Assadzadeh, A.; Arashpour, M.; Brilakis, I.; Ngo, T.; Konstantinou, E. Vision-based excavator pose estimation using synthetically generated datasets with domain randomization. *Autom. Constr.* **2022**, *134*, 104089. [CrossRef]
3. Zhou, Y.; Tuzel, O. Voxelnet: End-to-end learning for point cloud based 3d object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4490–4499.
4. Yan, Y.; Mao, Y.; Li, B. Second: Sparsely embedded convolutional detection. *Sensors* **2018**, *18*, 3337. [CrossRef] [PubMed]
5. Lang, A.H.; Vora, S.; Caesar, H.; Zhou, L.; Yang, J.; Beijbom, O. Pointpillars: Fast encoders for object detection from point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 12697–12705.
6. Yin, T.; Zhou, X.; Krahenbuhl, P. Center-based 3d object detection and tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 11784–11793.
7. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6569–6578.
8. *ISO 13849-1:2023*; Safety of Machinery—Safety-Related Parts of Control Systems—Part 1: General principles for design. International Organization for Standardization: Geneva, Switzerland, 2023.
9. *IEC 62061:2021*; Safety of Machinery—Functional Safety of Safety-Related Electrical, Electronic and Programmable Electronic Control Systems. International Electrotechnical Commission: Geneva, Switzerland, 2021.
10. *Directive 2006/42/EC of the European Parliament and of the Council of *on Machinery, and Amending Directive 95/16/EC**; Official Journal of the European Union: Brussels, Belgium, 17 May 2006.
11. Luo, W.; Yang, B.; Urtasun, R. Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3569–3577.
12. Weng, X.; Wang, J.; Held, D.; Kitani, K. 3d multi-object tracking: A baseline and new evaluation metrics. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 10359–10366.
13. Gu, S.; Yang, L.; Du, Y.; Chen, G.; Walter, F.; Wang, J.; Knoll, A. A review of safe reinforcement learning: Methods, theories and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 11216–11235. [CrossRef] [PubMed]
14. Lin, K.; Li, D.; Li, Y.; Chen, S.; Wu, X. FHCPL: An Intelligent Fixed-Horizon Constrained Policy Learning System for Risk-Sensitive Industrial Scenario. *IEEE Trans. Ind. Inform.* **2024**, *20*, 5794–5804. [CrossRef]
15. Lin, K.; Li, D.; Li, Y.; Chen, S.; Liu, Q.; Gao, J.; Jin, Y.; Gong, L. TAG: Teacher-Advice Mechanism with Gaussian Process for Reinforcement Learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 12419–12433. [CrossRef] [PubMed]
16. Zheng, Y.; Li, J.; Yu, D.; Yang, Y.; Li, S.E.; Zhan, X.; Liu, J. Safe offline reinforcement learning with feasibility-guided diffusion model. *arXiv* **2024**, arXiv:2401.10700.
17. Rivera, S.; Iannillo, A.K.; Lagraa, S.; Joly, C.; State, R. ROS-FM: Fast monitoring for the robotic operating system (ROS). In Proceedings of the 2020 25th International Conference on Engineering of Complex Computer Systems (ICECCS), Singapore, 28–31 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 187–196.

18. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]
19. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
20. He, X.; Cheng, J. Revisiting L1 loss in super-resolution: A probabilistic view and beyond. *arXiv* **2022**, arXiv:220110084.
21. Simonelli, A.; Bulo, S.R.; Porzi, L.; López-Antequera, M.; Kontschieder, P. Disentangling monocular 3d object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1991–1999.
22. Chen, Y.; Liu, J.; Zhang, X.; Qi, X.; Jia, J. Voxelnext: Fully sparse voxelnet for 3d object detection and tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 20–22 June 2023; pp. 21674–21683.
23. Shi, J.; Wang, X.; Fu, Y. Research on longitudinal collision hierarchical warning of vehicles based on Telematics. *J. Chongqing Jiaotong Univ. Nat. Sci. Ed.* **2020**, *39*, 1–7.
24. Wang, Z.; Zang, L.; Jiao, J.; Mao, Y. Research on hierarchical control strategy of automatic emergency braking system. *World Electr. Veh. J.* **2023**, *14*, 97. [CrossRef]
25. Gounis, K.; Bassiliades, N. In telligent momentary assisted control for autonomous emergency braking. *Simul. Model. Pract. Theory* **2022**, *115*, 102450. [CrossRef]
26. Park, J.; Kim, D.; Huh, K. Emergency collision avoidance by steering in critical situations. *Int. J. Automot. Technol.* **2021**, *22*, 173–184. [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.

sensors

MDPI

Article

Propagation Regimes and Signal Enhancement Mechanisms of Collinear Double-Pulse Plasma with Varying Inter-Pulse Delays

Yang Zhao [1], Lei Zhang [2,3,*], Zhihui Tian [4], Xiuqing Zhang [1], Jiandong Bai [1] and Wangbao Yin [2,3]

[1] School of Semiconductor and Physics, North University of China, Taiyuan 030051, China; yangzhao@nuc.edu.cn (Y.Z.); zhangxiuqing@nuc.edu.cn (X.Z.); jdbai@nuc.edu.cn (J.B.)
[2] State Key Laboratory of Quantum Optics Technologies and Devices, Institute of Laser Spectroscopy, Shanxi University, Taiyuan 030006, China; ywb65@sxu.edu.cn
[3] Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, China
[4] School of Physics and Electronic Information Engineering, Hubei Engineering University, Xiaogan 432000, China; tianzh@hbeu.edu.cn
* Correspondence: k1226@sxu.edu.cn

Abstract: Laser-induced breakdown spectroscopy (LIBS) is an in situ analytical technique. Compared to traditional single-pulse LIBS (SP-LIBS), collinear double-pulse LIBS (DP-LIBS) is a promising technique due to its lower limit of detection for trace elements. In this paper, we analyze the spectral and image information obtained from the emissions emitted by single/double pulse (SP/DP) laser-induced plasmas. The types of laser-supported absorption (LSA) waves of the plasmas were determined according to the interactions among the ablation vapor, the ambient gas, and the laser. Furthermore, the influence mechanisms of plasma shielding on DP-LIBS signal intensity enhancement with different inter-pulse delay were investigated. In our experimental conditions, the propagation regime of SP plasma is a laser-supported combustion (LSC) wave. The DP plasmas with short inter-pulse delays show the characteristics of a laser-supported detonation (LSD) wave, and the enhancement mechanism is mainly reheating for pre-plasma. On the contrary, the DP plasmas with longer inter-pulse delays show the characteristics of a LSC wave, and the increase in laser ablation is a major contributing factor to the signal improvement. In addition, the spectral lines, which are difficult to excite by SP-LIBS, can be obtained by selecting an appropriate inter-pulse delay and setting a short delay, which provides a new idea for the measurement of trace elements.

Keywords: collinear double-pulse laser-induced plasma; collinear double-pulse laser-induced breakdown spectroscopy (DP-LIBS); laser-supported absorption (LSA) wave; signal enhancement mechanisms

1. Introduction

Laser-induced breakdown spectroscopy (LIBS) is a developing analytical technique. A high-energy laser is used to ablate minute quantities of materials from the target, and the spectra emitted from the resulting high-temperature plasmas are analyzed to obtain the elemental composition of the target. LIBS is particularly suitable for industrial process monitoring and product quality control due to the advantages of high speed, little or no sample preparation, and multi-element analysis [1–4]. However, the detection accuracy and sensitivity of LIBS for trace elements have not yet met the requirements of certain industries, necessitating further advancements [5,6].

In recent years, many researchers have devoted a lot of energy to this issue and proposed a variety of signal enhancement schemes, such as double-pulse (DP) excita-

tion, picosecond and femtosecond pulse excitation, spark discharge-assisted excitation, microwave-assisted excitation, spatial confinement, and magnetic confinement [7–10]. Among these methods, DP-LIBS has emerged as a highly effective approach, as it uses two laser pulses (with inter-pulse delays ranging from hundreds of nanoseconds to a few microseconds) to more efficiently deliver laser energy to the target and plasma. DP-LIBS is a hotspot research area due to some of its merits, including an improved signal-to-noise ratio [11], increases in signal intensity by factors of two to thirty-three [12–15], reductions in the limit of detection (LOD) by almost tenfold [16–18], and substantial enhancements in quantitative analysis performance [13].

The generally acknowledged interpretations of the enhancement mechanisms of DP-LIBS are the following: 1. Due to the plasma shielding, the pre-plasma generated by the pre-pulse absorbs the energy from the reheating pulse. 2. The shockwave generated by the pre-pulse reduces ambient gas density, creating a low-pressure region at the target surface. Following the second excitation, the reheated plasma expands rapidly in this rarefied region according to fluid dynamics, resulting in a larger plume volume that reduces plasma shielding and increases ablation mass [19–22]. Additionally, Pasquini et al. [23] pointed out that, in addition to decreasing the breakdown threshold, effective laser–plasma coupling is also a crucial factor for enhancing DP-LIBS performance. Kepes et al. [24] found that reheating during ablation significantly improves the spatial homogeneity of the plasma temperature and electron density.

In general, most DP-LIBS research primarily uses larger signal acquisition delays and gate widths (>500 ns) [25]. However, understanding the structure and morphology of plasmas in the early stage is essential for comprehending their entire propagation and evolution. Furthermore, only a limited number of studies have combined the propagation regimes of laser-supported absorption (LSA) waves and species distribution to elucidate the enhancement mechanisms of DP-LIBS. The propagation mechanisms of plasma can be divided into two primary regimes: laser-supported combustion (LSC) waves and laser-supported detonation (LSD) waves. This classification depends on whether the laser energy is mainly absorbed by vapor or shocked gas. The shielding effect of plasma for LSD waves exists in the shock gas at the forefront of plasma, where the plasma exhibits a mixed structure. For LSC waves, the shielding effect occurs in the ablation vapor at the center of the plasma, resulting in a layered structure of plasma.

Based on the propagation direction and the temporal sequence of the laser pulses, the DP-LIBS can be categorized into three types: collinear, orthogonal pre-ablation, and orthogonal reheating. Among these schemes, the collinear DP-LIBS is more suitable for industrial applications due to its simple configurations and adjustability. In this study, the emission spectra and species emissivity images of collinear SP/DP plasmas with short delay times and narrow gates were obtained, and the interactions between vapor and ambient gases were analyzed. The purpose is to elucidate the propagation characteristics of the plasma and investigate the physical principles and signal enhancement mechanisms of collinear DP-LIBS. A novel idea is proposed to decrease the limit of detection for trace elements while providing theoretical and technological support for the application of DP-LIBS in online quality monitoring of industrial processes.

2. Experimental Setup

The DP-LIBS experimental setup, as shown in Figure 1, mainly consisted of three parts: the plasma excitation system, the dual-wavelength differential spectroscopic imaging system, and the spectra acquisition system.

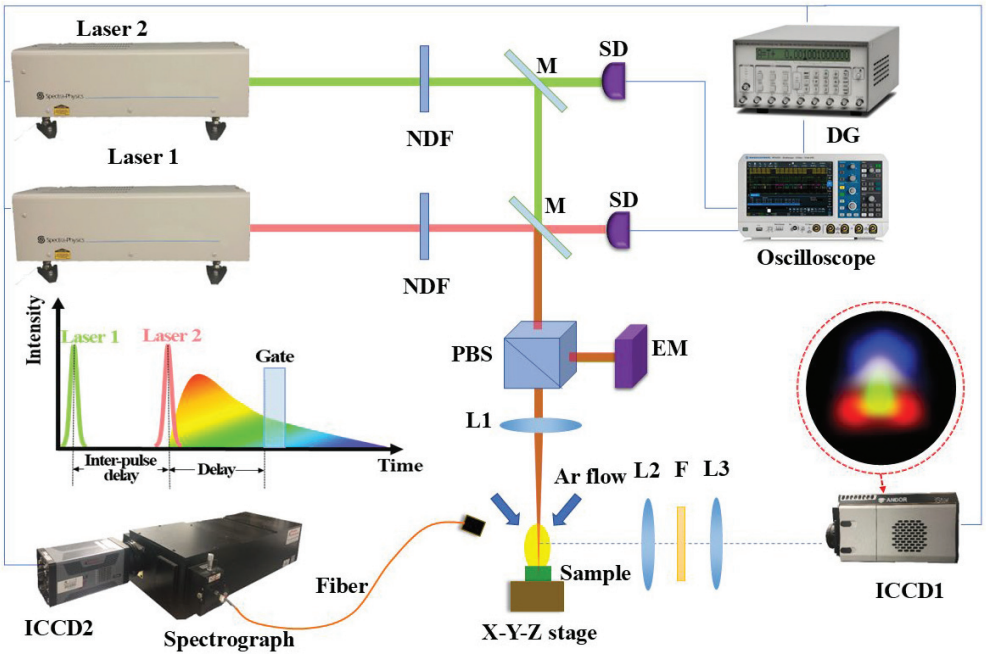


Figure 1. Schematic of the experimental DP-LIBS setup. Components include NDF (neutral density filters), M (mirror), PBS (polarizing beam splitter), EM (energy meter), DG (digital delay generator), SD (silicon detectors), ICCD (intensified CCD), L1, L2, L3, and L4 (lenses), and F (bandpass filter).

The excitation source for DP-LIBS was two high-energy pulsed lasers (Sunnyvale, CA, USA, Spectra Physics, INDIHG-20, wavelength: 1064 nm, pulse duration: 7 ns, and repetition frequency: 20 Hz). The digital delay generator (Sunnyvale, CA, USA, Stanford Research System, DG535) was used to synchronize two lasers and adjust inter-pulse delays, and the actual inter-pulse delays were monitored by two silicon detectors (Newton, NJ, USA, Thorlabs, DET10A/M) in real time. The neutral density filters (NDF) were used to adjust the laser energy to ensure that the pulse spots were evenly distributed. In order to monitor the incident laser energy, the fraction of the laser reflected by the polarizing beam splitter (PBS) was measured using an energy meter (Jerusalem, Israel, Ophir, Nova II) that was perfectly calibrated. A lens (L1) with a focal length of 50 mm was used for focusing the laser beams below the sample surface, and the spot size was about 0.6 mm. The sample was mounted on an X-Y-Z stage so that each laser pulse hit a different spot. To make certain that the plasma expanded in pure argon at atmospheric pressure, two gas tubes were positioned above the sample surface, and the argon continually blew close to the laser focus. This excitation system was capable of emitting laser in two modes. The single-pulse mode used only laser 1 with an irradiance of 2 GW/cm^2. The collinear dual-pulse mode utilized laser 1 and laser 2 with irradiances of 1 GW/cm^2 and 1 GW/cm^2, respectively. The timing and sequence of the laser pulses are illustrated in Figure 1, with inter-pulse delays between the two pulses ranging from 200 ns to 15000 ns.

The dual-wavelength differential spectroscopic imaging system was mounted on the right side of the plasma and perpendicular to the laser incident axis. Please refer to the published papers for more detailed information [26,27]. In summary, an ICCD1 camera (Belfast, United Kingdom, Andor, iStar DH334T-18U-03) was used to obtain the images of a certain species distribution in plasma through a 4F optical system composed of two lenses (L2 and L3) and a pair of bandpass filters (F1 and F2) with bandwidths of 6–10 nm. The center wavelength of F1 corresponded to the chosen emission line of the species, while the center wavelength of F2 was closely related to F1, and there were no lines in F2's passband. The ICCD1 recorded emission intensity images of the species through F1 while recording the continuous background using F2. In order to improve the signal-to-noise

ratio and reduce the fluctuation, the specific ICCD gains (see Table 1) were set for each species, and each image accumulated 60 plasmas. The obtained images were processed by transmittance correction, continuous background subtraction, and Abel inversion. Finally, the emissivity images of the species were obtained. This system can visually display the evolution process of various species in the plasma and directly obtain the morphology and structure of the plasma.

Table 1. The filter pairs used for the observation of Al and Ar atomic and ionic species. Additionally included are the ICCD gains for each species under different laser excitation modes.

Species	Emission Line (nm)	Center Wavelength of Filter F1 (nm)	Center Wavelength of Filter F2 (nm)	ICCD Gain	
				SP Plasma	DP Plasma
Al I	394.40	396	365	300	200
	396.15				
Al II	358.66	358	365	300	200
Ar I	763.51	764	786	500	300
Ar II	484.78	488	530	500	300

The spectra acquisition system was positioned on the left side of the plasma. The fluorescence was transmitted to the grating spectrograph (Trenton, NJ, USA, Princeton Instruments, SP-2750) equipped with ICCD2 (Trenton, NJ, USA, Princeton Instruments, PI-MAX4-1024i) through an all-silica optical fiber at a 45-degree angle. The time-resolved spectra can be obtained by adjusting the delay and gate width of the ICCD2. A low-pressure Hg lamp (Irvine, CA, USA, Newport, 6048) calibrated the spectrograph's wavelength, while a deuterium-halogen light source (Apeldoorn, Netherlands, Avantes, AvaLightDH-S-BAL) calibrated the intensity of spectra.

The materials used in this experiment were pure aluminum ($\geq$99.99%), with the surface polished using sandpaper before each measurement set. The spectral lines and center wavelengths of the filters used for imaging are listed in Table 1. Among them, the Al II line is a superposition of 15 lines, and Al II 358.66 nm listed in Table 1 is the one with the highest transition probability.

3. Results

3.1. Comparison of SP and DP LIBS Spectra

3.1.1. Effect of Spectral Acquisition Delays on Spectra

For DP-LIBS, only a few experiments have been conducted with short delay times and narrow gates. As a result, the spectral acquisition delays and gate widths of ICCD2 were set to 100 ns and 5 ns, respectively, in addition to the typical 1000 ns and 500 ns. Under the above acquisition conditions, the emission spectra of SP Al plasma (horizontal coordinate of zero) and DP Al plasma (inter-pulse delay 200 ns, 500 ns, 700 ns, 1500 ns, 2000 ns, 3000 ns, and 6000 ns) in the range of 340–420 nm are shown in Figure 2a. Figure 2b shows a magnified view of Figure 2a in the range of 350–370 nm, and the stacked lines are separated for clarity, so there is no vertical scale. Comparing the spectra at delays of 100 ns and 1000 ns, it can be found that the continuum background at 100 ns is significantly stronger, resulting in wider full width at half maximums (FWHMs) of the spectral lines. Notably, the spectral lines of doubly ionized Al species—Al III at 360.16 nm and 361.24 nm—are detectable at 100 ns, regardless of the method (SP or DP).

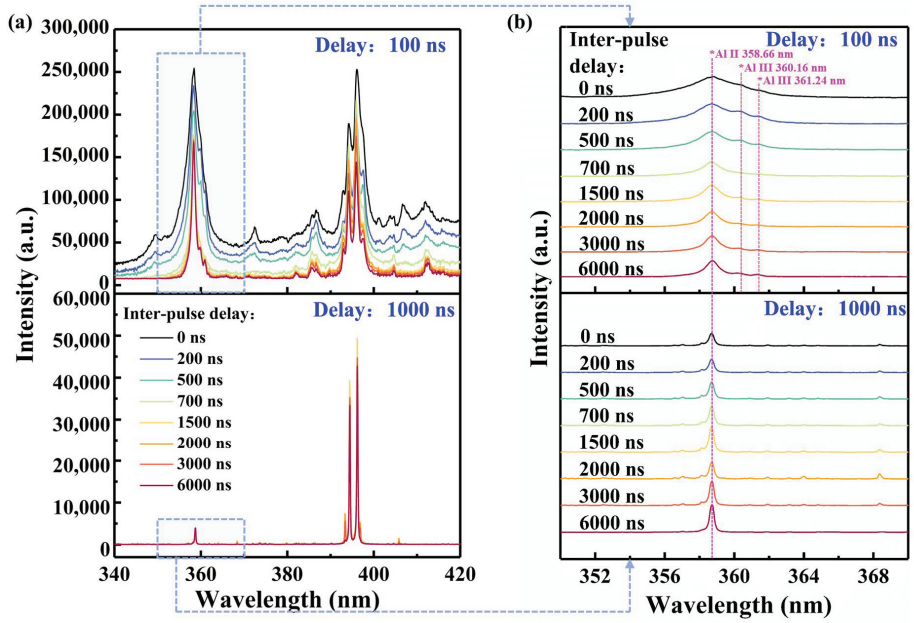


Figure 2. (**a**) Emission spectra of SP aluminum plasma (horizontal coordinate of zero) and DP aluminum plasma (inter-pulse delays of 200 ns, 500 ns, 700 ns, 1500 ns, 2000 ns, 3000 ns, and 6000 ns) in Ar ambient gas at delays of 100 ns and 1000 ns, corresponding to gate widths of 5 ns and 500 ns. (**b**) A magnified view of (**a**) in the range of 350–370 nm.

The electron number density *Ne* is the cause of the aforementioned phenomenon. The *Ne* was calculated by [28]

$$Ne(\mathrm{cm}^{-3}) = 10^{17}[\Delta\lambda_{1/2}(\mathrm{nm})/0.549]^{1.4713} \quad (1)$$

where $\Delta\lambda_{1/2}$ represents the FWHM of the Stark-broadened Hα line at 656.28 nm. The curves of *Ne* as a function of inter-pulse delays at 100 ns and 1000 ns are shown in Figure 3b. It can be seen that the *Ne* of the initial stage of plasma expansion is high due to a stronger continuous background brought on by Bremsstrahlung. This increased ionization allows the detection of the doubly ionized lines, although these lines have a very short lifetime and are not detectable at 200–300 ns.

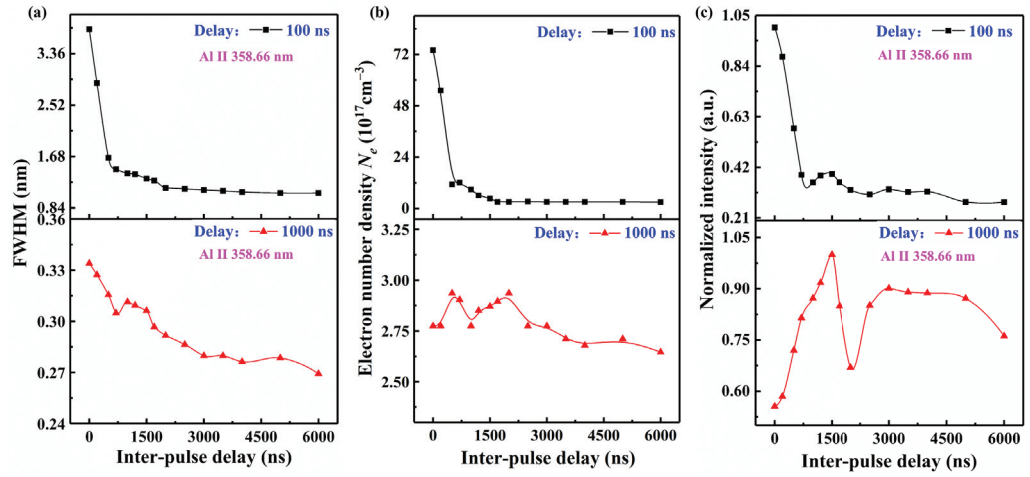


Figure 3. (**a**) FWHM of the Al II 358.66 nm spectral line emitted from SP and DP plasmas at inter-pulse delays of 100 ns (gate width: 5 ns) and 1000 ns (gate width: 500 ns) in an argon ambient gas environment. (**b**) *Ne* as a function of inter-pulse delay under the same experimental conditions. (**c**) Normalized signal intensity of Al II 358.66 nm plotted against inter-pulse delays, highlighting trends for SP and DP plasmas at 100 ns and 1000 ns acquisition delays.

3.1.2. Effect of Inter-Pulse Delays on Spectra

The spectra of SP and DP plasmas with varying inter-pulse delays were compared at the same delay. Compared to DP-LIBS, the spectra of SP-LIBS show the highest continuous background at a delay of 100 ns. The FWHM of Al II 358.66 nm is 3.76 nm for SP-LIBS. In contrast, the Al II 358.66 nm of DP-LIBS has a lower FWHM of 1.08–2.88 nm, and decreases with the inter-pulse delay. At a delay of 1000 ns, both SP and DP plasmas show a similarly low level of continuous background, and the FWHM of Al II 358.66 nm fluctuates within a narrow range of 0.27 to 0.33 nm. By comparing the intensity of spectral lines emitted from SP and DP plasma, we found maximum enhancement factors of 1.8 for Al II 358.66 nm, 1.6 for Al I 396.15 nm, and 1.4 for Al I 394.40 nm.

The effect of the inter-pulse delays on the continuous background and FWHM of the spectral lines is primarily due to the short inter-pulse delay, during which the second pulse interacts with the pre-plasma characterized by a high *Ne*. A portion of the energy from the second pulse is absorbed by free electrons, further increasing *Ne* through collisions between electrons and neutral or singly ionized atoms. Consequently, the *Ne* of plasma with the shorter inter-pulse delay is higher at 100 ns, as well as having a larger continuous background and wider FWHM. However, as the inter-pulse delays exceed 2000 ns, these influences decrease significantly. By contrast, the *Ne* drops to the same level at the delay of 1000 ns, resulting in minimal variation in the continuum background.

Figure 3c depicts the variation in the normalized intensity of Al II 358.66 nm as a function of inter-pulse delay. The intensity of Al II emitted from the SP plasma (with an inter-pulse delay of 0 ns) is notably lower than that emitted from the DP plasma at 1000 ns. However, the opposite trend is observed at 100 ns due to the larger continuous background in the SP plasma. The influence of the inter-pulse delays of DP plasma on the intensity of Al II was further investigated. At a 100 ns delay, the intensity of Al II for inter-pulse delays ranging from 100 to 700 ns is also significantly affected by the larger continuous background. After excluding these data points, the trends of the curves at delays of 100 ns and 1000 ns become consistent, both showing two peaks at inter-pulse delays of 1500 ns and 3000 ns. The underlying mechanisms for these observations will be discussed in Section 3.4, in conjunction with plasma imaging analysis. It is notable that the trends of the FWHM and normalized intensity of other Al II lines (e.g., Al II 466.31 nm) are similar to those of Al II 358.66 nm. The results derived from our analysis of Al II 358.66 nm can be extended to other Al II lines.

The effect of inter-pulse delays on the Al III 360.16 nm and 361.24 nm spectral lines is illustrated in Figure 4. Due to the strong continuous background of SP plasma, the signal-to-noise ratios (SNRs) of these doubly ionized lines of SP-LIBS are very low. However, in DP-LIBS, the SNRs improve significantly as the inter-pulse delay increases to 4000–5000 ns. In summary, DP-LIBS demonstrates the ability to excite spectral lines—such as highly ionized lines and trace element lines—that are typically inaccessible using conventional SP-LIBS. The aforementioned characteristics of DP-LIBS will offer a novel LIBS scheme for measuring trace elements: selecting a suitable inter-pulse delay to excite the plasma and setting a short delay time for spectral acquisition. The spectra with a low continuum background and a strong intensity of the spectral lines can be measured.

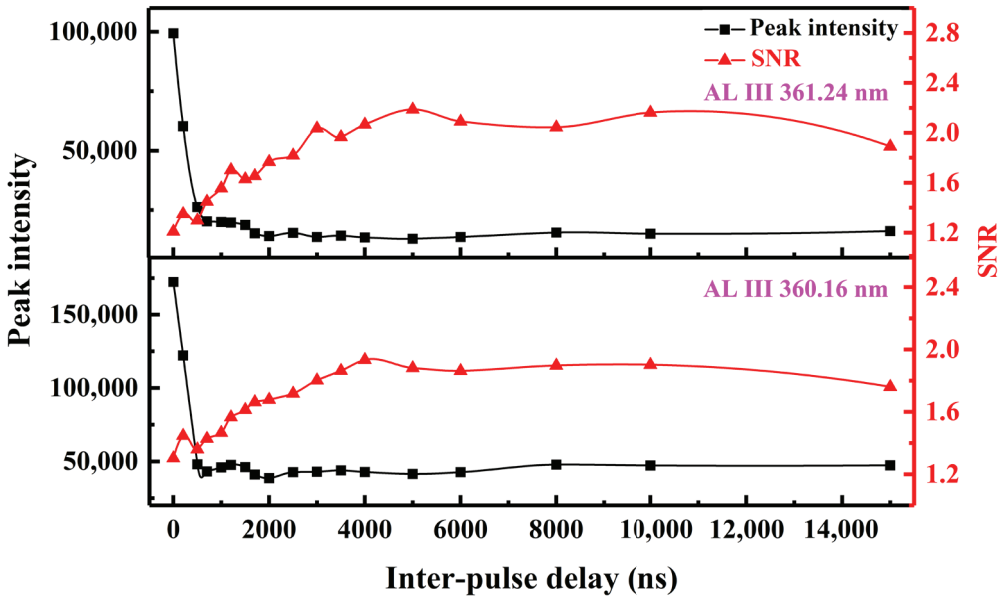


Figure 4. Peak intensity and signal to noise ratio (SNR) of the Al III 360.16 nm and 361.24 nm spectral lines as a function of inter-pulse delays.

3.2. Comparison of Space-Time Evolution of SP and DP Plasma

3.2.1. Morphology and Structure of SP Plasma

As shown in Figure 5, the dual-wavelength differential spectroscopy was used to obtain emissivity images of species with different colors: blue for argon atoms (Ar I), gray for argon ions (Ar II), red for aluminum atoms (Al I), and green for aluminum ions (Al II). To better represent the internal structure of the plasma, each emissivity image was normalized by its maximum value.

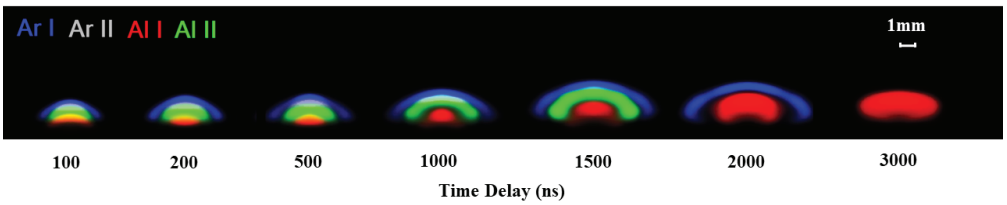


Figure 5. Normalized emissivity images of various species within the SP plasma at different delays. The species are represented by different colors: blue for argon atoms (Ar I), gray for argon ions (Ar II), red for aluminum atoms (Al I), and green for aluminum ions (Al II). The delay times investigated include 100 ns, 200 ns, 500 ns, 1000 ns, 1500 ns, 2000 ns, and 3000 ns, with corresponding gate widths of 5 ns, 7 ns, 10 ns, 50 ns, 100 ns, 150 ns, and 200 ns, respectively.

In this ablation mechanism, the plasma including all emitted species exhibits a hemispherical shape and a distinct layer structure, with the order of argon atoms, argon ions, aluminum ions, and aluminum atoms along the laser incidence direction (i.e., the symmetry axis of the plasma in the figure from top to bottom). The layer of argon ions covers the vapor plume, composed of aluminum ions and atoms, while the outermost layer of argon atoms envelopes the entire plume.

The ordered arrangement of species layer indicates a region of highest temperature exists between the excited argon atom layer and the front of the aluminum vapor. The global expansion and cooling of the aluminum vapor is more determined by fluid dynamics for the longer delay. The thin layer of argon ions dissipates after 1000 ns, and the aluminum ions vanish after 1500 ns.

As shown in Figure 6, the expansion radii of the plasma were measured by plotting two concentric semicircles. The centers of both semicircles aligned with the center of the laser spot, which corresponded to the location of the plasma center on the sample surface. *Ra* is the axial radius of the plasma, which accounts for its axial expansion, while *Rr* represents half of the maximum radial expansion range, reflecting radial expansion of the plasma. The *Ra* and *Rr* of SP and DP plasma, alongside the ratio of *Ra* to *Rr* are displayed at a delay of 100 ns in Figure 6. The representative values for analyzing the axial and radial expansion of plasmas are provided under different laser excitation conditions. For SP Al (horizontal coordinate of zero) at 100 ns, the measured values are *Ra* = 0.807 mm, *Rr* = 0.978 mm, and the ratio of *Ra* to *Rr* is 0.83. These results indicate that the SP plasma expands more quickly in the radial direction.

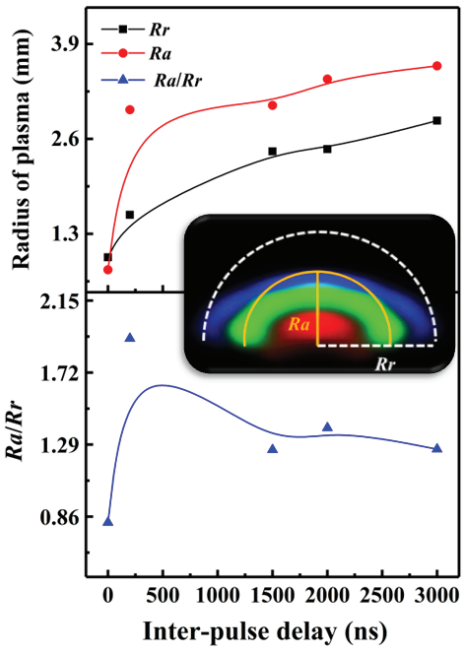


Figure 6. The axial expansion radius *Ra*, radial expansion radius *Rr*, and the ratio of *Ra* to *Rr* for SP plasma (horizontal coordinate of zero) and DP plasma (inter-pulse delay 200 ns, 1500 ns, 2000 ns, and 3000 ns) in argon ambient gas environment. The measurements were taken at a delay of 100 ns with a gate width of 5 ns.

3.2.2. Morphology and Structure of DP Plasma

The emissivity images of DP plasma with the inter-pulse delays of 200 ns, 1500 ns, 2000 ns, and 3000 ns (selected from Figure 3c) are shown in Figure 7. Similar to the SP plasma, the DP plasma exhibits a hemispherical shape. However, as shown in Figure 6, the *Ra* of the DP plasma can reach up to 4.5 times larger than that of the SP plasma, indicating significantly enhanced axial expansion. Both the *Ra* and *Rr* of the DP plasma increase with the inter-pulse delays, whereas the ratios of *Ra* to *Rr* decrease. The *Ra* is larger than the *Rr* for the DP plasma. Therefore, the DP plasma expands more quickly in the axial direction.

There are differences in the internal structure of the DP plasma with different inter-pulse delays. The first row of Figure 7 shows the emissivity images of DP plasma with a short inter-pulse delay of 200 ns. The argon atoms are completely ionized at the delay of 100 ns, and the argon ions overlap with the aluminum ions (creating a bright white region in Figure 7 due to the mixture of gray and green). This overlap results in the formation of a thick mixed zone, located behind the shockwave front. This structure indicates that the mixed thick zone with a high temperature is directly in contact with the undisturbed ambient gas, and the argon ions persist for a prolonged period. As the delay increases to

200 ns, the argon atoms are generated from a region near the target surface and gradually envelope the plume core.

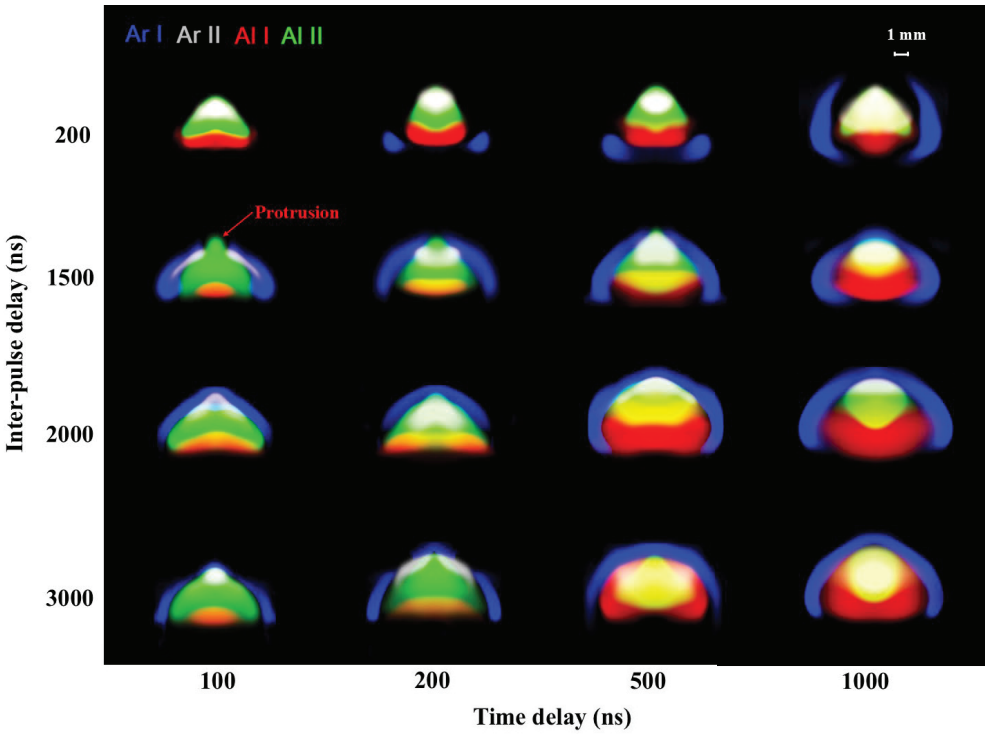


Figure 7. Emissivity images of species in DP plasma at inter-pulse delays of 200 ns, 1500 ns, 2000 ns, and 3000 ns. The species are represented by different colors: blue for argon atoms (Ar I), gray for argon ions (Ar II), red for aluminum atoms (Al I), and green for aluminum ions (Al II). The images were obtained at delay times of 100 ns, 200 ns, 500 ns, 1000 ns, 1500 ns, 2000 ns, and 3000 ns, with corresponding gate widths of 5 ns, 7 ns, 10 ns, 50 ns, 100 ns, 150 ns, and 200 ns, respectively.

The internal structure of DP plasma with inter-pulse delays of 1500 ns, 2000 ns, and 3000 ns is similar to that of SP plasma, exhibiting an obvious layer structure. In these cases, the argon ion layer and aluminum ion layer do not overlap or have only a small overlap region. However, a notable difference appears for DP plasma with the inter-pulse delay of 1500 ns: the axial expansion velocity of aluminum ions exceeds that of argon ions at the delay of 100 ns, causing the aluminum ions to breach the layer of argon ions and form a small "protrusion" on its originally hemispherical shape. The detailed explanation for this phenomenon will be provided in Section 3.4, combined with the propagation types of SP and DP plasma.

3.3. Propagation Types of SP and DP Plasma

In references [29–31] and our previous works [26,27,32], the experimentally observed propagation regimes of SP and DP plasmas were discussed, and the results are summarized in Table 2. SP plasma exhibits a distinct layered structure, and the degree of ionization in the shock gas layer is relatively low. Because of the low density and temperature of argon, the laser penetrates through the shocked gas layer and deposits on the ablation vapor plume. The excitation and ionization of argon are not directly caused by laser energy absorption, but by the interaction between argon and the heated vapor plasma core, such as compression, heat conduction, and radiation transfer. Due to the low density and temperature of argon, the laser penetrates through the shocked gas layer and deposits energy in the ablation vapor plume. Therefore, the laser absorption region is located in the middle of the plasma. These characteristics of SP plasma are in accordance with LSC wave.

Table 2. Propagation regimes and enhancement mechanisms of SP plasma and DP plasma with different inter-pulse delays.

	SP Plasma	DP Plasma with Different Inter-Pulse Delays			
		200 ns	1500 ns	2000 ns	3000 ns
Propagation regime	LSC wave	LSD wave	LSC wave	LSC wave	LSC wave
Enhancement mechanism	-	Reheating	Reheating	Reheating	Increase in ablation mass

For DP plasma with an inter-pulse delay of 200 ns, the shocked gas layer is widely distributed and exhibits a high degree of ionization. This is because the shocked gas layer of pre-plasma has high density and temperature during the brief expansion time (200 ns). There is strong coupling between the second laser pulse (laser 2) and the shocked gas layer of pre-plasma, leading to the direct absorption of a significant amount of laser energy. At this stage, the shocked gas layer not only propagates forward but also recoils backward, forming a mixed region with the ablation vapor. The laser-absorbing zone, which is located at the top of the plasma, is the driving force of the expansion process in the direction of the laser incidence, and the plasma appears elongated for this reason. In summary, the plasma propagation regime belongs to the LSD wave.

The internal layered structures of DP plasmas with longer inter-pulse delays (1500 ns, 2000 ns, and 3000 ns) are similar to those of SP plasma. This similarity arises because the prolonged expansion time of the pre-plasma reduces the density of the shocked gas layer. As a result, the ablation vapor becomes the primary region for laser energy absorption again, exhibiting characteristics consistent with the LSC wave propagation regime.

3.4. Signal Enhancement Mechanisms of DP-LIBS

The signal enhancement of the DP-LIBS can be attributed to two primary reasons: 1. due to the plasma shielding effect, the pre-plasma absorbs the energy of the second laser pulse (laser 2), leading to a re-heating process of the plasma; 2. the first laser (laser 1) improved environmental conditions for the laser 2. According to shadowgraph studies [33,34], the ambient gas is compressed into a thin layer behind the shockwave front after the initial plasma generation. The internal pressure in the plume decreases as the shockwave expands outwards from the target, creating more diluted conditions for the plasma induced by laser 2. Additionally, a stronger coupling between the laser 2 and the target leads to a higher ablation mass of material.

Referring to the intensity of the Al II line at a delay of 1000 ns (shown in Figure 3c), the reheating effect is more effective for shorter inter-pulse delays (200 ns and 1500 ns) due to the higher density of the pre-plasma. Consequently, the pre-plasma reheating is the main enhancement mechanism.

Specifically, for the inter-pulse delay of 200 ns, the diminished signal enhancement of the Al II line is owing to the strong coupling between laser 2 and the shocked gas layer of the pre-plasma. In contrast, for the inter-pulse delay of 1500 ns, the axial expansion speed of aluminum ions exceeds that of argon ions, indicating a stronger interaction between laser 2 and the ablation vapor of the pre-plasma. The argon ion layer is disrupted by the aluminum ions, forming a "protrusion" in the emissivity images, and the Al II line is greatly intensified as a result.

As the inter-pulse delay increases to 2000 ns, the enhancement mechanism remains dominated by pre-plasma reheating. However, a decrease in pre-plasma density reduces

the reheating efficiency of ablation vapor and limits the intensity enhancement of the Al II line, leading to a dip for 2000 ns as observed in Figure 3c.

With an even longer inter-pulse delay (3000 ns), the pre-plasma density further decreases, reducing its shielding effect on laser 2. The extremely rarefied ambient conditions enhance the laser–target interaction, making the increase in ablation mass the primary enhancement mechanism. As shown in Figure 3c, the intensity of the Al II line emitted from DP plasma with a 1500 ns inter-pulse delay is stronger than that with a 3000 ns inter-pulse delay. Thus, as an enhancement mechanism, plasma reheating is more effective than the increase in the ablation mass under the experimental conditions. The enhancement mechanisms for collinear DP-LIBS with different pulse delays are also summarized in Table 2.

4. Conclusions

In this study, a dual-wavelength differential spectroscopic imaging system and a time-resolved spectra acquisition system were employed to obtain both spectral and image information from SP and collinear DP aluminum plasmas with varying inter-pulse delays. The expansion dynamics of the vapor in the ambient gas and the relative distribution of different species were analyzed. Additionally, we investigated the propagation regimes of the plasmas and the signal enhancement mechanisms of DP-LIBS.

The results indicate that DP-LIBS can effectively enhance spectral quality compared to conventional SP-LIBS. The maximum enhancement factors of Al II 358.66 nm, Al I 396.15 nm and Al I 394.40 nm are 1.8, 1.6, and 1.4, respectively. The doubly-ionized lines of Al species can be emitted in the earliest stages of plasma lifetime. Under the experimental conditions in this research, the SP plasma exhibits an obvious layered structure with a lower ionization degree in the shocked gas layer. The characteristics indicate the propagation regime is an LSC wave. In contrast, DP plasmas with short inter-pulse delays exhibit a high ionization degree in the shocked gas layer, with an overlapping distribution of argon and aluminum ions observed at the plasma's top due to strong coupling between the second pulse and the shocked gas layer. In this condition, the propagation regime belongs to the LSD wave, and the signal enhancement mechanism relies on the high reheating efficiency of pre-plasma. For larger inter-pulse delays, internal layered structures of DP plasma are similar to those observed in SP plasma. This similarity arises because the longer expansion time of the pre-plasma reduces the density of the shocked gas layer and reduces the shielding effect. As a result, ablation vapor becomes the main area for laser energy absorption, presenting characteristics of LSC waves. At this time, the rarefaction within the shockwave can enhance coupling between the second pulse and the target, with the dominant enhancement mechanism being an increase in ablation mass.

The spectral lines that are difficult to excite using conventional SP-LIBS can be obtained using DP-LIBS by selecting an appropriate inter-pulse delay for the two pulses and setting a short delay time and narrow gate as acquisition parameters. This approach provides a new perspective for trace element measurement in LIBS, and we look forward to further technological breakthroughs.

In addition to DP-LIBS, several potential strategies exist for improving the analytical performance of LIBS in practical industrial applications. For example, by combining LIBS with high-reproducibility and high-sensitivity analytical techniques such as laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS), X-ray fluorescence spectrometry (XRF), and laser-induced fluorescence (LIF), the comprehensive quantitative analytical capabilities for multiple elements can be significantly enhanced while retaining the advantage of the broad elemental dynamic range of LIBS. Furthermore, introducing

advanced AI-based chemometric algorithms to optimize quantitative models also provides an effective approach for improving trace element detection accuracy.

Author Contributions: Conceptualization, Y.Z. and L.Z.; methodology, Y.Z.; software, Y.Z.; validation, Z.T.; formal analysis, X.Z. and J.B.; investigation, X.Z. and J.B.; writing—original draft preparation, Y.Z.; writing—review and editing, L.Z.; supervision, L.Z. and W.Y.; funding acquisition, L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: National Energy R&D Center of Petroleum Refining Technology (RIPP, SINOPEC); National Natural Science Foundation of China (NSFC) (12404467, 12374377, 12104417, 62471443, 61975103, and 627010407); Changjiang Scholars and Innovative Research Team in University of Ministry of Education of China (IRT_17R70); 111 Project (D18001); Fund for Shanxi "1331KSC"; Scientific Research Program of Hubei Provincial Department of Education (B2023162); Fundamental Research Program of Shanxi Province (202203021212122, 20210302124161, and 202103021223210), and Central Guidance on Local Science and Technology Development Fund of Shanxi Province (YDZJSX2024C010).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

Abbreviations

The following abbreviations are used in this manuscript:

LIBS	Laser-induced breakdown spectroscopy
SP-LIBS	Single-pulse laser-induced breakdown spectroscopy
DP-LIBS	Double-pulse laser-induced breakdown spectroscopy
SP	Single pulse
DP	Double pulse
LSA wave	Laser-supported absorption wave
LSC wave	Laser-supported combustion wave
LSD wave	Laser-supported detonation wave
LOD	Limit of detection
FWHM	Full width at half maximum
LA-ICP-MS	Laser ablation inductively coupled plasma mass spectrometry
XRF	X-ray fluorescence spectrometry
LIF	Laser-induced fluorescence

References

1. Sheta, S.; Afgan, M.S.; Hou, Z.; Yao, S.; Zhang, L.; Li, Z.; Wang, Z. Coal analysis by laser-induced breakdown spectroscopy: A tutorial review. *J. Anal. At. Spectrom.* **2019**, *34*, 1047–1082. [CrossRef]
2. Hou, Z.; Wang, Z.; Li, L.; Yu, X.; Li, T.; Yao, H.; Yan, G.; Ye, Q.; Liu, Z.; Zheng, H. Fast measurement of coking properties of coal using laser induced breakdown spectroscopy. *Spectrochim. Acta Part B* **2022**, *191*, 106406. [CrossRef]
3. Bie, N.; Wang, J.; Lv, P.; Zhang, Y.; Bai, Y.; Song, X.; Su, W.; Yu, G. In-situ release characteristic of alkali metals during co-pyrolysis of coal and biomass in a visual fixed bed combined with laser-induced breakdown spectroscopy. *Fuel* **2023**, *331*, 125868. [CrossRef]
4. Erler, A.; Riebe, D.; Beitz, T.; Löhmannsröben, H.-G.; Leenen, M.; Pätzold, S.; Ostermann, M.; Wójcik, M. Mobile Laser-Induced Breakdown Spectroscopy for Future Application in Precision Agriculture—A Case Study. *Sensors* **2023**, *23*, 7178. [CrossRef]
5. Palleschi, V. Understanding the source of signal fuctuations in laser-induced breakdown spectroscopy analytical applications. *Front. Phys.* **2020**, *16*, 22503. [CrossRef]
6. Yu, Z.; Yao, S.; Jiang, Y.; Chen, W.; Xu, S.; Qin, H.; Lu, Z.; Lu, J. Comparison of the matrix effect in laser induced breakdown spectroscopy analysis of coal particle flow and coal pellets. *J. Anal. At. Spectrom.* **2021**, *36*, 2473–2479. [CrossRef]

7. Giannakaris, N.; Siozos, P.; Piñon, V.; Banerjee, S.P.; Sentis, M.; Pedarnig, J.D.; Anglos, D. UV femtosecond single-pulse and collinear double-pulse laser-induced breakdown spectroscopy (LIBS) for depth-resolved characterization of nano-scaled films. *Appl. Surf. Sci.* **2023**, *640*, 158354. [CrossRef]
8. Chen, Y.; Wan, X.; Si, J.; Han, J.; Chen, A.; Jin, M. High-sensitivity analysis of trace elements in water using femtosecond LIBS with dry droplet pretreatment on a metallic substrate. *J. Anal. At. Spectrom.* **2024**, *39*, 1225–1234. [CrossRef]
9. Wang, Q.; Chen, A.; Chen, Y.; Jiang, Y.; Li, S.; Jin, M. Highly sensitive analysis of trace Pb in aqueous solution using electrodeposition and spark-discharge assisted laser-induced breakdown spectroscopy. *J. Anal. At. Spectrom.* **2021**, *36*, 1889–1894. [CrossRef]
10. Wakil, M.; Alwahabi, Z.T. Gated and non-gated silver detection using microwave-assisted laser induced breakdown spectroscopy. *J. Anal. At. Spectrom.* **2021**, *36*, 185–193. [CrossRef]
11. Corsi, M.; Cristoforetti, G.; Giuffrida, M.; Hidalgo, M.; Legnaioli, S.; Palleschi, V.; Salvetti, A.; Tognoni, E.; Vallebona, C. Three-dimensional analysis of laser induced plasmas in single and double pulse configuration. *Spectrochim. Acta Part B* **2004**, *59*, 723–735. [CrossRef]
12. Wang, J.; Li, X.; Li, H.; Yin, C. Analysis of the trace elements in micro-alloy steel by reheating double-pulse laser-induced breakdown spectroscopy. *Appl. Phys. B* **2017**, *123*, 131. [CrossRef]
13. Bhatt, C.R.; Hartzler, D.; Jain, J.C.; McIntyre, D.L. Evaluation of analytical performance of double pulse laser-induced breakdown spectroscopy for the detection of rare earth elements. *Opt. Laser Technol.* **2020**, *126*, 106110. [CrossRef]
14. Nicolodelli, G.; Senesi, G.S.; Ranulfi, A.; Watanabe, A.; Milori, D.M.B.P. Development of a double-pulse (DP) laser-induced breakdown spectroscopy (LIBS) setup in the orthogonal configuration for environmental applications. In Proceedings of the Latin America Optics & Photonics Conference, Medellin, Colombia, 22–26 August 2016.
15. Stratis, D.N.; Eland, K.L.; Angel, S.M. Dual-pulse LIBS using a pre-ablation spark for enhanced ablation and emission. *Appl. Spectrosc.* **2000**, *54*, 1270–1274. [CrossRef]
16. Rai, V.N.; Yueh, F.Y.; Singh, J.P. Time-dependent single and double pulse laser-induced breakdown spectroscopy of chromium in liquid. *Appl. Opt.* **2008**, *47*, 21–29. [CrossRef]
17. Burakov, V.S.; Tarasenko, N.V.; Nedelko, M.I.; Kononov, V.A.; Vasilev, N.N.; Isakov, S.N. Analysis of lead and sulfur in environmental samples by double pulse laser induced breakdown spectroscopy. *Spectrochim. Acta Part B* **2009**, *64*, 141–146. [CrossRef]
18. Sun, D.; Ma, Y.; Wang, Y.; Su, M.; Lu, Q.; Dong, C. Determination of the limits of detection for aluminum-based alloys by spatially resolved single- and double-pulse laser-induced breakdown spectroscopy. *Anal. Methods* **2018**, *10*, 2595–2603. [CrossRef]
19. Cui, M.; Deguchi, Y.; Wang, Z.; Fujita, Y.; Liu, R.; Shiou, F.J.; Zhao, S. Enhancement and stabilization of plasma using collinear long-short double-pulse laser-induced breakdown spectroscopy. *Spectrochim. Acta Part B* **2018**, *142*, 14–22. [CrossRef]
20. Cao, S.; Su, M.; Ma, P.; Wang, K.; Min, Q.; Sun, D.; Cheng, R.; Ying, Y.; Dong, C. Expansion dynamics and emission characteristics of nanosecond–picosecond collinear double pulse laser-induced Al plasma in air. *J. Quant. Spectrosc. Radiat. Transf.* **2020**, *242*, 106773. [CrossRef]
21. Qiu, Y.; Wang, A.; Liu, Y.; Huang, D.; Wu, J.; Li, J.; Zhang, Z.; Li, X.; Wu, Q. The effect of inter-pulse delay on the spectral emission and expansion dynamics of plasma in dual-pulse fiber-optic laser-induced breakdown spectroscopy. *Phys. Plasmas* **2020**, *27*, 083516. [CrossRef]
22. Tognoni, E.; Cristoforetti, G. Basic mechanisms of signal enhancement in ns double-pulse laser-induced breakdown spectroscopy in a gas environment. *J. Anal. At. Spectrom.* **2014**, *29*, 1318–1338. [CrossRef]
23. Pasquini, C.; Cortez, J.; Silva, L.; Gonzaga, F.B. Laser Induced Breakdown Spectroscopy. *J. Braz. Chem. Soc.* **2007**, *18*, 463–512. [CrossRef]
24. Kepes, E.; Gornushkin, I.; Porizka, P.; Kaiser, J. Tomography of double-pulse laser-induced plasmas in the orthogonal geometry. *Anal. Chim. Acta* **2020**, *1135*, 1–11. [CrossRef]
25. Nagli, L.; Gaft, M.; Gornushkin, I. Comparison of single and double-pulse excitation during the earliest stage of laser induced plasma. *Anal. Bioanal. Chem.* **2011**, *400*, 3207–3216. [CrossRef]
26. Zhao, Y.; Zhang, L.; Hou, J.; Ma, W.; Dong, L.; Yin, W.; Xiao, L.; Jia, S.; Yu, J. Species distribution in laser-induced plasma on the surface of binary immiscible alloy. *Spectrochim. Acta Part B* **2019**, *158*, 105644. [CrossRef]
27. Zhao, Y.; Zhang, L.; Wang, S.; Han, J.; Xia, G.; Ma, W.; Dong, L.; Yin, W.; Xiao, L.; Jia, S.; et al. Species distribution in laser-induced plasma on the surface of binary miscible alloy. *Spectrochim. Acta Part B* **2020**, *173*, 105987. [CrossRef]
28. Sherbini, A.M.E.; Hegazy, H.; Sherbini, T.M.E. Measurement of electron density utilizing the Hα-line from laser produced plasma in air. *Spectrochim. Acta Part B* **2006**, *61*, 532–539. [CrossRef]
29. Bai, X.; Cao, F.; Motto-Ros, V.; Ma, Q.; Chen, Y.; Yu, J. Morphology and characteristics of laser-induced aluminum plasma in argon and in air: A comparative study. *Spectrochim. Acta Part B* **2015**, *113*, 158–166. [CrossRef]
30. Bai, X.; Ma, Q.; Perrier, M.; Motto-Ros, V.; Sabourdy, D.; Nguyen, L.; Jalocha, A.; Yu, J. Experimental study of laser-induced plasma: Influence of laser fluence and pulse duration. *Spectrochim. Acta Part B* **2013**, *87*, 27–35. [CrossRef]

31. Ma, Q.; Motto-Ros, V.; Bai, X.; Yu, J. Experimental investigation of the structure and the dynamics of nanosecond laser-induced plasma in 1-atm argon ambient gas. *Appl. Phys. Lett.* **2013**, *103*, 204101. [CrossRef]
32. Zhao, Y.; Zhang, L.; Cheng, N.; Yin, W.; Hou, J.; Bai, C. Research on space-time evolutionary mechanisms of species distributionin laser induced binary plasma. *Spectrosc. Spect. Anal.* **2023**, *43*, 2067–2073.32.
33. Diwakar, P.K.; Harilal, S.S.; Freeman, J.R.; Hassanein, A. Role of laser pre-pulse wavelength and inter-pulse delay on signal enhancement in collinear double-pulse laser-induced breakdown spectroscopy. *Spectrochim. Acta Part B* **2013**, *87*, 65–73. [CrossRef]
34. Yu, J.; Ma, Q.; Motto-Ros, V.; Lei, W.; Wang, X.; Bai, X. Generation and expansion of laser-induced plasma as a spectroscopic emission source. *Front. Phys.* **2012**, *7*, 649–669. [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.

sensors

MDPI

Article

Multimodal MRI Image Fusion for Early Automatic Staging of Endometrial Cancer

Ziyu Zheng [1], Ye Liu [2], Longxiang Feng [3], Peizhong Liu [4], Haisheng Song [2], Lin Wang [5,*] and Fang Huang [6,*]

[1] Informatization Construction and Management Department, Huaqiao University, Quanzhou 362021, China; zhzy@hqu.edu.cn
[2] School of Physics and Electronic Engineering, Northwest Normal University, Lanzhou 730070, China; liuyeauto@163.com (Y.L.); songhs@nwnu.edu.cn (H.S.)
[3] College of Medicine, Huaqiao University, Quanzhou 362021, China; lxfengvon@gmail.com
[4] School of Engineering, Huaqiao University, Quanzhou 362021, China; pzliu@hqu.edu.cn
[5] Department of Radiology, Quanzhou First Hospital Affiliated to Fujian Medical University, Quanzhou 362000, China
[6] Radiology Department, The Second Affiliated Hospital of Fujian Medical University, Quanzhou 362000, China
* Correspondence: wanglin362000@163.com (L.W.); hf-dt@163.com (F.H.)

Abstract: This magnetic resonance imaging multimodal fusion study aims to automate the staging of endometrial cancer using deep learning and to compare the diagnostic performance of deep learning with that of radiologists in the staging of endometrial cancer. This study retrospectively investigated 122 patients with pathologically confirmed early EC from January 1, 2025 to December 31, 2021. Of these patients, 68 were in the International Federation of Gynecology and Obstetrics (FIGO) stage IA, and 54 were in FIGO stage IB. Based on the Swin transformer model and its proprietary SW-MSA (shift window multiple self-coherence) module, magnetic resonance imaging (MRI) images in each of the three planes (sagittal, coronal, and transverse) are cropped, enhanced, and classified, and fusion experiments in the three planes are performed simultaneously. Selecting one plane for the experiment, the accuracy of IA and IB classification was 0.988 in the sagittal, 0.96 in the coronal, and 0.94 in the transverse position, and classification accuracy after the fusion of three planes reached 1. Finally, the automatic classification method based on the Swin transformer has an accuracy of 1, a recall of 1, and a specificity of 1 for early EC classification. In this study, the multimodal fusion approach accurately classified early EC. It was comparable to what a radiologist would perform and simpler and more precise than previous methods that required segmenting followed by staging.

Keywords: endometrial cancer; transformer; deep learning; MRI multi-position; early staging

1. Introduction

Endometrial cancer arises from the endometrium (the lining of the uterus or womb) [1]. It results from abnormal growth of cells that can invade or spread to other parts of the body. The first sign is most often vaginal bleeding not associated with a menstrual period [2]. Other symptoms include pain with urination, pain during sexual intercourse, or pelvic pain [3]. Endometrial cancer occurs most commonly after menopause. It is the fourth most commonly diagnosed cancer in women and the second most common malignant tumor of the female reproductive system. According to the statistics of the International Agency for Research on Cancer (IARC) of the World Health Organization in 2024, the number of new cases of endometrial cancer reached 420,000 worldwide in 2022, while 97,000 people

died of the disease. Endometrial cancer ranked sixth among the cases of female tumors worldwide in that year and third among female reproductive tract tumors, as shown in the yellow bar graph in Figure 1 [4]. The EC incidence rate has increased about 1% per year since the mid-2000s among women aged 50 years and older and nearly 2% yearly since at least the mid-1990s in younger women [5]. The early stages of endometrial cancer can be categorized into stage IA and stage IB according to the Federation of Gynecology and Obstetrics (FIGO) staging [6]. For EC, the prognosis of patients in the early stage is relatively optimistic [7]. In Western countries, endometrial cancer has ranked first in the incidence of malignant tumors of the female reproductive system. Especially in the United States, it has become one of the few cancers with increasing incidence and mortality [8]. Due to the disease being accompanied by irregular vaginal bleeding, inflammation, and other clinical manifestations in the early stage [9], endometrial cancer is usually diagnosed in stage I when the tumor is confined to the uterine. Making an accurate staging prediction plays a crucial role in the treatment and prognosis of the patient [10]. The leading treatment option for endometrial cancer is abdominal hysterectomy (the total removal by surgery of the uterus), together with removal of the fallopian tubes and ovaries on both sides, called a bilateral salpingo-oophorectomy. In more advanced cases, radiation therapy, chemotherapy, or hormone therapy may also be recommended [11]. If the disease is diagnosed early, the outcome is favorable, and the overall five-year survival rate in the United States is greater than 80% [12].

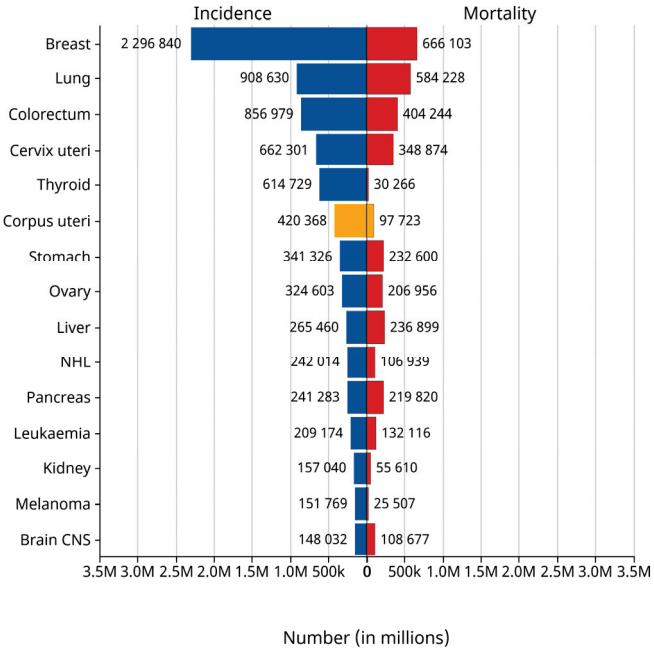


Figure 1. Female cancer cases worldwide by 2022. The orange represents the incidence and mortality of uterine cancer.

MRI has long been recognized as a useful preoperative diagnostic tool and is considered superior to ultrasound and CT in assessing the depth of myometrial and cervical infiltration [13,14]. MRI can show tumor size, indicate the extent of myometrial infiltration or cervical extension, detect enlarged lymph nodes suspected of metastasis, and identify tumors that invade adjacent organs and spread to distant organs. Due to the characteristics of multi-parameter, multi-sequence, multi-directional imaging, and high tissue resolution, MRI can comprehensively display the anatomical structure and signal characteristics of the uterus and surrounding tissues [15,16]. Therefore, the importance of MRI examination lies not only in the accurate localization and qualitative diagnosis of endometrial cancer by MRI

but also in clearly showing the infiltration depth, extent, and lymphatic metastasis of the tumor, and, thus, the tumor can be accurately staged [17]. MRI can guide the choice of clinical surgery and the postoperative evaluation. MRI is the best imaging method for displaying the female pelvic structure due to its numerous advantages. For endometrial cancer, MRI is necessary for staging [18]. The value of multimodal MRI is especially important in stage I, where patients have a 5-year survival rate of up to 90%. Early diagnosis and staging difficulty are limited to detecting small endothelial lesions and superficial muscle invasion. Deep learning demonstrates significant advantages over traditional machine learning techniques. It eliminates the tedious step of manually extracting image features. Instead, it relies on automatic learning mechanisms within the network to recognize features, thus greatly reducing the reliance on a deep background of specialized knowledge.

With the development of deep learning, more and more radiologists are beginning to analyze the MRIs of EC patients using deep learning methods. In the study by Jia et al [19,20], 4252 features from three sequences of T2WI, CE-T1WI, and ADC for each patient were used and tested under five machine learning models. Support vector machines performed best and were validated in both the internal test cohort and the external validation cohort. The corresponding AUCs were 0.875 (95% CI, 0.762–0.988) and 0.862 (95% CI, 0.781–0.942), respectively. However, although three sequences were used throughout the process, only the axial positions were selected, which did not allow a comprehensive analysis of the lesions. The process of manually delineating the ROIs was labor-intensive [21]. Teresa et al. used T2-weighted images (T2WI) and fused diffusion-weighted images (DWI) to assess the depth of myometrial invasion. Both results were compared with histopathological assessment. When comparing the two sets of images (T2WI and fused T2WIDWI) for the diagnosis of myometrial invasion, the fused images had a higher accuracy, and the difference was statistically significant ($p < 0.001$). The diagnostic accuracy rate for the T2WI analysis was 82.1% (70.6–88.7), whereas the fused images were diagnostically correct at 92.1% (79.5–97.2). However, they were also an orientation of the sequence chosen for the long axis of the uterine body [22].

This study proposes a Swin transformer-based framework for the early staging of endometrial cancer with multi-body position fusion. The framework firstly extracts the tumor feature information from three different body positions of the patient's MRI sequence; then, a spatial multi-axial self-attention mechanism is constructed to capture the contextual connections while preserving the positional information, which optimizes the feature characterization and reduces the computation amount; and finally, the classification vector is obtained and fed into the fully connected layer to carry out the classification, completing the early staging of endometrial cancer. Compared with other staging methods, this paper's proposed method considers the interconnectivity of the three body positions, and the staging results are more accurate.

2. Materials and Methods

2.1. Patients

This retrospective study included 122 patients (68 stage in IA and 54 stage in IB) diagnosed with early-stage EC by postoperative biopsy pathology at The Second Affiliated Hospital of Fujian Medical University and Quanzhou First Hospital Affiliated to Fujian Medical University from 1 January 2015 to 31 December 2021. The hospital's medical ethics approved the study, and the requirement for informed consent was waived for all participants. The patient data collected from the hospitals were stored in DICOM format. Out of the 385 patients who underwent MRI, 122 were diagnosed as stage I and had T2WI sequences in three planes. The exclusion criteria were as follows: (1) without a final

pathologic diagnostic statement; (2) absence of complete T2WI sequences in all three planes; and (3) simultaneous with other cancers.

All 122 patients underwent preoperative pelvic magnetic resonance imaging with a pathological diagnosis of stage I EC (age range 28–80 years, mean age: 55.7 years, standard deviation (SD): 7.9 years). The patient selection flowchart is shown in Figure 2, and patients' clinical and pathological data are summarized in Table 1.

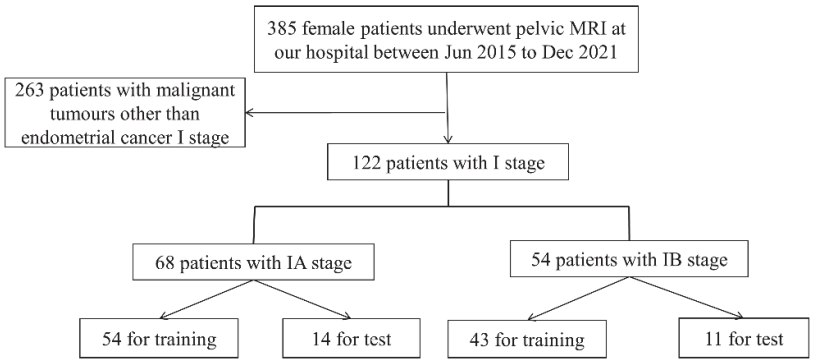


Figure 2. The patient selection process.

Table 1. Clinical and pathological data summaries. Patients' information (n = 122).

Parameter	Stage IA	Stage IB	p Value
Subpopulation	68	54	
Age (year)	51.5 ± 9.7	60.2 ± 8.3	
Endometrial type [#]			0.721
Grade 1	41	20	
Grade 2	24	26	
Grade 3	3	8	
Myometrial invasion			
<50%	65	4	0.873
≥50%	3	50	
Mixed carcinoma			0.875
No	65	51	
Yes	3	3	

[#] Histological grading of endometrioid carcinoma. According to the solid range of tumors, the classification criteria are as follows: Grade 1, solid growth area ≥5%; Grade 2, solid growth area accounts for 6% 50%; Grade 3, solid growth area > 50%.

The table indicates the presence of other tumors, such as clear cell carcinoma, uterine fibroids, etc.

2.2. MRI Examination Methods

All MRI examinations were performed using a 1.5T MRI scanner (Optima MR360, GE Healthcare, Chicago, IL, USA) and a phased-array coil. Patients were defecated with a glycerol enema before the scan and had adequate urine output (approximately half). No enemas or slowing agents were used for bowel movements to reduce bowel artifacts and motility artifacts caused by large bowel movements. Food is allowed (the food must not contain iron), and there is no need for intramuscular medication injections. It must be ensured that the patient has no contraindications to MRI and no metallic foreign bodies in the body.

2.3. Data Preparation and Processing

ITK-snap (version 4.0.1), a medical image reading, analysis, and manipulation software [23], was used to convert the patients' DICOM files into image data, with 19–24 slices per sequence. The sequence used in this study was T2WI, the radiologist's preferred reading for the patient's MRI [24]. For 7808 MRI images (2318 sagittal, 2928 transverse, and 2562 coronal). Some images with severe artifacts were eliminated to ensure the high quality of the training data. Radiologists usually select the anterior and posterior MRI slices of the largest tumor in one or more sequences when diagnosing. Based on this feature, we can localize tumors to layers 8 to 11 in the sagittal plane, 13 to 17 in the transverse plane, and 8 to 12 in the coronal plane. This resulted in 715 sagittal slices, 645 coronal slices, and 505 transverse slices. Experimental data were randomly divided 8:2 into training and test datasets. The training dataset consisted of 97 cases (54 in stage IA, 43 in stage IB) with 1492 images (572 in sagittal position, 516 in coronal position, and 404 in transverse position). The test dataset comprised 25 cases (14 IA, 11 IB) with 373 images (143 sagittal, 129 coronal, 101 transverse). The cohort selection flowchart is shown in Figure 1.

The T2WI sequence was one of the first sequences chosen by radiologists for MRI of patients with endometrial cancer [25,26]. It provides clear imaging with high resolution and allows clear differentiation of the molecular uterus, myometrium, and endometrial layers for staging. Conventional uterine T2WI can realistically show the three-layer structure of myometrium (proliferative stage—medium signal, secretory stage—high signal, postmenopausal stage 3- signal weakening and atrophy), myometrium conjugate (medial portion of myometrium, smooth muscle structure is tight, about 2~8 mm low-signal bands), and endometrial cavity (high signal). Some use a single sequence plane to extract tumor features [27–29]; in this case, we used three planes, the sagittal, coronal, and transverse T2WI, which are presented in Figure 3 below.

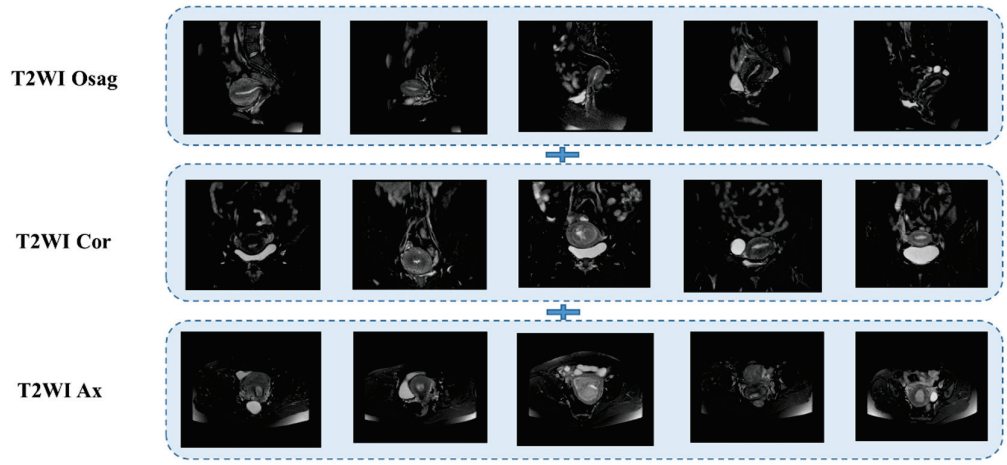


Figure 3. MRI T2WI sequence of endometrial cancer.

2.4. Deep Learning with Swin Transformer Model

The transformer is a natural language processing model that exploits the attention mechanism and is crucial in developing deep learning [30]. The Swin transformer introduces a key innovation in the form of SW-MSA, which linearly solves the problem of self-attentive complexity associated with image size [31]. This enhancement simplifies the processing of high-resolution images, such as MRI images, which are often complex and high-resolution. The cancer foci usually appear as grey and low-saturated pixels in images of T2WI sequences of MRI of endometrial cancer. The enhanced SW-MSA module

in the Swin transformer effectively improves self-attention and enables efficient detection of cancerous focal regions. The model structure of the Swin transformer is shown in Figure 4. The figure shows the Swin transformer block structure, where a Swin transformer block consists of a shift-window-based MSA module followed by a 2-layer MLP sandwiched between GeLU nonlinearities. A LayerNorm (LN) layer is applied before each MSA module and each MLP, and a residual bond is used after each module. LayerNorm (LN) is applied before each MSA module and MLP, with residual bonding after each module.

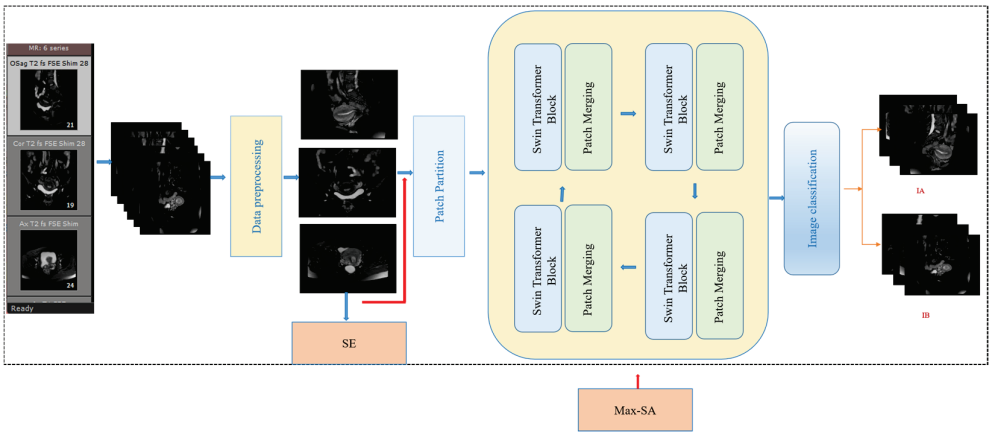


Figure 4. Swin transformer structure.

As shown in Figure 5, for the regular MSA module, the values of their q, k, and v are solved for each patch, and the q solved for any patch is a similar match to the other pixels k in the feature map. The disadvantage of this method is that the windows cannot interact with one another, so the detection field is reduced. Generally, W-MSA and SW-MSA are used in pairs; for example, in the bottom right corner of Figure 3, layer 1 uses W-MSA, and layer 1 + 1 uses SW-MSA, allowing information to be exchanged between windows. At the same time, there is a reduction in the amount of calculation work.

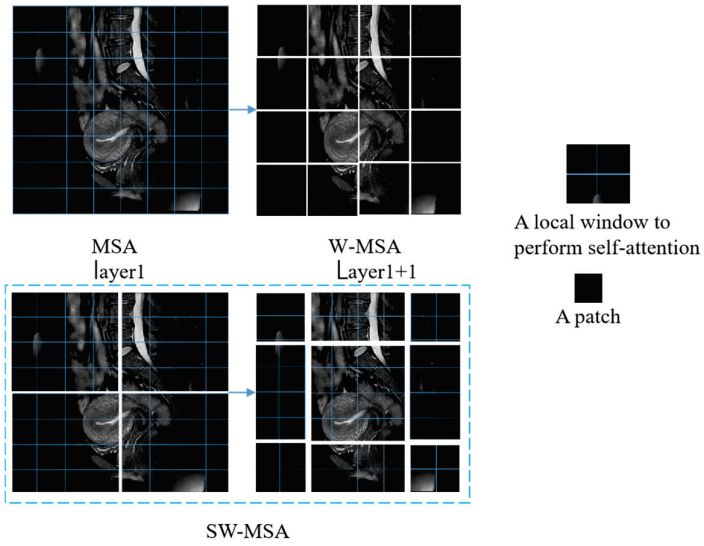


Figure 5. Shifted windows transformer block and sliding attention mechanism.

The calculation of two successive blocks of Swin transformers will be as follows:

$$\hat{Z}^l = W - MSA\left(\mathrm{LN}\left(Z^{l-1}\right)\right) + Z^{l-1} \quad (1)$$

$$\mathrm{Z}^{\mathrm{l}} = \mathrm{MLP}\left(\mathrm{LN}\left(\hat{\mathrm{Z}}^{\mathrm{l}}\right)\right) + \hat{\mathrm{Z}}^{\mathrm{l}} \tag{2}$$

$$\hat{Z}^{l+1} = SW - MSA\left(\mathrm{LN}\left(Z^{l}\right)\right) + Z^{l} \tag{3}$$

$$Z^{l+1} = MLP\left(LN\left(\hat{Z}^{l+1}\right)\right) + \hat{Z}^{l+1} \tag{4}$$

There is no position encoding module in the Swin transformer, and the position information is embedded by calculating the softmax in the attention module by adding the relative position bias B according to the following equation:

$$\mathrm{Attention}(\mathrm{Q},\mathrm{K},\mathrm{V}) = \mathrm{softmax}(\frac{QK^{T}}{\sqrt{d_{K}}} + \mathrm{B})\mathrm{V} \tag{5}$$

Q, K, and V are query, key, and value matrices, and B is taken from the bias matrix.

The trained images enter the Swin transformer module and are first partitioned by the patch partition layer. The partitioned data then enter the linear embedding layer for feature mapping. The feature-mapped data enter the Swin transformer block, which, together with the linear embedding, is known as stage 1. The following 2–4 stages are downsampled to produce the slices. Finally, the output module classifies the data passing through stage 4.

The Swin transformer employs a local window-based self-attention mechanism for feature modeling. The workflow of its core component, W-MSA (window-based multi-head self-attention), can be paraphrased as follows: The input feature map (assuming a spatial resolution of H $\times$ W) is partitioned into a grid of M $\times$ M pixel units of non-overlapping windows, where the window size M is a fixed parameter preset by the network. The feature map is then partitioned into (H/M) $\times$ (W/M) independent computational units, with each window performing a strictly constrained local self-attention computation. This windowing processing strategy makes the computational complexity of the model present a controllable feature, which is shown in Equations (6) and (7).

$$\Omega(MSA) = 4hwC^{2} + 2(hw)^{2}C \tag{6}$$

$$\Omega(W - MSA) = 4hwC^{2} + 2M^{2}hwC \tag{7}$$

where h represents the feature's height, w represents its width, and C represents its depth.

The sliding window-based transformer structure is shown in Figure 4; in the model structure diagram, shifted windows transformer block modules appear in even pairs, fusing the *W-MSA* module of the previous layer and the SW-MSA module of the latter layer. Meanwhile, Figure 5 demonstrate the image-based chunking and window-shifting strategy, where the original 8 $\times$ 8 feature map is uniformly divided into four 4 $\times$ 4 windows, each window being used for localized feature processing. Where the effect of window shifting on the feature map can be seen, each 4 $\times$ 4 window is moved in step 2 along the horizontal or vertical direction to form a new window position; this process is used to capture different local features of the image and reduce redundancy.

The Swin transformer attention mechanism faces a significant computational efficiency bottleneck when directly performing attention operations in the full spatial dimension. This bottleneck stems from the quadratic growth property of the attention operator's space complexity. To address this challenge, this study employs a multi-axis attention decomposition mechanism, which decomposes global attention into two sparsified computational phases, local and global, through spatial axis deconstruction.

As shown in Figure 6. This operation partitions the original feature map into non-overlapping window units of (P $\times$ P) size, and this chunking strategy forms an algorithmic-level correspondence with the established mouth attention mechanism. Under this frame-

work, the chunked attention computation inside each local window focuses on modeling the feature associations in the neighboring regions, thus efficiently enabling fine-grained spatial interactions. This multi-axis decoupled attention architecture maintains global perception capability while significantly reducing computational complexity through spatial constraints.

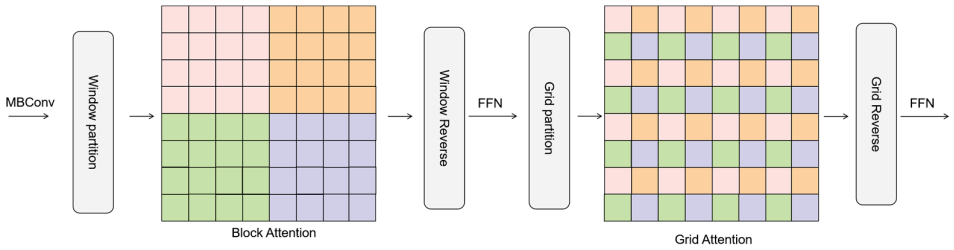


Figure 6. Multiaxial self-attention mechanism.

At the same time, the SE (Squeeze-and-Excitation) module is added, as shown in Figure 7. The core idea of the SE Block is to learn the feature weights through the network according to the loss so that the effective feature blocks are given more weight and the ineffective or ineffective feature blocks are given less weight to train the model to achieve better results. Therefore, this study incorporates the SE module to automatically adjust the feature weights according to the loss so that the model extracts the most useful endometrial cancer staging features from the three positions.

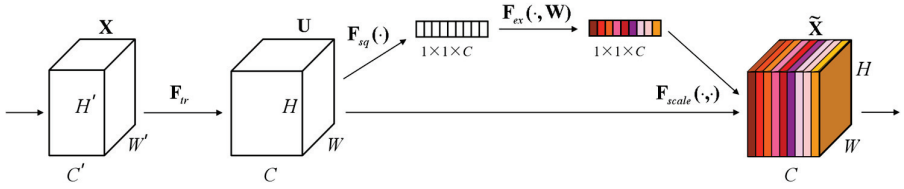


Figure 7. SE block structure.

In the squeeze stage, a global spatial aggregation strategy is used to achieve the squeeze of the null domain information and extract channel-level statistics through channel-by-channel global average pooling. Specifically, the input feature tensor X∈Rˆ(H × W × C) is compressed in the null domain along the spatial dimension $H \times W$. The two-dimensional feature maps of each channel are mapped into scalar form, thus generating the channel-level global description vector Z∈Rˆ(1 × 1 × C). The process can be formally expressed as shown in (8).

$$Z^c = F_{sq}(u_c) = \frac{1}{W * H} \sum_{i=1}^{W} \sum_{j=1}^{H} u_c(i,j) \tag{8}$$

where Z^c denotes the global spatial response intensity of the cth channel; this null domain compression mechanism effectively captures each channel's feature distribution characteristics in the receptive field's full domain, providing a statistical a priori for the subsequent feature recalibration. The conversion of feature characterization from dense distribution in the null domain to compact encoding in the channel dimension is realized by eliminating the redundant spatial dimension information.

The excitation process is the core of SE, which is used to inscribe the weights of the C feature blocks in the tensor U. The formula is shown in (9).

$$\mathrm{S} = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \sigma(W_1 z)) \tag{9}$$

where $\sigma(g(z, W))$ is the Sigmoid function to achieve probabilistic compression; this weight is learned through these previously fully connected and nonlinear layers. The role of these two fully connected layers is to fuse the feature block information of each channel because the previous squeeze operates inside the feature block of a certain channel.

2.5. Comparison of Diagnoses with Radiologists

The 60 patient test images were independently reviewed in random order by three radiologists with 21, 16, and 4 years of experience interpreting pelvic MRI (reader 1, reader 2, and reader 3). Two are associate physicians. One is a registrar. They examined the original images to determine whether the patient was in stage IA or IB. They were blinded to the pathological and clinical results of the study. We will compare the diagnostic results of clinicians with the diagnostic results of the model and statistically evaluate the effectiveness of the depth model as a tool for early staging of endometrial cancer.

3. Results

We chose metrics commonly used in classification tasks, such as precision, recall, the confusion matrix, and the ROC curve [32], to express the accuracy of model detection. *TP* stands for true positive, meaning that the prediction is a positive sample and the prediction is correct. *TN* stands for true negative, meaning that a negative sample is predicted and predicted is correct. *FN* stands for false positive, meaning that a positive sample is predicted but the prediction is incorrect. Finally, *FP* stands for false negative, meaning that a negative sample is predicted but the prediction is wrong. The calculation of the accuracy, precision, recall, and specificity is shown in Equations (10)–(15).

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$Specificity = \frac{TN}{FP + TN} \tag{13}$$

$$\mathrm{TPR} = \frac{TP}{TP + FN} \tag{14}$$

$$\mathrm{FPR} = \frac{FP}{FP + TN} \tag{15}$$

The final results of the experiment are shown in Table 2. The selected and improved model is suitable for our experimental data, as the training accuracies for all body positions are above 0.9. The accuracy of the sagittal position is 0.99, and the accuracy of the multi-somite fusion is 1. However, the specificity of the multi-somite fusion is slightly better than that of the sagittal position. The experiment met and exceeded the expected results. The results in the table are presented after several experiments and model improvements. We chose the Swin transformer light training weight Swin-t for the experiments, and the epoch of the experiments was found to be best at epoch 50 after 20, 50, 100, and 200 experiments, and the time was also relatively short. As a rule, the larger the epochs, the better the training results, but with increasing accuracy, the loss increases and may lead to overfitting. Ultimately, we chose the higher accuracy epoch, where the loss is minimal. The trial results indicate that multi-somatic fusion is more effective than monochromatic training.

Table 2. Different position' experiment results (epochs = 50).

		Accuracy	Precision	Recall	Specificity
Sagittal T2WI	IA	0.99	0.988	1.0	0.98
	IB		1.0	0.98	1.0
Coronal T2WI	IA	0.96	0.967	0.95	0.967
	IB		0.95	0.967	0.95
Axial T2WI	IA	0.94	0.937	0.967	0.9
	IB		0.947	0.9	0.967
Fusion Position	IA	1.0	1.0	1.0	1.0
	IB		1.0	1.0	1.0

We used the loss in precision and confusion matrices to visualize the results. Figure 8 shows the confusion matrix and the accuracy loss. The top half of Figure 8 shows the results of the confusion matrix, the most commonly used metric for classifying. The positive diagonal shows the number of samples correctly classified, and the negative diagonal shows the incorrect classifications. As can be seen from the confusion matrix of the experiment, both the sagittal fusion and the multi-somatic fusion were correctly classified. There were misclassifications in both the coronal and transverse positions. The accuracy loss in the bottom half of Figure 8 shows, from left to right, the results of the sagittal training, the coronal training, and the transverse training and the results of the multi-site fusion. From the graph of the loss in accuracy, it can be seen that they all start from 0.6, and the accuracy shows an increasing trend, and the loss shows a decreasing trend. The accuracy of the sagittal position reaches 0.9 at epoch 20, the coronal position and the transverse position reach 0.9 at epoch 30, and we find that the accuracy of the multi-site fusion training reaches 0.9 at epoch 10. This shows that the multi-site fusion training converges faster. Meanwhile, the zigzag degree of the curve indicates that the training process of multi-site fusion is softer, and the effect of the transverse position is the worst.

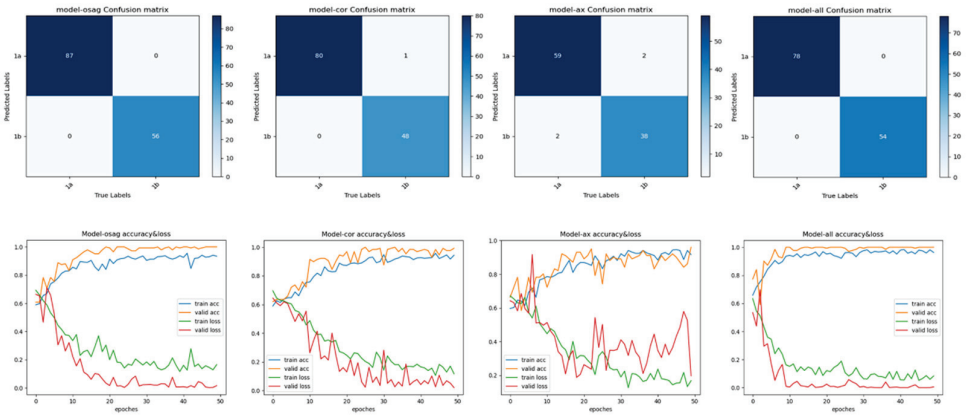


Figure 8. Confusion matrix and accuracy loss in different positions.

Comparisons with some classification models, such as ViT-B/16, ResNet50, and SVM + HOT, are made in this subsection; in the four classification deep learning models explored in this section, we consistently use a uniform set of training strategies to perform the training and testing staging tasks. Specifically, we set each batch to contain eight images (the batch size is 8) and performed 300 rounds of iterative training (the number of epochs is 300) to ensure that the loss function can converge effectively. We chose Adam as the optimization algorithm for all models and set the learning rate to 0.0001. The results are shown in Table 3, where the performance metrics for each model show that the improved

model Swin transformer in this chapter achieves the best performance for all metrics in the endometrial cancer multi-site fusion task. Figure 8 shows the ROC plots for the sagittal, coronal, and transverse positions, respectively, from which it can be seen that the improved Swin transformer model outperforms several other deep learning models, with multi-site fusion outperforming the other three positions in staging.

Table 3. Different models' experiment results (epochs = 300).

Model	Sagittal T2WI	Coronal T2WI	Axial T2WI	Fusion Position
SwinT	93.5	89.1	92.6	94.5
ViT-B/16	86.2	83.2	84.3	86.3
ResNet50	90.4	85.4	88.4	90.1
SVM + HOG	77.5	73.1	76.2	82.6

This paper designs and implements a systematic ablation experiment to investigate the contribution of each module to model performance. The experiments focused on the effects of three core modules, Max-SA multi-axis self-attention, the dynamic weight fusion SE (Squeeze-and-Excitation) module, and the hierarchical pruning strategy, on model performance. The experimental results are shown in Table 4, where the combination of modules significantly affects the model accuracy.

Table 4. Ablation experiment.

Model	Max-SA	SE	MPK	Accuracy
Model (1)	——	——	——	0.876
Model (2)	√			0.935
Model (3)	√	√		0.947
Model (4)	√		√	0.978
Model (5)		√	√	0.953
Model (6)	√	√	√	1.0

Performance analysis of the base model: Model (1), the base model without introducing any optimization module, has an accuracy of 0.876. This result reflects the performance benchmark of the base model without any optimization and provides an important reference for evaluating the subsequent modules.

Single-module performance impact: When the Max-SA multi-axis self-attention mechanism is introduced, the accuracy of Model (2) is significantly improved to 0.935, 6.7 percentage points higher than the base model. This result indicates that the Max-SA multi-axis self-attention mechanism can effectively capture the global contextual information and local detailed features in the image, thus significantly improving the feature extraction capability of the model.

Performance analysis of dual-module combination: In Model (3), the Max-SA multi-axis self-attention mechanism and the SE module are introduced simultaneously, and the model accuracy is further improved to 0.947. This result indicates that the SE module can adaptively adjust the weight distributions of the different feature layers, which creates a good synergy effect with the Max-SA multi-axis self-attention mechanism. Model (4) examines the Max-SA multi-axis self-attention mechanism, which can capture the image's global context information and local details, thus significantly improving the feature extraction capability; on the other hand, examining the effect of combining the Max-SA multiaxial self-attention mechanism with medical prior knowledge (MPK), it achieves an accuracy of 0.978. This significant enhancement suggests that medical a priori knowledge provides domain-specific constraints and guidance for the model, that the a priori information of anatomical structures helps correct staging errors, and that clinical knowledge enhances

the interpretability and reliability of the model. Notably, Model (5), which only used the SE module and medical priori knowledge, resulted in an accuracy of 0.953. Although this result is superior to that of a single module, there is still a gap compared to the combination that includes the Max-SA multiaxial self-attention mechanism, highlighting the attention mechanism's critical role in feature extraction.

Performance analysis of the full-module combination: Model (6) simultaneously incorporates the Max-SA multiaxial self-attention mechanism, the SE module, and medical prior knowledge and achieves a perfect performance of accuracy = 1.0. This result fully demonstrates that the Max-SA multi-axis self-attention mechanism provides a multi-level feature representation, the SE module realizes adaptive integration of features, and the medical a priori knowledge enhances the interpretability and reliability of the model.

Combining the three modules fully exploits their respective advantages and generates significant synergistic effects, ultimately improving model performance overall. This finding provides important guidance for subsequent model optimization. Figure 9 shows the ROC curves of different ablation models.

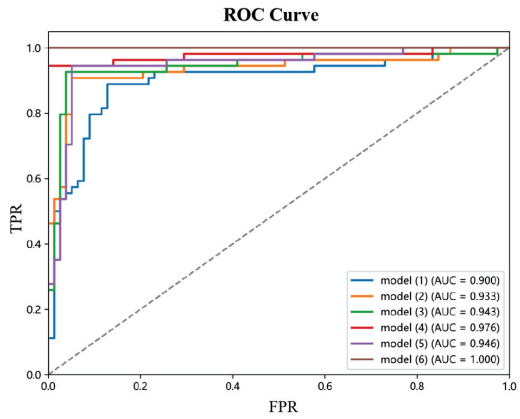


Figure 9. ROC for different ablation models.

4. Discussion

Figure 10 shows the final staging results. The upper side shows the segmentation image of the three positions of T2WI of a patient with stage IA, and the lower side shows the segmentation of the three positions of a patient with stage IB, where the red is the region of cancer foci. The segmentation image shows that the model correctly recognizes the cancerous areas of the three body positions.

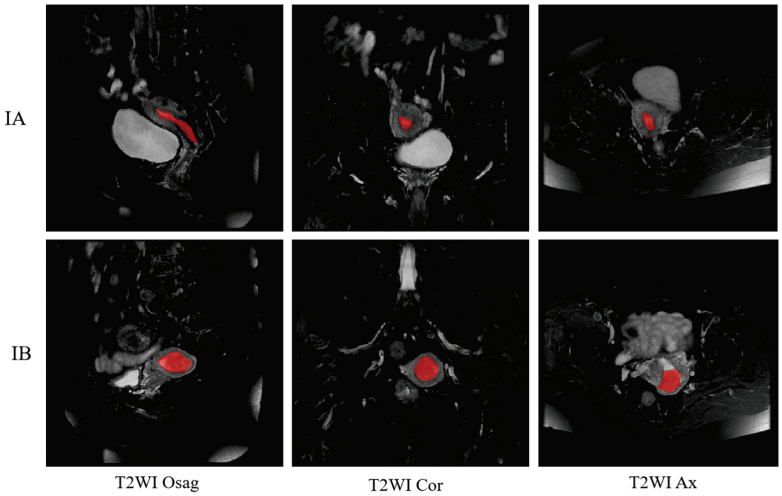


Figure 10. Classification results of T2WI for stage IA and IB. The red represents the cancerous area.

There was also a comparison of the results of three radiologists of different grades and experience with the results of the model. The statistical results are shown in Table 5. A total of 30 patients each were externally validated for IA and IB. The model finally classified two patients from IA to IB and correctly classified all 30 patients from IB. The two radiologists at the associate director level would have misclassified both IA to IB and IB to IA, and the registrar would have misclassified one or two more patients than the director.

Table 5. Test results of the model and readers.

	Model	Reader 1	Reader 2	Reader 3
IA	28	28	27	26
IB	30	28	29	27
IA→IB	2	2	3	4
IB→IA	0	2	1	3

Deep learning, which is particularly suited to image and speech recognition, is a new research direction in machine learning that can automatically learn the intrinsic laws and symbolic levels of sample data. It has been widely used in medical-assisted diagnosis in recent years due to its excellent image-processing capabilities [33–35]. As one of the three most common gynecological tumors in women, endometrial cancer is becoming increasingly common at younger and younger ages. Early staging of this disease is crucial. Early staging and diagnosis are very helpful for young patients who want to preserve their fertility [36]. At the same time, the different stages of endometrial cancer lead to other surgical methods and prognoses. Therefore, it is important to use deep learning to stage endometrial cancer at an early stage.

On one hand, our research shows that multi-position fusion can facilitate the model to extract more features and perform better staging than a single position. On the other hand, our model showed slightly better diagnostic performance than the two associate physicians compared to three radiologists. The statistical results showed that the model would misclassify two patients with stage IA as stage IB and that stage IB was correct. However, radiologists have misdiagnosed both. The size of the uterus and a person's age and whether they are menopausal or not have an effect on the thickness of the endometrium, and the cancer foci in stage I patients are inside the uterus and usually around the endometrium. Stage IA and IB are classified according to whether the infiltration of the myometrium reaches 50%, but the machine is not as accurate for patients at the critical point. The radiologists are better in this respect; they can do an excellent job of classification based on years of experience. However, the radiologists will show that IA is classified as IB and some IB is classified as stage II. Because I told the doctors at the time of the experiment that they could only grade II, they did not grade some IB as IA, but the doctors still thought some IB patients were already stage IB based on their experience and understanding of what that means. Cervical invasion and should be grade II. Although deep learning for medical imaging is not mixed with human factors and can be as objective as possible, its shortcomings are undeniable. It does not have a global view and experience in different cases compared to the doctor. Therefore, rather than replacing the doctor who ultimately makes the diagnosis, our proposed model is more of an aid. Deep learning can significantly reduce a physician's time to read a chart, improving efficiency.

Our study differs from previous studies in that, firstly, there is no need for extensive experience and manual labeling of tumors for segmentation. Instead, it directly and automatically extracts image features relevant to lesion detection. This eliminates the need for extensive time spent on manual labeling. Secondly, most previous studies have been performed using a single-body position. We used a three-body fusion of MRI scans, which

also means that the extraction of the lesion will be more comprehensive and accurate, and it also provides some basis for the accuracy of staging. Thirdly, we compared and validated with radiologists of different experiences and finally reached a level comparable to that of an associate physician.

At the same time, there are still some areas for improvement in our research work: Firstly, our sample of data is relatively small; we aim to build up an extensive database of endometrial cancer for future studies, and we are also actively collecting data from other hospitals. Secondly, although the T2WI sequences we have chosen are said to be the first sequences doctors choose for diagnosis, doctors usually choose several sequences that are read together to ensure that the most appropriate staging diagnosis is made. Our follow-up study will also add DWI and ADC sequences for preliminary data collection.

5. Conclusions

In this study, the Swin transformer used for early endometrial staging was modified, and a fusion experiment was performed using three T2WI positions. The results were superior to those of independent experiments using a single position, with the modified model accurately identifying cancer features in all three positions and staging at a level comparable to that of a resident with approximately 21 years of experience. This study significantly reduces the radiologist's reading workload and improves the radiologist's productivity. The model will improve the accuracy of endometrial cancer staging. This will help determine the best surgical approach for patients. With the continuous development of science and technology, deep learning will likely be applied to more high-end medical diagnostic tools.

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

Funding: This research was funded by the Joint Funds for the innovation of science and Technology, Fu jian province (Grant number: 2024Y9376).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the School of Medicine, Huaqiao University(protocol code: M2024006; date: May 2024).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Amant, F.; Moerman, P.; Neven, P.; Timmerman, D.; Van Limbergen, E.; Vergote, I. Endometrial cancer. *Lancet* **2005**, *366*, 491–505. [CrossRef] [PubMed]
2. Zhou, X.; Zeng, Y.; Zheng, R.; Wang, Y.; Li, T.; Song, S.; Zhang, S.; Huang, J.; Ren, Y. Natural products modulate cell apoptosis: A promising way for treating endometrial cancer. *Front. Pharmacol.* **2023**, *14*, 1209412. [CrossRef] [PubMed]
3. Farzaneh, F.; Ashtiani, A.J.; Hashemi, M.; Hosseini, M.S.; Arab, M.; Ashrafganjoei, T.; Chayjan, S.H. Endometrial cancer in women with abnormal uterine bleeding: Data mining classification methods. *Casp. J. Intern. Med.* **2023**, *14*, 526.
4. Siegel, R.L.; Miller, K.D.; Wagle, N.S.; Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin.* **2023**, *73*, 17–48. [CrossRef]
5. Shen, Y.; Yang, W.; Liu, J.; Zhang, Y. Minimally invasive approaches for the early detection of endometrial cancer. *Mol. Cancer* **2023**, *22*, 53. [CrossRef]
6. Lewin, S.N. Revised FIGO staging system for endometrial cancer. *Clin. Obstet. Gynecol.* **2011**, *54*, 215–218. [CrossRef]

7. Beavis, A.L.; Blechter, B.; Najjar, O.; Fader, A.N.; Katebi Kashi, P.; Rositch, A.F. Identifying women 45 years and younger at elevated risk for endometrial hyperplasia or cancer. *Gynecol. Oncol.* **2023**, *174*, 98–105. [CrossRef]
8. Marín-Jiménez, J.A.; García-Mulero, S.; Matías-Guiu, X.; Piulats, J.M. Facts and hopes in immunotherapy of endometrial cancer. *Clin. Cancer Res.* **2022**, *28*, 4849–4860. [CrossRef]
9. Zhang, S.; Gong, T.T.; Liu, F.H.; Jiang, Y.T.; Sun, H.; Ma, X.X.; Zhao, Y.H.; Wu, Q.J. Global, regional, and national burden of endometrial cancer, 1990–2017: Results from the Global Burden of Disease study, 2017. *Front. Oncol.* **2019**, *9*, 1440. [CrossRef]
10. Jiang, X.; Song, J.; Zhang, A.; Cheng, W.; Duan, S.; Liu, X.; Chen, T. Preoperative Assessment of MRI-Invisible Early-Stage Endometrial Cancer With MRI-Based Radiomics Analysis. *J. Magn. Reson. Imaging* **2023**, *58*, 247–255. [CrossRef]
11. Degez, M.; Caillon, H.; Chauviré-Drouard, A.; Leroy, M.; Lair, D.; Winer, N.; Thubert, T.; Le Floch, P.; Desroys du Roure, V.; Randet, M.; et al. HE4 in the diagnostic approach of endometrial cancer in patients with postmenopausal bleeding, the METRODEC protocol: Protocol for a multicenter prospective study. *Diagnostics* **2021**, *11*, 1274. [CrossRef] [PubMed]
12. Braun, M.M.; Overbeek-Wager, E.A.; Grumbo, R.J. Diagnosis and management of endometrial cancer. *Am. Fam. Physician* **2016**, *93*, 468–474. [PubMed]
13. Faria, S.C.; Devine, C.E.; Rao, B.; Sagebiel, T.; Bhosale, P. Imaging and staging of endometrial cancer. In *Seminars in Ultrasound, CT and MRI*; WB Saunders: Philadelphia, PA, USA, 2019; Volume 40, No. 4. pp. 287–294.
14. Antonsen, S.L.; Jensen, L.N.; Loft, A.; Berthelsen, A.K.; Costa, J.; Tabor, A.; Qvist, I.; Hansen, M.R.; Fisker, R.; Andersen, E.S.; et al. MRI, PET/CT and ultrasound in the preoperative staging of endometrial cancer—A multicenter prospective comparative study. *Gynecol. Oncol.* **2013**, *128*, 300–308. [CrossRef] [PubMed]
15. Rei, M.; Costa-Santos, C.; Bernardes, J.; Costa, A. Preoperative staging of uterine cancer: Can transvaginal ultrasonography play a role? *Front. Oncol.* **2023**, *13*, 1089105. [CrossRef]
16. Lin, Z.; Wang, T.; Li, Q.; Bi, Q.; Wang, Y.; Luo, Y.; Feng, F.; Xiao, M.; Gu, Y.; Qiang, J.; et al. Development and validation of MRI-based radiomics model to predict recurrence risk in patients with endometrial cancer: A multicenter study. *Eur. Radiol.* **2023**, *33*, 5814–5824. [CrossRef]
17. Grevenkamp, F.; Kommoss, F.; Kommoss, F.; Lax, S.; Fend, F.; Wallwiener, D.; Schönfisch, B.; Krämer, B.; Brucker, S.Y.; Taran, F.A.; et al. Second opinion expert pathology in endometrial cancer: Potential clinical implications. *Int. J. Gynecol. Cancer* **2017**, *27*, 289–296. [CrossRef]
18. Nougaret, S.; Horta, M.; Sala, E.; Lakhman, Y.; Thomassin-Naggara, I.; Kido, A.; Masselli, G.; Bharwani, N.; Sadowski, E.; Ertmer, A.; et al. Endometrial cancer MRI staging: Updated guidelines of the European Society of Urogenital Radiology. *Eur. Radiol.* **2019**, *29*, 792–805. [CrossRef]
19. Mao, W.; Chen, C.; Gao, H.; Xiong, L.; Lin, Y. Quantitative evaluation of myometrial infiltration depth ratio for early endometrial cancer based on deep learning. *Biomed. Signal Process. Control* **2023**, *84*, 104685. [CrossRef]
20. Lefebvre, T.L.; Ciga, O.; Bhatnagar, S.R.; Ueno, Y.; Saif, S.; Winter-Reinhold, E.; Dohan, A.; Soyer, P.; Forghani, R.; Siddiqi, K.; et al. Predicting histopathology markers of endometrial carcinoma with a quantitative image analysis approach based on spherical harmonics in multiparametric MRI. *Diagn. Interv. Imaging* **2023**, *104*, 142–152. [CrossRef]
21. Jia, Y.; Hou, L.; Zhao, J.; Ren, J.; Li, D.; Li, H.; Cui, Y. Radiomics analysis of multiparametric MRI for preoperative prediction of microsatellite instability status in endometrial cancer: A dual-center study. *Front. Oncol.* **2024**, *14*, 1333020. [CrossRef]
22. Neves, T.R.; Correia, M.T.; Serrado, M.A.; Horta, M.; Caetano, A.P.; Cunha, T.M. Staging of Endometrial Cancer Using Fusion T2-Weighted Images with Diffusion-Weighted Images: A Way to Avoid Gadolinium? *Cancers* **2022**, *14*, 384. [CrossRef] [PubMed]
23. Yushkevich, P.A.; Gao, Y.; Gerig, G. ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 3342–3345.
24. Karaca, L.; Özdemir, Z.M.; Kahraman, A.; Yılmaz, E.; Akatlı, A.; Kural, H. Endometrial carcinoma detection with 3.0 Tesla imaging: Which sequence is more useful. *Eur. Rev. Med. Pharmacol. Sci.* **2022**, *26*, 8098–8104. [PubMed]
25. Bi, Q.; Chen, Y.; Chen, J.; Zhang, H.; Lei, Y.; Yang, J.; Zhang, Y.; Bi, G. Predictive value of T2-weighted imaging and dynamic contrast-enhanced MRI for assessing cervical invasion in patients with endometrial cancer: A meta-analysis. *Clin. Imaging* **2021**, *78*, 206–213. [CrossRef] [PubMed]
26. Feng, L.; Chen, C.; Wang, L.; Zhang, J.; Li, Y.; Yang, T.; Fan, Y.; Liu, P.; Sun, P.; Huang, F. ECMS-NET: A multi-task model for early endometrial cancer MRI sequences classification and segmentation of key tumor structures. *Biomed. Signal Process. Control* **2024**, *93*, 106223. [CrossRef]
27. Yamada, I.; Miyasaka, N.; Kobayashi, D.; Wakana, K.; Oshima, N.; Wakabayashi, A.; Sakamoto, J.; Saida, Y.; Tateishi, U.; Eishi, Y. Endometrial carcinoma: Texture analysis of apparent diffusion coefficient maps and its correlation with histopathologic findings and prognosis. *Radiol. Imaging Cancer* **2019**, *1*, e190054. [CrossRef]
28. Stanzione, A.; Cuocolo, R.; Del Grosso, R.; Nardiello, A.; Romeo, V.; Travaglino, A.; Raffone, A.; Bifulco, G.; Zullo, F.; Insabato, L.; et al. Deep myometrial infiltration of endometrial cancer on MRI: A radiomics-powered machine learning pilot study. *Acad. Radiol.* **2021**, *28*, 737–744. [CrossRef]

29. Xu, X.; Li, H.; Wang, S.; Fang, M.; Zhong, L.; Fan, W.; Dong, D.; Tian, J.; Zhao, X. Multiplanar MRI-based predictive model for preoperative assessment of lymph node metastasis in endometrial cancer. *Front. Oncol.* **2019**, *9*, 1007. [CrossRef]
30. Han, K.; Xiao, A.; Wu, E.; Guo, J.; Xu, C.; Wang, Y. Transformer in transformer. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 15908–15919.
31. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
32. Singh, S.; Mittal, N.; Singh, H. Classification of various image fusion algorithms and their performance evaluation metrics. In *Computational Intelligence for Machine Learning and Healthcare Informatics*; De Gruyter Brill: Berlin, Germany, 2020; pp. 179–198.
33. Volinsky-Fremond, S.; Horeweg, N.; Andani, S.; Barkey Wolf, J.; Lafarge, M.W.; de Kroon, C.D.; Ørtoft, G.; Høgdall, E.; Dijkstra, J.; Jobsen, J.J.; et al. Prediction of recurrence risk in endometrial cancer with multimodal deep learning. *Nat. Med.* **2024**, *30*, 1962–1973. [CrossRef]
34. Chen, X.; Wang, Y.; Shen, M.; Yang, B.; Zhou, Q.; Yi, Y.; Liu, W.; Zhang, G.; Yang, G.; Zhang, H. Deep learning for determining myometrial invasion depth and automatic lesion identification in endometrial cancer MR imaging: A preliminary study in a single institution. *Eur. Radiol.* **2020**, *30*, 4985–4994. [CrossRef]
35. Wang, C.W.; Muzakky, H.; Firdi, N.P.; Liu, T.C.; Lai, P.J.; Wang, Y.C.; Yu, M.H.; Chao, T.K. Deep learning to assess microsatellite instability directly from histopathological whole slide images in endometrial cancer. *NPJ Digit. Med.* **2024**, *7*, 143. [CrossRef]
36. Schubert, M.; Mettler, L.; Deenadayal Tolani, A.; Alkatout, I. Fertility Preservation in Endometrial Cancer—Treatment and Molecular Aspects. *Medicina* **2023**, *59*, 221. [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.

sensors

MDPI

Article

Simulation Research and Analysis of Wavelength Modulation Off-Axis Integrated Cavity Output Spectrum Measurement System

Tao Wu [1,2], Xiao Zhang [1,2], Xiao Chen [1,2], Wangwang Liu [1,2], Yan Han [1,2,*], Yubin Zhong [1,2], Dan Zhao [1,2], Zhen Fang [1,2], Linxin Pan [1,2], Feiyang Wang [1,2] and Hang Xu [1,2]

[1] Deep Space Exploration Laboratory, Hefei 230000, China
[2] National Key Laboratory of Deep Space Exploration, Hefei 230000, China
* Correspondence: hanyan@ustc.edu

Abstract: Wavelength modulation spectroscopy off-axis integrated cavity output spectroscopy (WMS-OA-ICOS) is an in situ detection technique suitable for analyzing trace gases in the atmospheres, characterized by its high sensitivity and ease of integration. However, in current practical applications, the design and optimization of WMS-OA-ICOS systems primarily rely on empirical knowledge, lacking systematic quantitative methodologies. To address this limitation, this study conducts comprehensive modeling and simulation research on WMS-OA-ICOS spectroscopy, proposing a novel modeling approach. The spot distribution simulation results obtained from the self-developed model are validated against those generated using Tracepro. Furthermore, based on the self-developed model, an in-depth investigation is conducted into the effects of cavity length tolerance, beam waist matching, modulation depth, and laser linewidth on signal quality. The findings provide valuable insights for designing and optimizing miniaturized systems with high signal-to-noise ratios.

Keywords: gas sensor; WMS-OA-ICOS; extraterrestrial planets; self-developed model; simulation; designing and optimizing

1. Introduction

One of the most significant scientific objectives of deep space exploration is the search for evidence of life on other planets [1]. This quest primarily involves detecting environmental parameters associated with life, such as methane isotopes and other gases that play critical roles in creating conditions suitable for life [2–5]. Early efforts to detect methane on Mars relied on ground-based high-resolution infrared spectroscopy via observatories, but this approach suffered from low spatial resolution and reliability issues due to the significant detection distance [6]. To address these challenges, the planetary Fourier spectrometer technology aboard Mars orbiters was employed for subsequent missions for occultation remote sensing of the atmospheric composition. However, this method's sensitivity was significantly affected by Martian surface visibility (e.g., sandstorms and other climatic conditions), making it difficult to detect trace gas concentration variations on the Martian surface [7]. In the 2011 "Mars Science Laboratory (MSL)" mission, NASA's Curiosity rover deployed a tunable laser spectrometer (TLS) with a Herriott cell of 16.8 m effective absorption path length payload for surface exploration, capturing a pulse concentration of approximately 7×10^{-9}. This milestone represented the first and, so far, only successful real-time in situ monitoring of trace methane on the Martian surface [8]. Normally, methane

concentrations on Mars are extremely low (less than 1 ppbv). Therefore, a pulse of methane concentration of 7×10^{-9} means that there may have been a special emission source at this time, which has important implications for the study of whether life existed on Mars. To further confirm whether this methane pulse is the source of life, it is necessary to measure the isotopic gas of methane. However, due to interference noise between light spots, the effective absorption path length of tunable laser absorption spectroscopy (TDLAS) is typically limited to a few hundred meters, making it insufficient for measuring methane isotopes on Mars (e.g., a confocal optical multipass gas cell with an effective optical path of 580 m was proposed by Xia et al.) [9]. The development of off-axis integrated cavity output spectroscopy (OA-ICOS) significantly extended the absorption path length to several kilometers. This technique introduces the laser beam into the cavity at an angle to the central axis, exciting high-order transverse modes (TEM_{mn}) within the cavity. This increases the cavity's spectral mode density, suppresses cavity mode effects, and produces nearly continuous absorption signals, overcoming the strict optical alignment requirements of traditional ICOS [10]. To further enhance the detection capability of OA-ICOS systems, some research groups have integrated wavelength modulation spectroscopy (WMS) with OA-ICOS in practical applications, achieving superior detection performance [11,12].

The intensity differences between fundamental and higher-order modes, as well as interference effects arising from off-axis incident spot distributions, make the input beam characteristics and cavity parameters critical factors influencing the gas absorption signals in OA-ICOS systems [13]. Traditional approaches often estimate the number of spots based on the mirror dimensions and perform preliminary cavity length calculations using the "re-entrant" theory [14,15]. However, as factors such as spot number, cavity length, and laser linewidth can all impact interference fringe noise differently, such estimation methods lack a quantitative basis and are unsuitable for optimizing system parameters. To address this issue, several research groups have explored modeling approaches for OA-ICOS. In 2013, Romanini et al. [16] decomposed the input beam into a superposition of transverse electromagnetic (TEM) modes of the cavity and used a small-angle approximation combined with a recursive algorithm to calculate the projection coefficients of the excitation field on each eigenfunction of the system. Multiplying these coefficients by the corresponding cavity response functions yielded the overall optical field distribution within the cavity. In 2015, Kasyutich et al. [17] developed a steady-state model for an ideal off-axis integrated cavity to investigate the influence of laser linewidth on the transmitted signal. However, this model did not account for interference effects between off-axis spot distributions. In 2018, Shen et al. [18] addressed the inaccuracies in Romanini's model caused by the small-angle approximation when the off-axis radius was large and examined the influencing factors of off-axis integrated cavities under direct absorption. In 2021, Zheng et al. [19] built upon this model to propose the MIMO (multi-input multi-output)-OA-ICOS method, achieving a denser cavity mode structure and thereby improving the system's signal-to-noise ratio (SNR). Despite these advancements, the development of practical WMS-OA-ICOS systems still largely relies on empirical methods. For example, similar to TDLAS measurement techniques, parameters related to second harmonic settings were experimentally optimized, but these studies did not systematically consider the effect of parameter configuration on second harmonic signals, especially when it comes to cavity mode effects [20,21]. Systematic studies on the effects of optical modeling and system parameter characteristics on second harmonic signals remain scarce and a theoretical guidance model capable of facilitating system parameter design and optimization is still lacking.

This study proposes a modeling approach for the WMS-OA-ICOS optical system and, based on the developed model, simulates both direct absorption and wavelength modulation signals under various conditions for target gas absorption. By performing

fitting analysis on the absorption signals, this study investigates the effects of factors such as cavity length tolerance, beam waist matching, modulation depth, and laser linewidth on interference between light spots, and evaluates their respective impacts on the measurement results of the wavelength modulation off-axis integrated cavity output spectroscopy (WMS-OA-ICOS) method. The proposed modeling approach and research provide a straightforward and intuitive means for the preliminary design of WMS-OA-ICOS systems based on specific parameter requirements. This work offers valuable guidance for the construction and optimization of WMS-OA-ICOS systems through both theoretical analysis and modeling simulation.

2. Modeling Methodology

Under coaxial incidence, the free spectral range (FSR) of the transmitted signal can be calculated using Equation (1). When the beam is introduced off-axis, higher-order transverse modes (TEM_{mn}) are excited within the cavity. If the "re-entrant" condition is satisfied, the excited TEM_{mn} modes form degenerate mode groups with equal frequency intervals within the original FSR of the cavity. The FSR in this case can be determined using Equation (2), which is reduced by a factor of M compared to the coaxial incidence scenario. Consequently, the density of coupled cavity modes is significantly increased, resulting in higher spectral resolution.

$$FSR = \frac{c}{2d} \tag{1}$$

$$FSR' = \frac{c}{2Md} \tag{2}$$

In the aforementioned equation, c signifies the velocity of light, whereas d indicates the length of the cavity. To delve deeper into the factors that influence the absorption spectroscopic signals within the WMS-OA-ICOS system, a mathematical model of the WMS-OA-ICOS system is formulated. Within the framework of the wavelength-modulated off-axis integrated cavity system, it is postulated that the electric field reflectivity (R_e) and transmittance (T_e) of the high-reflectance mirrors positioned at both ends of the cavity are such that the loss of the cavity mirrors is negligible, yielding the relationship $R_e = 1 - T_e$. Consequently, the reflectivity (R) and transmittance (T) of the light intensity can be expressed as $R = {R_e}^2$ and $T = {T_e}^2$, respectively. To ascertain the aggregate light field exiting the cavity, it is imperative to consider the summation of all optical pathways within the cavity. The duration for a light beam to complete one round trip within the cavity is denoted as t_d ($t_d = 2\,d/c$), and the amplitude of the light beam passing through at a specific instant in time is ${T_e}^2$. Upon each round trip within the high-reflectance cavity, the amplitude of the light field diminishes to ${R_e}^2$ times its preceding value. By summing the electric fields within the cavity over an infinite sequence of round trips, the intra-cavity light field $E(z,t)$ at position z and the output electric field $E_{out}(t)$ (measured beyond the output mirror at $z = d$) are obtained. The superposition of light fields in the coaxial scenario has been previously derived [22]. In contrast, in the off-axis scenario, an M-fold re-entrant spot distribution forms at the output mirror of the high-reflectance cavity, with the electric field of the beam initially reaching the output mirror serving as the initial condition. Considering M re-entrant events as constituting one complete cycle, and denoting the number of cycles by n, M spot distributions are formed on the rear cavity mirror. Within a single cycle, the beam arrives at the rear cavity mirror m times. $E_{out,m}(t)$ represents the output field intensity at the m-th spot of the first period on the output mirror, while $E_{out,m(n)}(t)$ corresponds to

the output field at the mmm-th spot during the n-th period. The total output field intensity of the rear cavity mirror is given by $E_{\mathrm{out}}(t)$.

$$\begin{aligned} E(z,t) = & \sum_{n=1}^{\infty}\sum_{m=1}^{M} T_{\mathrm{e}}R_{\mathrm{e}}{}^{2(n-1)M}E_{\mathrm{out,m}(1)}(t-(\frac{2ndM}{c})-\frac{z}{c})- \\ & \sum_{n=1}^{\infty}\sum_{m=1}^{M} T_{\mathrm{e}}R_{e}{}^{2(n-1)M+1}E_{\mathrm{out,m}(1)}(\frac{2(n-1)dM}{c}+\frac{d-z}{c}) \\ = & \sum_{m=1}^{M}\sum_{n=1}^{\infty} T_{\mathrm{e}}R_{\mathrm{e}}{}^{2(n-1)M}E_{\mathrm{out,m}(1)}(t-\frac{2(n-1)dM-d}{c})- \\ & \sum_{m=1}^{M}\sum_{n=1}^{\infty} T_{\mathrm{e}}R_{\mathrm{e}}{}^{2nM+1}E_{\mathrm{out,m}(1)}(\frac{(2(n-1)M+1)d-z}{c}) \end{aligned} \tag{3}$$

$$E_{\mathrm{out,m}}(t) = \sum_{n=1}^{\infty} T_{e}{}^{2}R_{e}{}^{2(n-1)M}E_{\mathrm{out,m}(1)}(t-(n-1)Mt_{\mathrm{d}}) \tag{4}$$

$$E_{\mathrm{out}}(t) = \sum_{m=1}^{M} E_{\mathrm{out,m}}(t) = \sum_{m=1}^{M}\sum_{n=1}^{\infty} T_{e}{}^{2}R_{e}{}^{2(n-1)M}E_{\mathrm{out,m}(1)}(t-(n-1)Mt_{\mathrm{d}}) \tag{5}$$

By performing a Fast Fourier Transform (FFT) on the optical field at the output end, the transmission spectrum of the cavity is obtained:

$$\begin{aligned} \hat{E}_{\mathrm{out}}(\omega) = & \frac{1}{\sqrt{2\pi}}\int E_{\mathrm{out}}(t)\exp(-j\omega t)dt \\ = & \frac{1}{\sqrt{2\pi}}\int \sum_{m=1}^{M}\sum_{n=1}^{\infty} T_{\mathrm{e}}{}^{2}R_{\mathrm{e}}{}^{2(n-1)M}E_{\mathrm{out,m}(1)}(t-(n-1)Mt_{\mathrm{d}})\exp(-j\omega t)dt \end{aligned} \tag{6}$$

By setting $n = n' + 1$, $t = t' + n'Mt_d$, Equation (6) can be manipulated to derive Equation (7).

$$\begin{aligned} \hat{E}_{\mathrm{out}}(\omega) = & \frac{1}{\sqrt{2\pi}}\int \sum_{m=1}^{M}\sum_{n'=0}^{\infty} T_{\mathrm{e}}{}^{2}R_{\mathrm{e}}{}^{2n'M}E_{\mathrm{out,m}(1)}(t')\exp(-j\omega(t'+n'Mt_{\mathrm{d}}))dt' \\ = & \sum_{n'=0}^{\infty} T_{\mathrm{e}}{}^{2}R_{\mathrm{e}}{}^{2n'M}\exp(-j\omega n'Mt_{\mathrm{d}})\sum_{m=1}^{M}\frac{1}{\sqrt{2\pi}}\int \sum_{n'=1}^{\infty} E_{\mathrm{out,m}(1)}(t')\exp(-j\omega t')dt' \\ = & T_{\mathrm{e}}{}^{2}\sum_{m=1}^{M}\hat{E}_{\mathrm{out,m}(1)}(\omega)\sum_{n'=0}^{\infty}\left(R_{\mathrm{e}}{}^{2M}\exp(-j\omega Mt_{\mathrm{d}})\right)^{n'} \\ = & \frac{T_{\mathrm{e}}{}^{2}}{1-R_{\mathrm{e}}{}^{2M}\exp(-j\omega Mt_{\mathrm{d}})}\sum_{m=1}^{M}\hat{E}_{\mathrm{out,m}(1)}(\omega) \end{aligned} \tag{7}$$

In the integrated cavity absorption spectroscopy system, the Gaussian laser source is most commonly utilized. To obtain $\sum_{m=1}^{M}\hat{E}_{\mathrm{out,m}(1)}(\omega)$, the normalized Gaussian beam field at an arbitrary z-plane is introduced, as described by Equation (8) [23].

$$\begin{aligned} \hat{E}(\omega,x,y,z) = & (\frac{2}{\pi})^{1/2}\frac{\hat{q}_0}{\omega_0\hat{q}(z)}\exp(-jkz-jk\frac{x^2+y^2}{2\hat{q}(z)}) \\ = & (\frac{2}{\pi})^{1/2}\frac{\exp(-jkz+j\phi(z))}{\omega(z)}\exp(-\frac{x^2+y^2}{\omega^2(z)}-jk\frac{x^2+y^2}{2R(z)}) \end{aligned} \tag{8}$$

In the aforementioned equation, $\hat{q}_0$ represents the complex radius of curvature at the beam waist, as shown in Equation (9), while $\hat{q}(z)$ denotes the complex radius of curvature at the z-plane, as indicated in Equation (10). ω_0 is the beam waist of the Gaussian beam, $R(z)$ is the wavefront radius at a distance z from the beam waist, and k is the wavenumber of the laser, as expressed in Equation (11).

$$\hat{q}_0 = j\frac{\pi\omega_0^2}{\lambda} \tag{9}$$

$$\hat{q}(z) = \hat{q} + z = z + jz_R \tag{10}$$

$$\phi(z) = \tan^{-1}(z/z_R) \tag{11}$$

In Equations (9)–(11), z represents the position relative to the beam waist, z_R is the Rayleigh length, and λ is the wavelength. The ABCD matrix is utilized to describe paraxial optical propagation systems. For off-axis optical systems, due to the incident light having a certain angle of inclination and position of incidence, a complex displacement needs to be introduced into Equation (8). The revised equation is shown in Equation (12) [24]:

$$\hat{E}(\omega, x, y, z) = (\frac{2}{\pi})^{1/2} \frac{\hat{q}_0}{\omega_0 \hat{q}(z)} \exp(-jkz - jk\frac{(x-\hat{p}_x)^2 + (y-\hat{p}_y)^2}{2\hat{q}(z)}) \tag{12}$$

In the aforementioned equation, $\hat{p}_x$ and $\hat{p}_y$ denote the complex displacements along the x-axis and y-axis directions, respectively, as illustrated in Equations (13) and (14).

$$\hat{p}_x = x_d - jx_0 \tag{13}$$

$$\hat{p}_y = y_d - jy_0 \tag{14}$$

In Equations (13) and (14), $x_0 = x_0' z_R$ and $y_0 = y_0' z_R$, where x_0' and y_0' represent the incident angles in the x-direction and y-direction, respectively. The complex radius of curvature, complex amplitude, and complex displacement at the output mirror surface can be obtained through the application of the ABCD matrix, along with the corresponding values at the incident end, as demonstrated in Equations (15)–(17) [25].

$$\hat{q}_{\mathrm{out,m(n)}} = \frac{A\hat{q}_{\mathrm{in}} + B}{C\hat{q}_{\mathrm{in}} + D} \tag{15}$$

$$\frac{\hat{E}_{\mathrm{out,m(n)}}}{\hat{E}_{\mathrm{in}}} = \frac{1}{A + B/\hat{q}_{\mathrm{in}}} \tag{16}$$

$$\hat{p}_{\mathrm{out,m(n)}} = \frac{\hat{p}_{\mathrm{in}}}{C\hat{q}_{\mathrm{in}} + D} \tag{17}$$

After the incident light passes through the front cavity mirror, it traverses the cavity and reaches the rear cavity mirror. The light is then reflected by the rear mirror, propagates back through the cavity, and returns to the surface of the front mirror. Under the condition that the re-entrant criteria are satisfied, multiple reflections result in the formation of M-uniformly distributed spots on the surface of the rear cavity mirror. The M-fold re-entrant reflections are considered one complete cycle, and the number of re-entrant cycles is denoted as n. Within a single cycle, the M-fold re-entrant reflections produce M-spot distributions on the rear mirror, with m representing the number of arrivals at the rear mirror during a single cycle. The optical field distribution on the rear mirror can be derived from the incident beam's optical field and the optical transfer matrix. The optical transfer matrix is provided in Equation (18).

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = K^{nM+m-1}K' = \left(\begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2/r & 1 \end{bmatrix} \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2/r & 1 \end{bmatrix} \right)^{nM+m-1} \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \tag{18}$$

In the above equation, K represents the single-pass optical round-trip propagation matrix, while K' denotes the optical propagation matrix from the front cavity mirror to the rear cavity mirror. d is the cavity length, and r is the radius of curvature of the high-reflectivity mirror. The output optical field of the m-th light spot on the rear cavity mirror during the n-th cycle is expressed as shown in Equation (19).

$$\hat{E}_{\mathrm{out,m(n)}}(\omega,x,y,z) = \hat{E}_{\mathrm{out,m(n)}} {T_e}^2 {R_e}^{(n-1)M+m-1} \exp(-jkd - jk\cdot 2((n-1)M+m-1)d)\cdot \exp(-jk\frac{(x-\hat{p}_{\mathrm{x,out,m(n)}})^2+(x-\hat{p}_{\mathrm{y,out,m(n)}})^2}{2\hat{q}_{\mathrm{out,m(n)}}}) \quad (19)$$

In the above equation, $\hat{E}_{\mathrm{out,m(n)}}$, $\hat{p}_{\mathrm{x,out,m(n)}}$, $\hat{p}_{\mathrm{x,out,m(n)}}$ and $\hat{q}_{\mathrm{out,m(n)}}$ can all be derived from Equations (8)–(17). By substituting Equation (19) into Equation (7), one can obtain $\hat{E}_{\mathrm{out}}(\omega)$.

$$\hat{E}_{\mathrm{out}}(\omega) = \frac{{T_e}^2}{1-{R_{\mathrm{e}}}^{2M}\exp(-j\omega M t_{\mathrm{d}})}\sum_{m=1}^{M}\hat{E}_{\mathrm{out,m(1)}}(\omega) \quad (20)$$

Therefore, the transmittance of the resonant cavity, as expressed in Equation (21), is defined as the ratio of the output power to the input power of the resonant cavity.

$$T_{cav}(v) = \frac{P_{\mathrm{out}}(v)}{P_{\mathrm{in}}(v)} = \frac{\iint I_{\mathrm{out}}dxdy}{\iint \mathrm{I_{in}}dxdy} = \frac{\iint \hat{E}_{\mathrm{out}}(\omega)\hat{E}^*_{\mathrm{out}}(\omega)dxdy}{\iint \hat{E}_{\mathrm{in}}(\omega)\hat{E}^*_{\mathrm{in}}(\omega)dxdy} \quad (21)$$

In the above equations, I_{in} and I_{out} represent the incident and transmitted light intensities, respectively. Integrating these intensities over the area of the rear cavity mirror yields the power, with the calculation accuracy depending on the granularity of the grid division applied to the rear mirror. To obtain the direct absorption spectroscopy (DAS) signal and WMS signal from the output of the off-axis integrated cavity system, two modulation schemes are applied to the incident laser. A low-frequency sawtooth wave modulation generates the DAS signal, while a combination of low-frequency sawtooth wave and high-frequency sinusoidal wave modulation produces the WMS signal. The modulated signals of power and wavenumber in both cases are expressed in Equations (22)–(25).

$$v_{\mathrm{das}}(t) = v_{\mathrm{c}} + \Delta v\cdot \mathrm{sawtooth}(t,1) \quad (22)$$

$$I_{\mathrm{das}}(t) = I_0(1 + p_\Omega\cdot\Delta v\cdot \mathrm{sawtooth}(t,1)) \quad (23)$$

$$v_{\mathrm{wms}}(t) = v_{\mathrm{c}} + \Delta v\cdot \mathrm{sawtooth}(t,1) + m\Delta v_{\mathrm{line}}\cos(2\pi f t) \quad (24)$$

$$I_{\mathrm{wms}}(t) = I_0[1 + p_\Omega\cdot\Delta v\cdot \mathrm{sawtooth}(t,1)] + p_{\mathrm{w}}\cdot m\Delta v_{\mathrm{line}}\cos(2\pi f t) \quad (25)$$

Here, sawtooth(t,1) represents a low-frequency sawtooth carrier function with unit amplitude, $\Delta\nu$ denotes the range of wavenumber variation, f is the frequency of the high-frequency sinusoidal modulating wave, p_Ω and p_{w} are the optical cavity conversion coefficients under low-frequency and high-frequency modulation respectively, and m is the modulation index, which is equal to $\Delta\nu/\Delta\nu_{\mathrm{line}}$. The absorption line shape of the gas can be derived from Equation (26):

$$g(v(t),v_g) = \frac{\Delta v_{\mathrm{line}}}{\pi}\frac{1}{(\Delta v_{\mathrm{line}})^2 + (v(t)-v_{\mathrm{g}})^2} \quad (26)$$

By combining Lambert–Beer's law with Equations (21)–(23), and (26), the transmission signal under DAS can be obtained, as shown in Equation (27):

$$I_{\mathrm{das,trans}}(v_0(t)) = \int_0^{+\infty} I_{\mathrm{das}}\varphi(v,v_0)T_{\mathrm{cav}}(v)\exp[-g(v-v_g)SN_{\mathrm{gas}}L]dv \quad (27)$$

In the above equation, S represents the line strength of gas absorption, N_{gas} denotes the gas concentration, and L stands for the effective absorption path length of the off-axis integrated cavity. After detection by a photodetector, the electrical signal of DAS can be obtained.

$$A_{\mathrm{das,trans}}(t) = K_{\mathrm{det}}G_{\upsilon}I_{\mathrm{das,trans}}(t) \quad (28)$$

In Equation (28), K_{det} represents the detection efficiency of the photodetector, and G_v denotes the photoelectric conversion gain. By performing a Fourier series expansion on the absorption line shape function of the gas, the following can be obtained:

$$\begin{aligned} g(v(t), v_g) &= \frac{\Delta v_{\text{line}}}{\pi} \frac{1}{(\Delta v_{\text{line}})^2 + (v(t) - v(g))^2} \\ &= \frac{1}{\pi \Delta v_{\text{line}}} \cdot \frac{1}{1 + (\Delta v \cdot \text{sawtooth}(t,1)/\Delta v_{\text{line}} - m\cos(\omega t)^2} \\ &= \frac{1}{\pi \Delta v_{\text{line}}} \cdot \frac{1}{1 + (x_c - m\cos(\omega t)^2} \end{aligned} \tag{29}$$

where $x_c = \Delta v \cdot \text{sawtooth}(t,1)/\Delta v_{\text{line}}$ denotes the dimensionless deviation between the static wavenumber of the laser output and the wavenumber at the center of the gas absorption line. By performing a Fourier series expansion on Equation (29), the following expression is obtained:

$$g(v(t), v_{\text{abs_c}}) = \frac{1}{\pi \Delta v_{\text{line}}} \cdot \sum_{u=0}^{\infty} H_u(m, x_c) \cos(u\omega t) \tag{30}$$

Based on the relevant Fourier expansion calculations, the correlation coefficient can be derived as:

$$H_0(m, x_c) = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{1}{1 + (x_c - m\cos(\omega t))^2} d(\omega t) \tag{31}$$

$$H_h(m, x_c) = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{1}{1 + (x_c - m\cos(h\omega t))^2} \cos(h\omega t) d(h\omega t) \tag{32}$$

The Fourier coefficients $H_1(m, x_c)$ and $H_1(m, x_c)$ can be calculated, as shown in Equations (33) and (34) [26].

$$H_1(m, x_c) = \frac{\sqrt{2}}{m} \cdot \frac{-x_c \sqrt{\sqrt{Q^2 + 4{x_c}^2} + Q}}{\sqrt{Q^2 + 4{x_c}^2}} + \frac{\text{sign}(x_c) \sqrt{\sqrt{Q^2 + 4{x_c}^2} - Q}}{\sqrt{(Q)^2 + 4{x_c}^2}} \tag{33}$$

$$\begin{aligned} H_2(m, x_c) = \; & -\tfrac{4}{m^2} + \tfrac{\sqrt{2}}{m^2} \cdot \tfrac{(\sqrt{Q^2+4{x_c}^2}+1-x_c)\sqrt{\sqrt{Q^2+4{x_c}^2}+1-{x_c}^2+m}}{\sqrt{Q^2+4{x_c}^2}} \\ & + \tfrac{2|x_c|\sqrt{\sqrt{Q^2+4{x_c}^2}-Q}}{\sqrt{Q^2+4{x_c}^2}} \end{aligned} \tag{34}$$

In the above equation, $Q = 1 - {x_c}^2 + m^2$. The modulated transmitted light can be obtained from Equations (21) and (24)–(26).

$$I_{\text{wms,trans}}(t) = I_{\text{wms}} \cdot T_{\text{cav}}(v) \exp(-g(v - v_c) S N_{\text{gas}} L) \tag{35}$$

When $-g(v - v_c) S N_{\text{gas}} L \ll 1$ is satisfied, $I_{\text{wms,trans}}(t)$ can be approximated by:

$$\begin{aligned} I_{\text{wms,trans}}(t) &= I_{\text{wms}} \cdot T_{\text{cav}}(v)[1 - g(v - v_c) S N_{\text{gas}} L] \\ &= I_{\text{wms}}(t) T_{\text{cav}}(v)(1 - \frac{S N_{\text{gas}} L}{\pi \Delta v_{\text{line}}} \sum_{u=0}^{\infty} H_{\text{u}}(m, x_c) \cos(u\omega t)) \\ &= I_0 i_1 T_{\text{cav}}(v)[1 - \frac{S N_{\text{gas}} L}{\pi \Delta v_{\text{line}}}((H_0(m, x_c) - i_2 H_1(m, x_c))] \\ &\quad - T_{\text{cav}}(v)[I_0 i_1 \frac{S N_{\text{gas}} L}{\pi \Delta v_{\text{line}}}(H_1(m, x_c) + \frac{i_2}{2} H_2(m, x_c)) - I_0 i_2 (1 - \frac{S N_{\text{gas}} L}{\pi \Delta v_{\text{line}}} H_0(m, x_c))] \cos(\omega t) \\ &\quad - T_{\text{cav}}(v) I_0 i_1 \cdot \frac{S N_{\text{gas}} L}{\pi \Delta v_{\text{line}}}[H_2(m, x_c) + \frac{i_2}{2}(H_1(m, x_c) + H_3(m, x_c))] \cos(2\omega t) + \ldots \end{aligned} \tag{36}$$

Consequently, the fundamental wave and second harmonic signals are, respectively, expressed as:

$$I_{1f}(t) = -I_0 i_1 T_{\text{cav}}(v) \frac{SN_{\text{gas}}L}{\pi \Delta v_{\text{line}}} [H_1(m, x_c) + \frac{i_2}{2} H_2(m, x_c)) - I_0 i_2 (1 - \frac{SN_{\text{gas}}L}{\pi \Delta v_{\text{line}}} H_0(m, x_c)] \quad (37)$$

$$I_{2f}(t) = -I_0 i_1 T_{\text{cav}}(v) \cdot \frac{SN_{\text{gas}}L}{\pi \Delta v_{\text{line}}} [H_2(m, x_c) + \frac{i_2}{2} (H_1(m, x_c) + H_3(m, x_c))] \quad (38)$$

In the above equation, $i_1 = 1 + p_\Omega \cdot \Delta v_{\text{line}} x_c$ and $i_2 = p_w \cdot m \Delta v_{\text{line}}$, which are optical signals transmitted through the cavity, are detected by the photodetector after gas absorption and subsequently converted into electrical signals.

$$A_{\text{wms,trans}}(t) = K_{\text{det}} G_\upsilon I_{\text{wms,trans}}(t) \quad (39)$$

To model the lock-in amplifier for the direct extraction of the first and second harmonic signals of gas absorption under wavelength modulation, as illustrated in Equations (40)–(45):

$$A_{1f,x}(t) = [A_{\text{trans}}(t) \cdot S_{\text{ref}}(2f, 0)]|_{\text{LPF}} \quad (40)$$

$$A_{1f,y}(t) = [A_{\text{trans}}(t) \cdot S_{\text{ref}}(2f, 2\pi)]\Big|_{\text{LPF}} \quad (41)$$

$$A_{2f,R}(t) = \sqrt{A_{2f,x}(t)^2 + A_{2f,y}(t)^2} \quad (42)$$

$$A_{2f,x}(t) = [A_{\text{trans}}(t) \cdot S_{\text{ref}}(2f, 0)]|_{\text{LPF}} \quad (43)$$

$$A_{2f,y}(t) = [A_{\text{trans}}(t) \cdot S_{\text{ref}}(2f, 2\pi)]\Big|_{\text{LPF}} \quad (44)$$

$$A_{2f,R}(t) = \sqrt{A_{2f,x}(t)^2 + A_{2f,y}(t)^2} \quad (45)$$

In the aforementioned equations, $S_{\text{ref}}(1f,0)$ represents the fundamental frequency reference signal with an initial phase of 0, $S_{\text{ref}}(1f,\pi)$ represents the fundamental frequency reference signal with an initial phase of π, Sref$(2f,0)$ represents the second harmonic reference signal with an initial phase of 0, and Sref$(2f,\pi)$ represents the second harmonic reference signal with an initial phase of π.

3. Results and Discussion

3.1. Model Verification

To validate the established integral cavity optical model, a case study involving 11 reentrants was conducted. The self-developed model was compared with the light spot distribution results obtained from the Tracepro 7.3.4 software. The Tracepro simulations were configured with the following critical parameters: the light source was set to operate at 5.68 μm wavelength with a beam waist diameter of 0.907 mm, positioned at the cavity's beam waist location (coordinates: x = 0 mm, y = 5.05 mm, z = 0 mm). The beam was introduced at an incident angle of $-0.636°$ along the x-axis. The optical cavity was constructed using high-reflectance mirrors (thickness: 6.35 mm; surface curvature radius: 1000 mm for both mirrors) with a total cavity length of 58.46 cm. The front mirror's first surface was centered at the origin (0, 0, 0), while the rear mirror's first surface was positioned at (0, 0, 590.93 mm), as shown in Figure 1a,b. The light spot distribution had a diameter of 12 mm, and both the waist position and size were consistent with the cavity dimensions. Consequently, the light spot distribution was uniform, and the simulation results for both models showed good agreement. These results confirm the validity of the self-developed model.

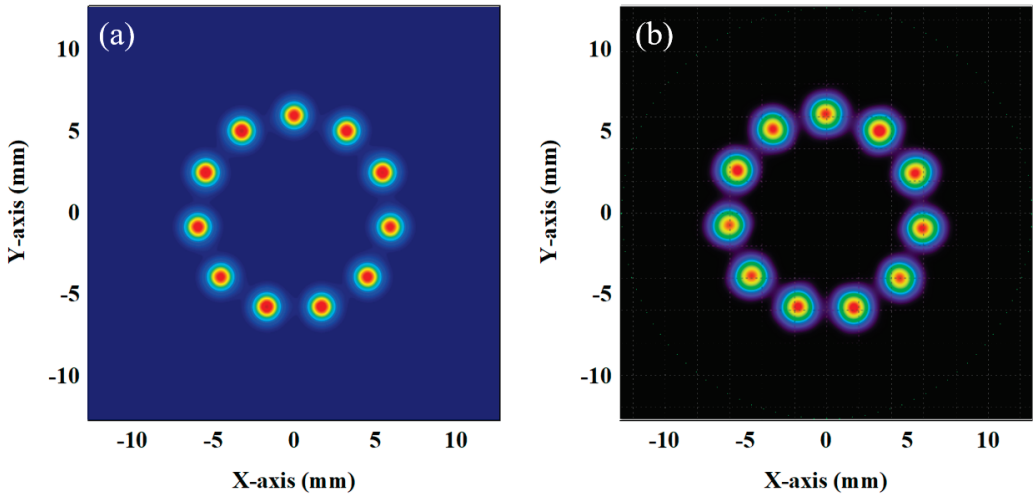


Figure 1. Simulation of off-axis integrated cavity light spot distribution: (**a**) simulation results from a self-developed model; (**b**) simulation results from Tracepro.

3.2. Simulation of Off-Axis Integrating Cavity Under Standard Conditions

Based on the self-developed model, the effects of different system parameters of WMS-OA-ICOS on the measurement signal were investigated. The line strength of the HCHO gas at 25 °C was 6.095×10^{-20} cm^{-1}/molecule·cm^{-2}. First, the optical properties of the laser source were described, and the output characteristics of laser power and wavelength were simulated using Equations (22)–(25). The laser source output characteristics of WMS-OA-ICOS are shown in Figure 2.

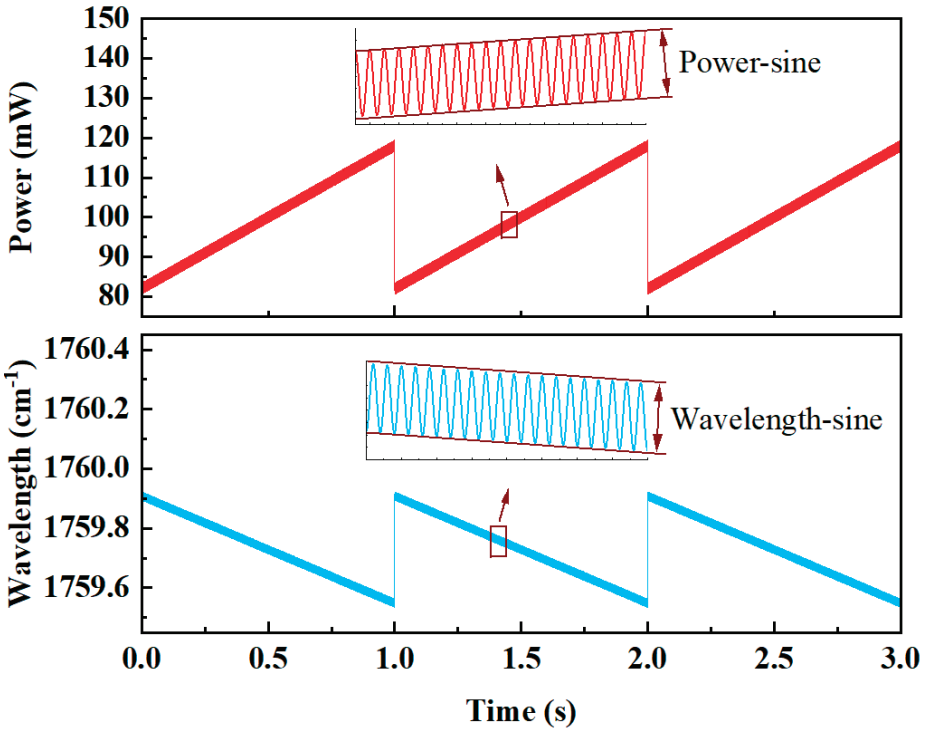


Figure 2. Relationship between power and wave number as a function of scanning time.

Under standard conditions, other simulation parameters are shown in Table 1.

Table 1. Simulation parameters for WMS-OA-ICOS.

Parameter	Value	Parameter	Value
Cavity length	58.46 cm	Center wavelength of laser	5682 nm
Incident coordinates	(3.24 mm, 5.05 mm)	Power and wavelength conversion coefficient of laser	100 mW/cm^{-1}
incident angle	($-0.636°$, 0)	Sawtooth modulation frequency	1 Hz
Waist size	0.907 mm	Sinewave modulation frequency	1 kHz
Girdle position	Cavity center	Laser linewidth	30 MHz
Curvature radius	1000 mm	Modulation depth	2.2
Cavity pressure and environment temperature	5 kPa, 25 °C	Detector photoelectric conversion coefficient	5×10^4 V/W
Reflectivity of mirrors	0.995	Detector detection efficiency	0.1

As shown in Figure 3a, under standard conditions with 11 re-entrant reflections, the position and intensity distribution of the output end mirror's spot are uniform. The simulation results indicate that the free spectral range (FSR) of the optical cavity is approximately 23 MHz, consistent with the value calculated using Equation (2). This consistency further validates the accuracy of the simulation results from another perspective. From the cavity mode simulation results in the lower panel of Figure 3b, it can be observed that the intensity and position distribution of the cavity mode signals are relatively uniform. When high-frequency sinusoidal modulation is disabled, and only low-frequency sawtooth wave scanning of the laser source is applied, continuous output of the transmission signal can be achieved. However, noticeable cavity noise is still observed in the transmission signal in the upper panel of Figure 3b. This may be attributed to an insufficient number of internal reflections, which leads to an inadequate excitation of high-order degenerate modes in the off-axis cavity. The large FSR spacing and uneven coupling between the laser and cavity modes during scanning can cause fluctuations in transmission power. Figure 3b also presents the normalized signal of the 2 ppmv target gas absorption and its Lorentzian fit. The standard deviation of the fitting is 6.5×10^{-5}, and the detection limit is 80.8 ppbv (1σ).

Based on the established optical model, the absorption signal behavior of the WMS-OA-ICOS system under standard conditions was further investigated. High-frequency sinusoidal modulation was enabled, and the direct absorption signal of HCHO gas under wavelength modulation is shown in the upper panel of Figure 4b. This signal was obtained by superimposing a 1 kHz high-frequency modulation on a 1 Hz sawtooth wave scan. The second harmonic (2f) signal demodulated from the direct absorption signal is shown in the middle panel of Figure 4a. Fluctuating noise induced by cavity mode effects is also evident in the 2f signal. However, it can be readily observed that the noise characteristics in the 2f signal differ from those in the direct absorption signal depicted in Figure 3b. In the direct absorption signal, cavity mode noise exhibits high temporal consistency, whereas in the wavelength modulation signal, the envelope of cavity mode noise shows a distinct periodic variation. This difference is attributed to the different coupling mechanisms between the laser and the cavity modes in the two scenarios. This observation suggests the potential for suppressing cavity mode noise by altering the coupling mechanism. This aspect will be further explored in subsequent sections based on the WMS-OA-ICOS model. The lower panel of Figure 4a shows the 1f signal, which is used to correct the intensity of the 2f signal. After correction, the peak-to-peak value is 2.055, corresponding to a detection sensitivity of

44.2 ppbv (1σ). Figure 4b illustrates the relationship between the corrected 2f simulation signal's peak-to-peak value and the target gas concentration. A strong correlation is observed, with the fitted R^2 value reaching 0.9993, indicating excellent linearity.

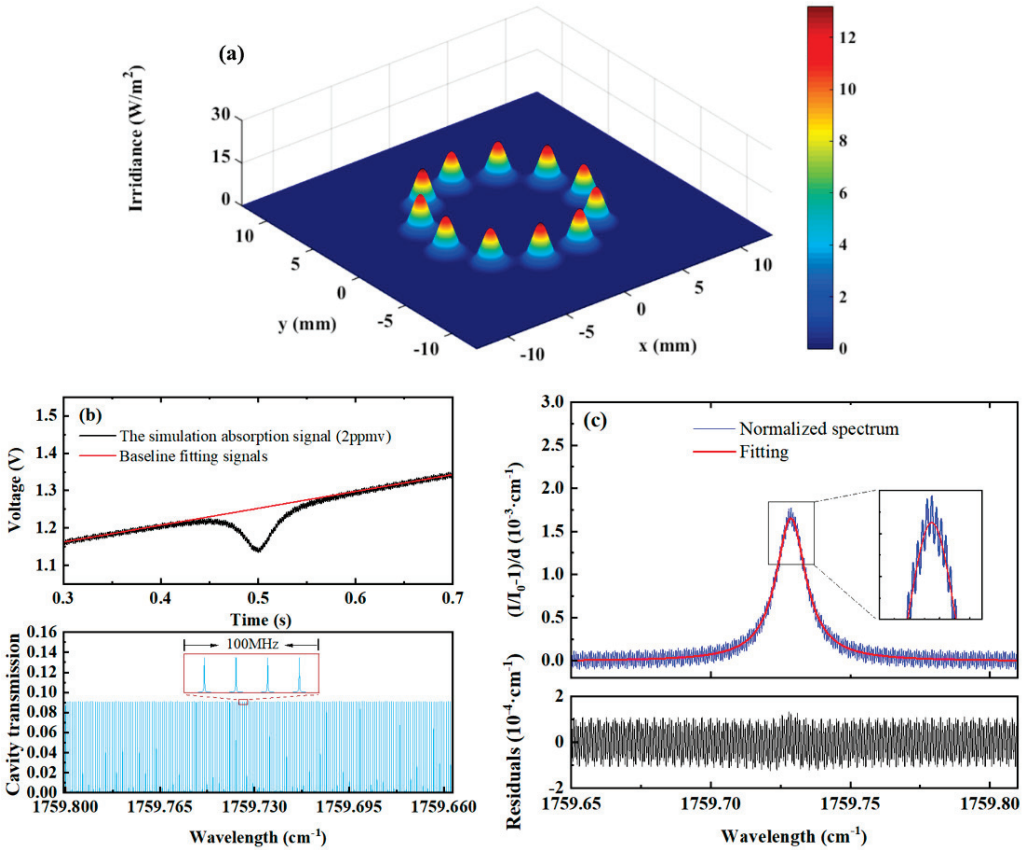


Figure 3. Simulation of spot distribution and direct absorption under standard conditions. (**a**) Three-dimensional representation of spot distribution, (**b**) direct absorption signal of a target gas at 2 ppmv, (**c**) normalized and fitted absorption signal for a target gas at 2 ppmv.

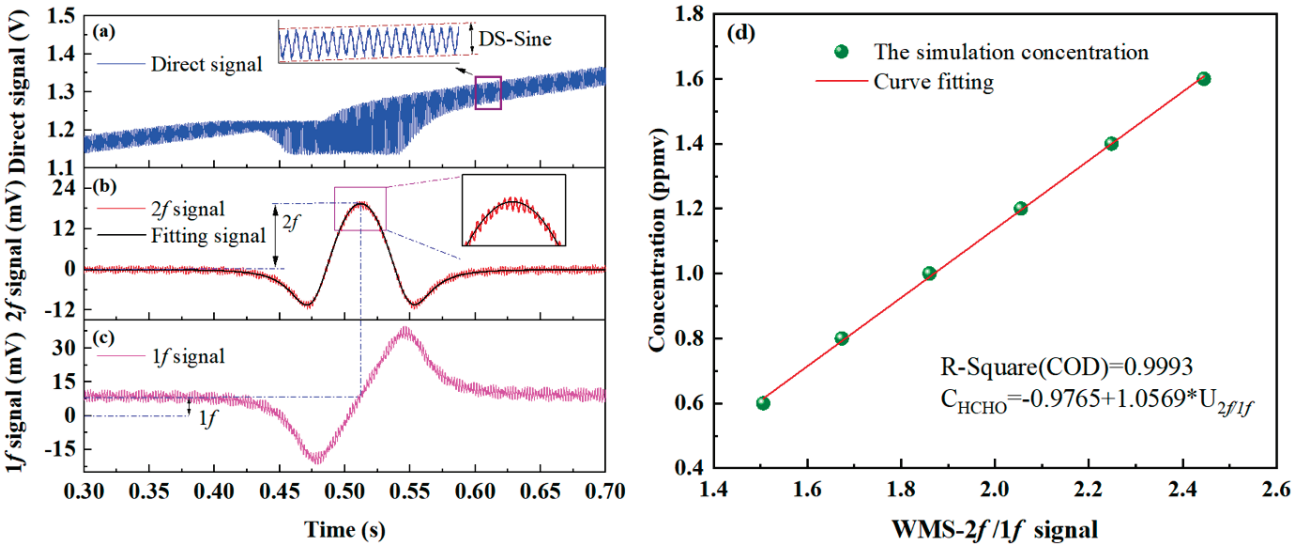


Figure 4. Wavelength modulation simulations under standard conditions: (**a**) absorption signal with high-frequency modulation, (**b**) 2f signal, (**c**) 1f signal, and (**d**) calibration fit between the target gas concentration and the 2f/1f ratio.

3.3. Impact of Mode-Matching on the WMS-OA-ICOS System

The interference effects of beams with different characteristic parameters within the resonant cavity can vary, potentially leading to different impacts on the 2f signal of the WMS-OA-ICOS system. Therefore, this issue was further investigated based on the self-developed model. Typically, the waist size of the laser source is determined during production. To explore the influence of the matching relationship between the laser source beam waist and the resonant cavity on measurement results, the beam waist of the resonant cavity was first calculated using the formula provided in Equation (46).

$$w_0 = \sqrt{\frac{d\lambda}{\pi}}\left[\frac{g_1 g_2(1-g_1 g_2)}{(g_1+g_2-2g_1 g_2)^2}\right]^{0.25} \tag{46}$$

In the equation, $g_1 = 1 - \mathrm{d}/\mathrm{R}_1$, $g_2 = 1 - \mathrm{d}/\mathrm{R}_2$, where λ represents the wavelength of the incident laser, d denotes the length of the multipass gas cell, and R_1 and R_2 are the radii of curvature of the two high-reflectance mirrors, respectively. Under standard conditions with 11 times re-entrant light, the length of the off-axis integrated cavity is chosen to be 58.46 cm, at which point the beam waist size of the resonant cavity is 0.907 mm. With a step size of 0.2 mm, simulations were conducted to investigate the distribution of the laser spot on the mirrors of the integrated cavity for beam waist sizes of 0.307 mm, 0.507 mm, 0.707 mm, 1.107 mm, 1.307 mm, and 1.507 mm, as illustrated in Figure 5a–f. A comparison between the beam spot distribution and cavity mode signals under the standard waist size in Figure 3 clearly demonstrates that increased deviation between the laser beam waist size and the resonant cavity waist size leads to progressively non-uniform spot distribution on the mirror.

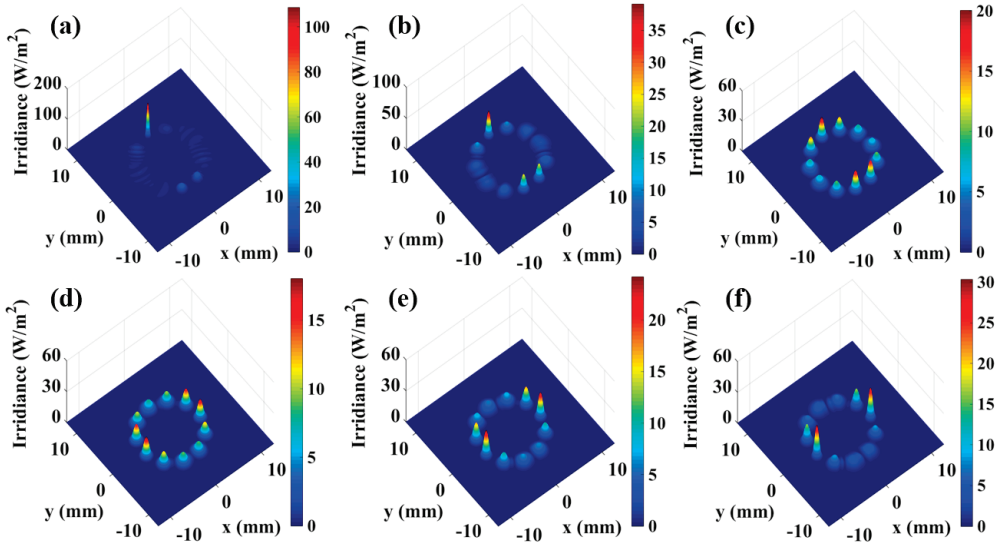


Figure 5. Spot distribution with different beam waist sizes. (**a**) Spot distribution with a beam waist size of 0.307 mm, (**b**) spot distribution with a beam waist size of 0.507 mm, (**c**) spot distribution with a beam waist size of 0.707 mm, (**d**) spot distribution with a beam waist size of 1.107 mm, (**e**) spot distribution with a beam waist size of 1.307 mm, (**f**) spot distribution with a beam waist size of 1.507 mm.

Figure 6 displays the cavity mode signals and transmission signals under varying beam waist sizes. As illustrated in Figure 6a–f, the uniformity of cavity mode intensity deteriorates progressively as the beam waist deviates from the ideal size of 0.907 mm. This pattern is similarly reflected in the transmission signals of the integrating cavity, as shown in Figure 6g–l. Combined with the simulation results under standard conditions in Figure 3, it is evident that deviations of the actual beam waist size from the ideal value result in significant degradation of the transmission signal quality.

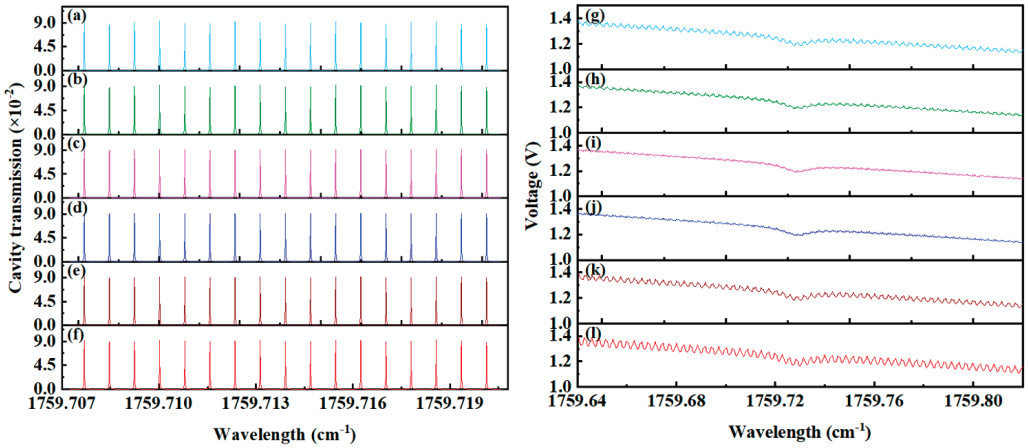


Figure 6. Cavity mode signals and transmitted signal with different beam waist sizes. (**a**) Cavity mode with a beam waist size of 0.307 mm, (**b**) cavity mode with a beam waist size of 0.507 mm, (**c**) cavity mode with a beam waist size of 0.707 mm, (**d**) cavity mode with a beam waist size of 1.107 mm, (**e**) cavity mode with a beam waist size of 1.307 mm, (**f**) cavity mode with a beam waist size of 1.507 mm, (**g**) transmitted signal of 0.307 mm, (**h**) transmitted signal of 0.507 mm, (**i**) transmitted signal of 0.707 mm, (**j**) transmitted signal of 1.107 mm, (**k**) transmitted signal of 1.307 mm, (**l**) transmitted signal of 1.507 mm.

Simulations of 2f signals formed by beams with different waist sizes undergoing stable resonance after absorption by a target gas at the same concentration are shown in Figure 7a. The relationships between the peak-to-peak value and SNR of the 2f signal and the beam waist size are illustrated in Figure 7b. It can be observed that the 2f signal peak is minimally affected by variations in the laser source beam waist size. However, the SNR increases initially and then decreases as the beam waist size increases, reaching a peak at 0.907 mm. This indicates that inconsistencies between the laser source beam waist size and the resonant cavity beam waist size directly lead to a reduction in SNR. Therefore, in the practical design of WMS-OA-ICOS systems, it is crucial to ensure proper matching between the waist of the laser source and the resonant cavity to avoid significant deviations between the incident laser beam waist and the cavity beam waist.

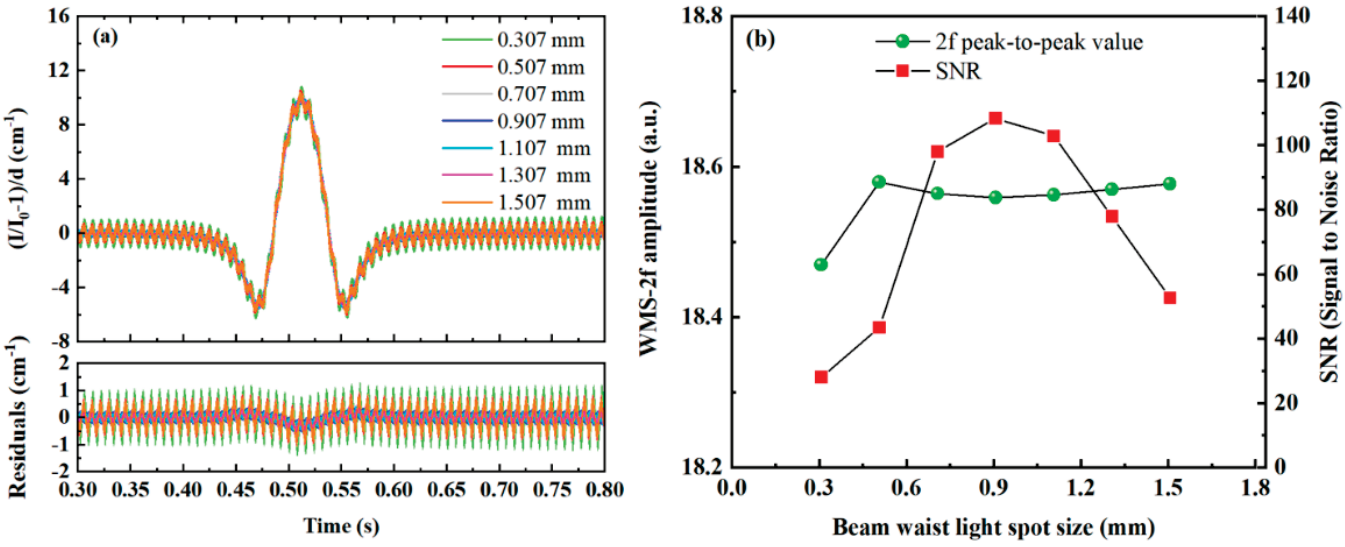


Figure 7. Effects of different beam waist sizes on the second harmonic absorption signal: (**a**) second harmonic signals for various beam waist sizes; (**b**) relationships between the peak-to-peak value and SNR of the second harmonic signal and the beam waist size.

3.4. Influence of Cavity Length on WMS-OA-ICOS System

In WMS-OA-ICOS, re-entrance is achieved by adjusting the incident position and angle based on the pre-defined cavity length of the integrated cavity. When the conditions for re-entrance are satisfied, a uniform and stable light field distribution is formed within the cavity, enabling the excitation of uniform cavity modes. However, changes in the cavity length can affect the light field distribution within the cavity. To investigate the effects of cavity length on the mirror light field distribution, transmitted cavity modes, and 2f signals in the integrated cavity, simulations were conducted using the WMS-OA-ICOS model. Under standard conditions with 11 re-entrances, the off-axis integrated cavity length was set to 58.46 cm. Simulations were performed for cavity lengths of 58.16 cm, 58.26 cm, 58.36 cm, 58.56 cm, 58.66 cm, and 58.76 cm with a step size of 0.1 cm. The corresponding mirror spot distributions are shown in Figure 8a–f. A comparison of the beam spot distribution and cavity mode signals under the standard waist size in Figure 3 clearly demonstrates that as the deviation between the laser beam waist size and the resonant cavity waist size increases, the spot distribution on the mirror becomes progressively non-uniform, with overlapping between spots becoming evident.

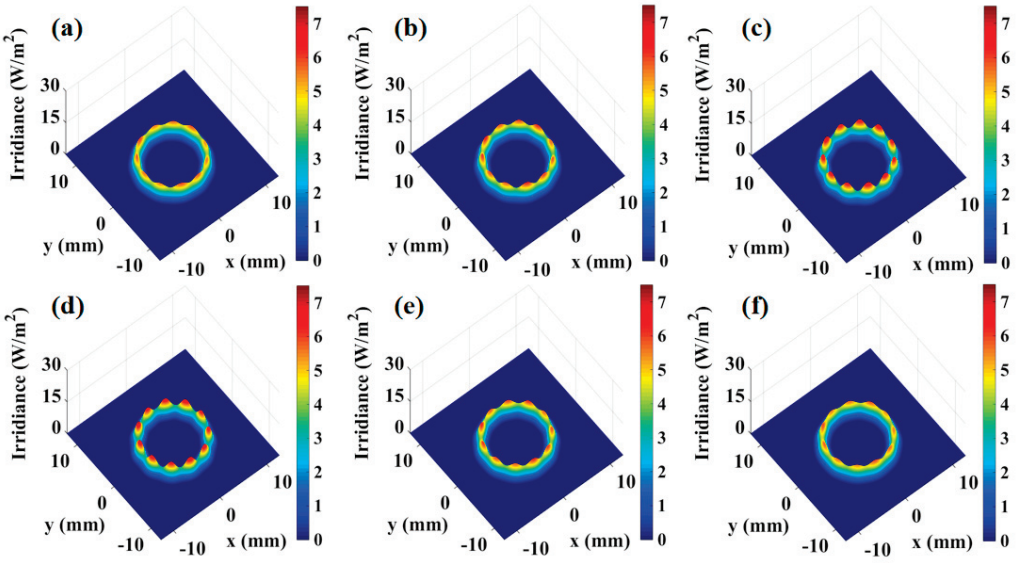


Figure 8. Spot distribution at different cavity lengths: (**a**) spot distribution at a cavity length of 58.16 cm, (**b**) spot distribution at a cavity length of 58.26 cm, (**c**) spot distribution at a cavity length of 58.36 cm, (**d**) spot distribution at a cavity length of 58.56 cm, (**e**) spot distribution at a cavity length of 58.66 cm, (**f**) spot distribution at a cavity length of 58.76 cm.

Figure 9 displays the cavity mode signals and transmission signals under different cavity lengths. As shown in Figure 9a–f, deviations from the ideal cavity length of 58.46 cm lead to a progressive deterioration in the uniformity of cavity mode intensity, accompanied by the excitation of an increasing number of higher-order modes. In the ideal configuration, the off-axis integrating cavity—compared to its coaxial counterpart—excites uniform higher-order modes due to the multi-spot distribution under the "re-entrant" condition. This results in an equal partitioning of both the free spectral range (FSR) width and the cavity mode intensities, in contrast to the coaxial case. Here, although minor deviations in cavity length induce higher-order cavity modes, these excited higher-order modes remain spectrally adjacent to the fundamental mode, and the total cavity mode intensity remains consistent with the ideal scenario. Consequently, during the wavelength tuning of the incident laser, the convolution between the laser spectrum and the cavity modes is minimally affected. This behavior is corroborated by the transmission signals shown in Figure 9g–l.

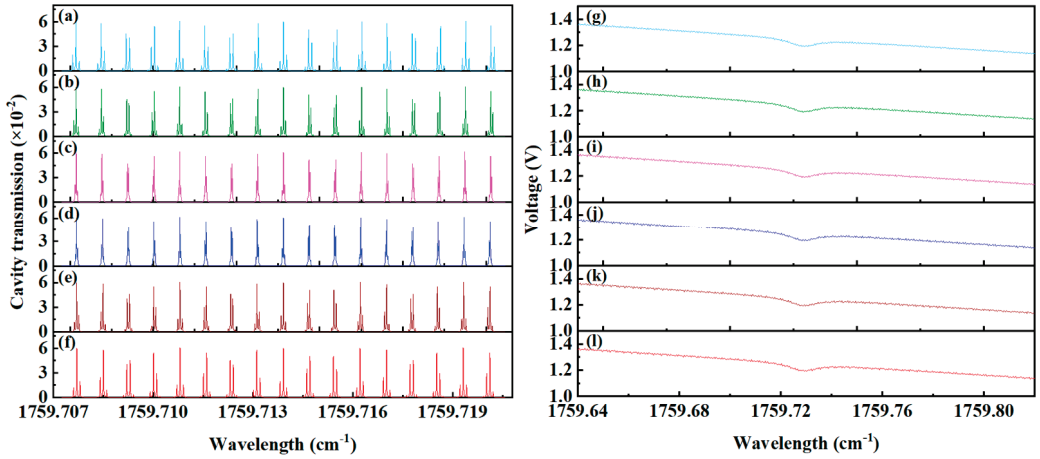


Figure 9. Cavity mode signals and transmitted signal at different cavity lengths: (**a**) cavity mode at a cavity length of 58.16 cm, (**b**) cavity mode at a cavity length of 58.26 cm, (**c**) cavity mode at a cavity length of 58.36 cm, (**d**) cavity mode at a cavity length of 58.56 cm, (**e**) cavity mode at a cavity length of 58.66 cm, (**f**) cavity mode at a cavity length of 58.76 cm, (**g**) transmitted signal at a cavity length of 58.16 cm, (**h**) transmitted signal at a cavity length of 58.26 cm, (**i**) transmitted signal at a cavity length of 58.36 cm, (**j**) transmitted signal at a cavity length of 58.56 cm, (**k**) transmitted signal at a cavity length of 58.66 cm, (**l**) transmitted signal at a cavity length of 58.76 cm.

Compared to the spot distribution and cavity mode signals under the standard cavity length shown in Figure 3, Figure 10 demonstrates that as the cavity length deviates further from the standard length, interference between spots becomes more pronounced, leading to the excitation of additional higher-order cavity modes. Simulations of the 2f signals generated by beams undergoing stable resonance after absorption by a target gas of the same concentration under varying cavity lengths are shown in Figure 10a. The relationships between the 2f signal peak-to-peak value, SNR, and the integrated cavity length are illustrated in Figure 10b. Both the peak-to-peak value and SNR of the 2f signal are minimally affected by changes in the cavity length. This is because the linewidth of the laser source is significantly broader than the free spectral range (FSR) of the off-axis integrated cavity with 11 re-entrant times. Even though the cavity length variation excites more higher-order cavity modes, the total cavity mode intensity within each FSR remains unaffected, resulting in negligible impact on coupling strength. On a finer scale, differences in the convolution process between the laser spectrum of the laser source and the higher-order cavity modes under different cavity lengths introduce slight variations in SNR. The optimal SNR depends on multiple factors, including the number of re-entrant times, the types of higher-order cavity modes excited, mode spacing, and the linewidth of the laser source. Therefore, in practical applications, millimeter-scale manufacturing tolerances near the standard cavity length have an extremely small effect on measurement results and can be considered negligible.

It is noteworthy that significant deviation of the cavity length from the standard value will lead to remarkable degradation in signal quality. For instance, when the cavity length measures 55.96 cm (representing a −2.50 cm deviation from the standard cavity length), the standard deviation of 2f signal fitting residuals increases to 0.665. Compared with the standard cavity length condition (1σ = 0.171), this corresponds to a 3.88-fold amplification of noise. This phenomenon occurs because substantial higher-order cavity modes with varying intensities become excited during cavity length variations, completely deviating from the original fundamental mode position. During the scanning process of the laser source, the light source couples with different cavity modes, manifesting as intense fluctuations in both transmission signals and 2f signals, as illustrated in Figure 11.

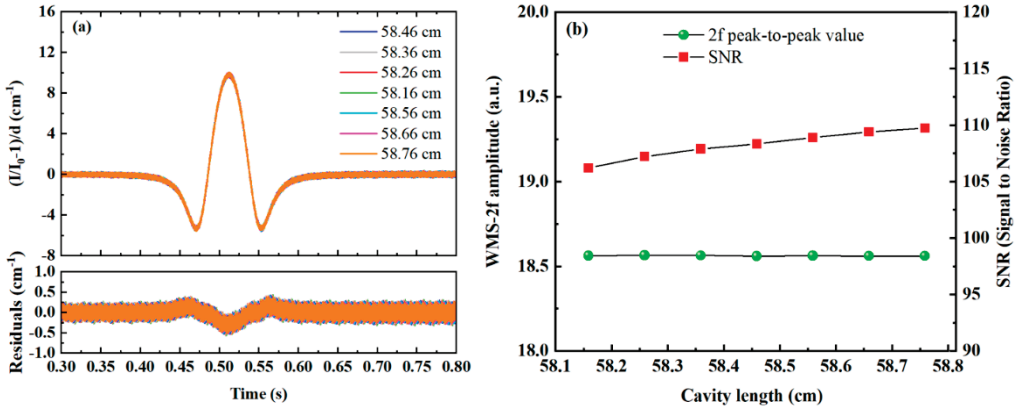


Figure 10. Influence of various cavity lengths on the second harmonic absorption signal: (**a**) second harmonic signals at different cavity lengths, (**b**) relationship between the peak-to-peak value and SNR of the second harmonic signal and the cavity length.

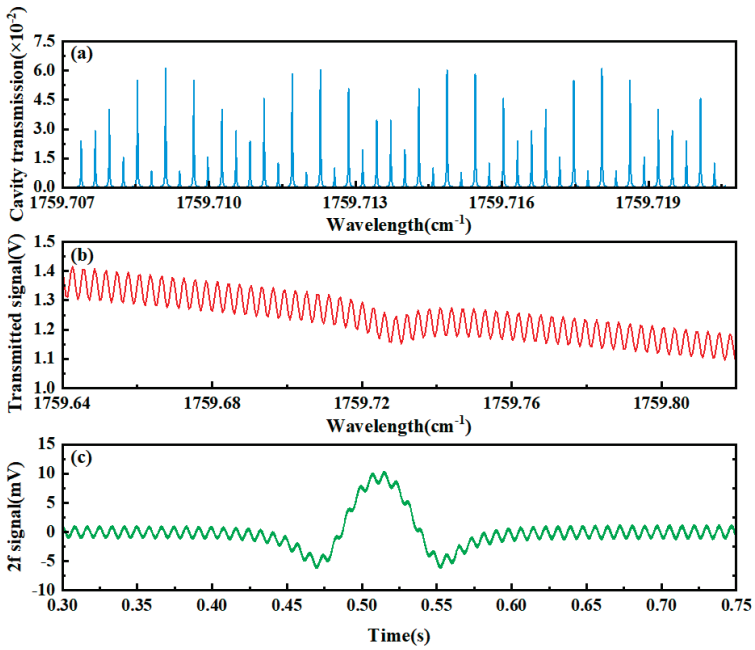


Figure 11. Cavity modes and absorption characteristics under a cavity length of 55.96 cm. (**a**) Cavity mode signals, (**b**) transmission signals, (**c**) second harmonic signals.

3.5. Impact of Modulation Coefficient on the Performance of WMS-OA-ICOS Systems

In tunable absorption spectroscopy systems, the issue of cavity mode coupling is absent, and typically, the maximum peak-to-peak value of the second harmonic can be obtained at a modulation coefficient of 2.2 [27]. However, in WMS-OA-ICOS systems, the modulation signal influences the convolution of the laser source spectrum with the modes of the off-axis integrated cavity. Therefore, the impact of the modulation coefficient on the second harmonic absorption signal is further investigated by setting different modulation coefficients. Figure 12a shows the second harmonic gas absorption signals at various modulation coefficients, while Figure 12b illustrates the relationship between the peak-to-peak value and SNR of the second harmonic signal and the modulation coefficient. It can be observed that the peak-to-peak value of the second harmonic signal first increases and then decreases as the modulation coefficient increases, peaking at a modulation coefficient of 2.2, which is consistent with the results observed in tunable absorption spectroscopy systems. However, when considering noise induced by cavity mode effects, the SNR initially increases with the modulation coefficient, reaching a peak at a modulation co-

efficient of 2.2 before declining. As the modulation coefficient reaches 3, the SNR of the system begins to improve again, reaching a second peak when the modulation coefficient reaches 3.8, followed by a downward trend. This phenomenon arises because, although the amplitude of the second harmonic signal exhibits a monotonically decreasing trend when the modulation coefficient exceeds 2.2, the noise intensity induced by the coupling between the incident laser and the cavity mode varies with the modulation depth. When m = 3.8, the standard deviation of the fitting residual for the second harmonic signal is 0.1654, while when m = 3.4 and m = 4.2, the corresponding standard deviations are 0.1719 and 0.1656, respectively. Consequently, a second peak emerges at m = 3.8.

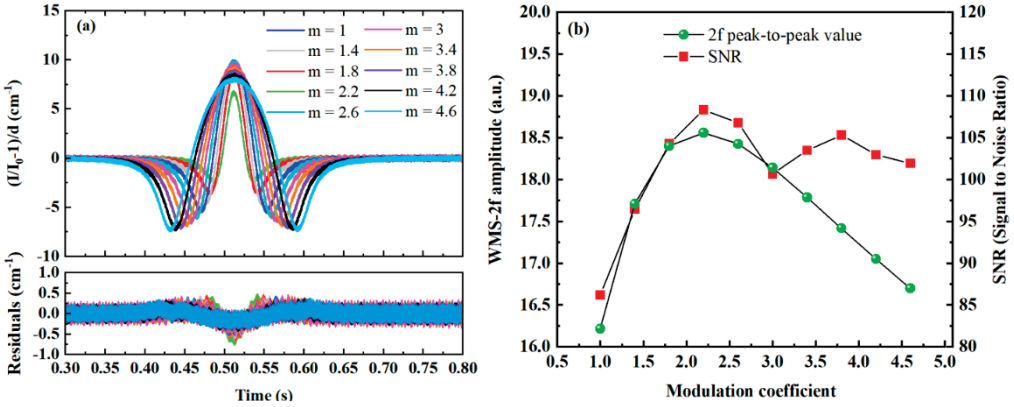


Figure 12. Influence of different modulation coefficients on the second harmonic absorption signal; (**a**) second harmonic signals at various modulation coefficients; (**b**) relationship between the peak-to-peak value and SNR of the second harmonic signal and the modulation coefficient.

3.6. Impact of Laser Source Linewidth on the Performance of WMS-OA-ICOS Systems

In WMS-OA-ICOS systems, cavity mode effects introduce periodic interference noise in the 2f signal. Reducing the free spectral range (FSR) can partially mitigate this issue. This is because a narrower FSR allows the laser to simultaneously overlap with more cavity modes, minimizing fluctuations during wavelength scanning. The progression from on-axis to off-axis measurement leverages this principle. According to the FSR calculation formula (see Equation (2)), increasing the number of off-axis re-entries and extending the cavity length both lead to a narrower FSR. However, in many practical applications, constraints on instrument size limit the ability to increase cavity length to enhance cavity mode density. Increasing the number of off-axis re-entries effectively excites a larger number of high-order cavity modes, thereby narrowing the FSR. However, this also reduces the intensity of cavity mode signals, which in turn affects the intensity of the transmitted signal. Therefore, when the FSR is fixed, an alternative approach is to increase the spectral linewidth of the laser source to improve its overlap with cavity modes. The second harmonic gas absorption signals under different laser linewidths are shown in Figure 13a, and the relationships between the 2f signal peak-to-peak value, SNR, and laser linewidth are illustrated in Figure 13b.

As shown in Figure 13b, the SNR increases initially and then decreases with the increase in laser linewidth, reaching a peak at approximately 240 MHz. The reason for this trend is that although interference noise in the system is significantly reduced as the linewidth increases, the intensity of the 2f signal also gradually decreases. The decline in signal intensity becomes particularly pronounced when the linewidth exceeds 60 MHz. Moreover, the figure reveals that when the laser linewidth is smaller than the free spectral range (FSR), the absorption signal exhibits significant noise. However, when the linewidth increases to 30 MHz, this noise is effectively suppressed. This is because a linewidth of 30 MHz, which is substantially larger than the FSR of the cavity (23 MHz), enables

continuous cavity mode scanning. These findings indicate that, in practical applications, system cavity mode noise can be effectively mitigated by appropriately broadening the laser linewidth. The optimal linewidth broadening is directly determined by the FSR of the cavity.

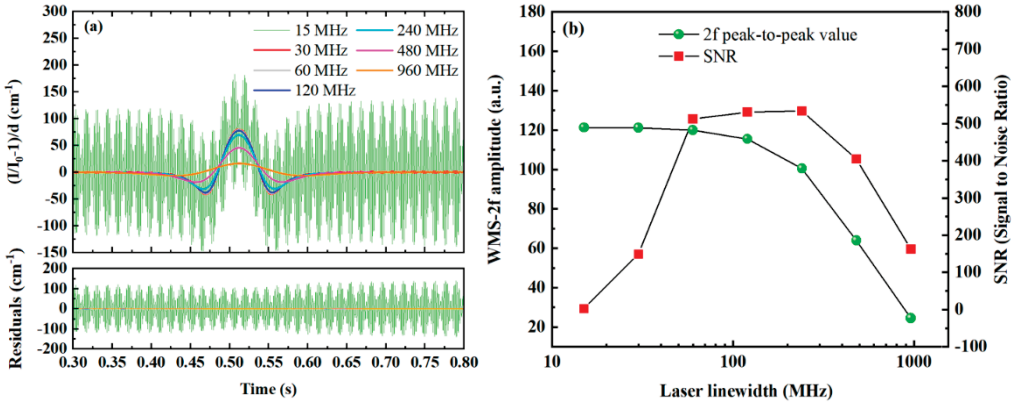


Figure 13. Influence of different laser linewidths on second harmonic absorption signals: (**a**) second harmonic signals at various laser linewidths, (**b**) relationships between the peak-to-peak value and SNR of the second harmonic signal and the laser linewidth.

4. Summary

This study proposes a wavelength modulation off-axis integrated cavity output spectroscopy (WMS-OA-ICOS) model to investigate the effects of mode matching, cavity length variations, modulation index, and laser linewidth on the quality of 2f gas absorption signals. Initially, the WMS-OA-ICOS model was established through theoretical analysis, and the simulated beam profile results were compared with those obtained using TracePro to validate the accuracy and reliability of the model. Based on simulations of 2f absorption signals under varying characteristic parameters, the following conclusions can be drawn:

(1) In the WMS-OA-ICOS system, the degree of mode-matching between the laser source beam waist and the cavity waist significantly impacts the SNR of the 2f absorption signal. Hence, in practical system design, it is crucial to avoid large deviations between the beam waist of the incident laser and the cavity waist.

(2) When the cavity length deviates from the standard value by millimeter-scale errors, the SNR of the 2f signal is minimally affected. Since the manufacturing and assembly of the cavity generally meet such tolerance requirements, this factor does not require primary consideration in practical applications.

(3) The maximum SNR of the 2f signal is achieved at a modulation index of 2.2.

(4) As the laser linewidth increases, the 2f signal amplitude gradually decreases. However, the increase in linewidth effectively suppresses cavity mode noise in the signal. Therefore, appropriately broadening the laser linewidth can serve as an effective strategy for mitigating cavity mode noise in the WMS-OA-ICOS system.

The proposed WMS-OA-ICOS simulation model and its analysis of influencing factors provide valuable guidance for the development and optimization of WMS-OA-ICOS systems in practical applications. This study has great application potential in exoplanet atmosphere exploration missions with small size and high sensitivity detection requirements.

Author Contributions: All the authors made contributions to this work. Conceptualization, T.W. and Y.H.; Funding acquisition, resources and project administration, T.W., X.C., H.X. and W.L.; Investigation, T.W., H.X. and Y.Z.; Methodology and scheme, T.W., Y.H. and X.Z.; Writing—original draft, T.W.; Writing—Revising, T.W. and W.L.; Processing program, D.Z., X.Z. and Z.F.; Optical model

construction T.W., F.W. and L.P.; Validation, D.Z., Z.F., Y.Z. and L.P. 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 (42405133), Key Technology Research Project of Planetary Protection for the Tianwen-3 mission (TW3010), Advanced scientific research project of Deep Space Exploration Laboratory(GC03ZX0001ZC3XT-2325).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Hou, Z.; Liu, J.; Xu, Y.; Pang, F.; Wang, Y.; Qin, L.; Liu, Y.; Zhao, Y.-Y.S.; Wei, G.; Xu, M.; et al. The search for life signatures on Mars by the Tianwen-3 Mars sample return mission. *Natl. Sci. Rev.* **2024**, *11*, nwae313. [CrossRef]
2. Edwards, J.K.; Solsona, B.; Ntainjua, E.; Carley, A.F.; Herzing, A.A.; Kiely, C.J.; Hutchings, G.J. Strong release of methane on mars in northern summer 2003. *Science* **2009**, *323*, 1037–1041. [CrossRef]
3. Horita, J.; Berndt, M.E. Abiogenic methane formation and isotopic fractionation under hydrothermal conditions. *Science* **1999**, *285*, 1055–1057. [CrossRef]
4. Ueno, Y.; Yamada, K.; Yoshida, N.; Maruyama, S.; Isozaki, Y. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **2006**, *440*, 516–519. [CrossRef]
5. Lefevre, F.; Forget, F. Observed variations of methane on Mars unexplained by known atmospheric chemistry and physics. *Nature* **2009**, *460*, 720–723. [CrossRef]
6. Villanueva, G.; Mumma, M.; Novak, R.; Radeva, Y.; Käufl, H.; Smette, A.; Tokunaga, A.; Khayat, A.; Encrenaz, T.; Hartogh, P. A sensitive search for organics (CH_4, CH_3OH, H_2CO, C_2H_6, C_2H_2, C_2H_4), hydroperoxyl (HO_2), nitrogen compounds (N_2O, NH_3, HCN) and chlorine species (HCl, CH_3Cl) on Mars using ground-based high-resolution infrared spectroscopy. *Icarus* **2013**, *223*, 11–27. [CrossRef]
7. Liuzzi, G.; Villanueva, G.L.; Mumma, M.J.; Smith, M.D.; Daerden, F.; Ristic, B.; Thomas, I.; Vandaele, A.C.; Patel, M.R.; Lopez-Moreno, J.-J.; et al. Methane on Mars: New insights into the sensitivity of CH_4 with the NOMAD/ExoMars spectrometer through its first in-flight calibration. *Icarus* **2019**, *321*, 671–690. [CrossRef]
8. Webster, C.R.; Mahaffy, P.R.; Atreya, S.K.; Flesch, G.J.; Mischna, M.A.; Meslin, P.-Y.; Farley, K.A.; Conrad, P.G.; Christensen, L.E.; Pavlov, A.A.; et al. Mars methane detection and variability at Gale crater. *Science* **2015**, *347*, 415–417. [CrossRef]
9. Xia, J.; Feng, C.; Zhu, F.; Ye, S.; Zhang, S.; Kolomenskii, A.; Wang, Q.; Dong, J.; Wang, Z.; Jin, W.; et al. A sensitive methane sensor of a ppt detection level using a mid-infrared interband cascade laser and a long-path multipass cell. *Sens. Actuators B Chem.* **2021**, *334*, 129641. [CrossRef]
10. Hua, Y.; Zhang, L.; Zhou, Y.; Yu, L.; Zheng, K.; Song, F.; Liu, M.; Yang, Y.; Zheng, C.; Wang, Y. Near-Infrared Off-Axis integrated cavity output spectroscopic sensor system for ammonia leakage measurement in cold store using embedded Multi-Core processor. *Infrared Phys. Technol.* **2024**, *137*, 105150. [CrossRef]
11. Malara, P.; Witinski, M.F.; Capasso, F.; Anderson, J.G.; De Natale, P. Sensitivity enhancement of off-axis ICOS using wavelength modulation. *Appl. Phys. B Lasers Opt.* **2012**, *108*, 353–359. [CrossRef]
12. Yuan, Z.; Huang, Y.; Zhao, Q.; Zhang, L.; Lu, X.; Huang, J.; Qi, G.; Luo, T.; Cao, Z. Dual-path coupling V-shaped structure off-axis integrated cavity output spectroscopy (V-OA-ICOS) for water vapor stable isotope detection at 3.66 μm. *Sens. Actuators B Chem.* **2024**, *410*, 135676. [CrossRef]
13. Ngo, M.-N.; Nguyen-Ba, T.; Dewaele, D.; Cazier, F.; Zhao, W.; Nähle, L.; Chen, W. Wavelength modulation enhanced off-axis integrated cavity output spectroscopy for OH radical measurement at 2.8 μm. *Sens. Actuators A Phys.* **2023**, *362*, 114654. [CrossRef]
14. Malara, P.; Maddaloni, P.; Gagliardi, G.; De Natale, P. Combining a difference-frequency source with an off-axis high-finesse cavity for trace-gas monitoring around 3 μm. *Opt. Express* **2006**, *14*, 1304–1313. [CrossRef]
15. Herriott, D.; Kogelnik, H.; Kompfner, R. Off-axis paths in spherical mirror interferometers. *Appl. Opt.* **1964**, *3*, 523–526. [CrossRef]
16. Romanini, D. Modelling the excitation field of an optical resonator. *Appl. Phys. B Lasers Opt.* **2014**, *115*, 517–531. [CrossRef]
17. Kasyutich, V.L.; Sigrist, M.W. Optimisation of laser linewidth and cavity alignment in off-axis cavity-enhanced absorption spectroscopy. *Infrared Phys. Technol.* **2015**, *71*, 179–186. [CrossRef]

18. Shen, G.; Chao, X.; Sun, K. Modeling the optical field in off-axis integrated-cavity-output spectroscopy using the decentered Gaussian beam model. *Appl. Opt.* **2018**, *57*, 2947–2954. [CrossRef]
19. Zheng, K.; Zheng, C.; Zhang, H.; Guan, G.; Zhang, Y.; Wang, Y.; Tittel, F.K. A novel gas sensing scheme using near-infrared multi-input multi-output off-axis integrated cavity output spectroscopy (MIMO-OA-ICOS). *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *256*, 119745. [CrossRef]
20. Wang, F.; Wu, J.; Liang, R.; Wang, Q.; Wei, Y.; Cheng, Y.; Li, Q.; Cao, D.; Xue, Q. Ultra-Stable temperature controller-based laser wavelength locking for improvement in WMS methane detection. *Sensors* **2023**, *23*, 5107. [CrossRef]
21. Tian, X.; Cao, Y.; Chen, J.; Liu, K.; Wang, G.; Tan, T.; Mei, J.; Chen, W.; Gao, X. Dual-gassensor of CH_4/C_2H_6 based on wavelength modulation spectroscopy coupled to a home-made compact dense-pattern multipass cell. *Sensors* **2019**, *19*, 820. [CrossRef]
22. Lehmann, K.K.; Romanini, D. The superposition principle and cavity ring-down spectroscopy. *J. Chem. Phys.* **1996**, *105*, 10263–10277. [CrossRef]
23. Siegman, A.E. Physical Properties of Gaussian Beams. In *Laser*; University Science Books: Herndon, VA, USA, 1986; pp. 663–697.
24. Al-Rashed, A.-A.R.; Saleh, B.E.A. Decentered gaussian beams. *Appl. Opt.* **1995**, *34*, 6819–6825. [CrossRef]
25. Siegman, A.E. Complex paraxial wave optics. In *Laser*; University Science Books: Herndon, VA, USA, 1986; pp. 777–814.
26. Arndt, R. Analytical line shapes for lorentzian signals broadened by modulation. *J. Appl. Phys.* **1965**, *36*, 2522–2524. [CrossRef]
27. Zhu, G.; Zhu, H.; Yang, C.; Gui, W. Parameter optimization of a short open optical path's oxygen concentration detection system based on WMS. *Chin. J. Electron.* **2017**, *26*, 797–802. [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.

sensors

MDPI

Article

Research on the Transformer Failure Diagnosis Method Based on Fluorescence Spectroscopy Analysis and SBOA Optimized BPNN

Xueqing Chen [1], Dacheng Li [2,3] and Anjing Wang [2,3,*]

[1] Institutes of Physical Science and Information Technology, Anhui University, Hefei 230601, China; q22301351@stu.ahu.edu.cn
[2] Anhui Institute of Optics and Fine Mechanics, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China; dcli@aiofm.ac.cn
[3] Key Laboratory of General Optical Calibration and Characterization Technology, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China
* Correspondence: ajwang@aiofm.ac.cn

Abstract: The representative dissolved gases analysis (DGA) method for transformer fault detection faces many shortcomings in early fault diagnosis, which restricts the application and development of fault detection technology in the field of transformers. In order to diagnose early failure in time, fluorescence analysis technology has recently been used for the research of transformer failure diagnosis, which makes up for the shortcomings of DGA. However, most of the existing fluorescence analyses of insulating oil studies combined with intelligent algorithms are a qualitative diagnosis of fault types; the quantitative fault diagnosis of the same oil sample has not been reported. In this study, a typical fault simulation experiment of the interval discharge of insulating oil was carried out with the new Xinjiang Karamay oil, and the fluorescence spectroscopy data of insulating oil under different discharge durations were collected. In order to eliminate the influence of noise factors on the spectral analysis and boost the accuracy of the diagnosis, a variety of spectral preprocessing algorithms, such as Savitzky–Golay (SG), moving median, moving mean, gaussian, locally weighted linear regression smoothing (Lowess), locally weighted quadratic regression smoothing (Loess), and robust (RLowess) and (Rloess), are used to smooth denoise the collected spectral data. Then, the dimensionality reduction techniques of principal component analysis (PCA), kernel principal component analysis (KPCA), and multi-dimensional scale (MDS) are used for further processing. Based on various preprocessed and dimensionally reduced data, transformer failure diagnosis models based on the particle swarm optimization algorithm (PSO) and the secretary bird optimization algorithm (SBOA) optimized BPNN are established to quantitatively analyze the state of insulating oil and predict the durations of transformer failure. By using the mathematical evaluation methods to comprehensively evaluate and compare the effects of various algorithm models, it was found that the Loess-MDS-SBOA-BP model has the best performance, with its determination coefficient (R^2) increasing to 99.711%, the root mean square error (RMSE) being only 0.27144, and the other evaluation indicators also being optimal. The experimental results show that the failure diagnosis model finally proposed in this paper can perform an accurate diagnosis of the failure time; the predicted time is closest to the true value, which lays a foundation for the further development of the field of transformer failure diagnosis.

Keywords: transformer failure diagnosis; fluorescence spectroscopy; back propagation neural network (BPNN); secretary bird optimization algorithm (SBOA)

1. Introduction

The transformer is an important infrastructure for transmission and distribution in a power system, monitoring its operation status, strengthening the ability to detect faults, especially in the early stage of faults, the timely detection of faults, and early warnings, which is conducive to allowing maintenance personnel to quickly take countermeasures, minimize losses, and keep the power system operating stably [1,2]. Insulating oil is an insulating material stored in the oil tank of an oil-immersed transformer, and when the transformer is at fault, the state of the insulating oil changes, and based on this mechanism, the dissolved gas analysis method (DGA) has been widely used to study transformer failure, and the faults can be diagnosed according to the gas change in the oil [3,4].

However, in early fault warnings, DGA is constrained by the following:

1. The characteristic gas content does not change significantly, does not exceed the gas standard warning value, and cannot issue an early warning for latent hidden dangers in time;
2. The intelligence is low and the auxiliary discrimination method is not flexible;
3. The processing process is complex and the detection cycle time is long, which makes it impossible to quickly perform an early fault diagnosis [5–7].

The research on this technology needs to be gradually improved or replaced by other more efficient new technology methods, such as the latest research that has proposed a new method for the detection of dissolved gases and the use of fiber-enhanced Raman spectroscopy technology to detect a variety of decomposition gases in environmentally friendly gas-insulated power equipment, so as to perform the operation status evaluation of gas-insulated power equipment [8].

Facing the shortcomings of gas analysis methods in identifying the early faults of oil-immersed transformers, fluorescence technology has been applied to the field of transformer failure detection in some studies, due to its advantages of the large amount of information, fast detection speed, high sensitivity, and that fact that it does not need to be in direct contact with oil samples. Studies have shown that the generation of and change in fluorescence signals in transformer oil are related to the change in the concentration of aromatic hydrocarbons, which are fluorescent substances [9,10]. When some electrical and thermal failures arise within transformers, the aromatic substances in the oil crack and the spectral intensity weakens, so the fluorescence analysis method of insulating oil can realize the timely monitoring of transformer oil deterioration [11–14]. At present, most of the reported research on the failure detection of power transformers is qualitative diagnosis. Pengcheng Yan, Rongying Dai, and Feng H et al. used laser-induced fluorescence (LIF), built a LIF system to collect the insulating oil spectra of different transformer fault types, and fused a variety of intelligent algorithms to establish a classification model to classify and identify the common fault types of power transformers (electrical faults, thermal faults, local moisture faults, lightning faults, water inlet damage faults, and short circuit faults) [15–18].

Due to the different causes of faults, the positions and intensities of oil fluorescence spectral characteristic peaks of different fault types are often very different and can even be effectively identified by the eye alone, so they can often achieve a good fault classification effect [19–21]. However, there are few fault quantitative diagnostic studies on transformer oil of the same type of fault with no obvious spectral difference, which cannot accurately predict the fault time and evaluate the severity of the fault type. For the quantitative detection of faults, the fluorescence double-color ratio method proposed by Zhao Yue et al. and the ultra-high-sensitivity fault detection method proposed by Mu L et al. on this basis are relatively new methods that have been reported so far. However, they all rely on fluorescence bi-characteristic peak information, which requires the spectra to exhibit the

characteristics of doublets, and this approach may not be suitable for complex fluorescence spectra with single or multiple peaks [22,23].

In order to solve the problem of the quantitative diagnosis of transformer early fault duration time, this paper proposes a universally applicable and accurate quantitative diagnosis method of transformer faults by combining fluorescence spectral characteristic analysis and the artificial intelligence algorithm.

2. Materials and Methods

The technical route is shown in Figure 1. Firstly, the interval discharge experiment was simulated. After the early fault oil samples were obtained, the fluorescence spectral data under different fault durations were collected, and several preprocessing algorithms were used for smooth denoising: Savitzky–Golay (SG), moving median, moving mean, gaussian filtering, locally weighted linear regression smoothing (Lowess), locally weighted quadratic regression smoothing (Loess), and robust (RLowess) and (Rloess) signal smoothing algorithms. Then, redundant information was removed and the data dimension was reduced by principal component analysis (PCA), kernel principal component analysis (KPCA) [24], and multi-dimensional scale (MDS) dimensionality reduction algorithms. Based on the preprocessed data, an error back propagation neural network (BPNN) model was established, then the network's weight and threshold parameters were optimized by using the secretary bird intelligence optimization algorithm (SBOA). Compare the diagnostic results of different models to verify the effectiveness of the transformer failure diagnosis model based on the SBOA optimized BPNN proposed in this study and perform the quantitative diagnosis of the early failure time of transformers.

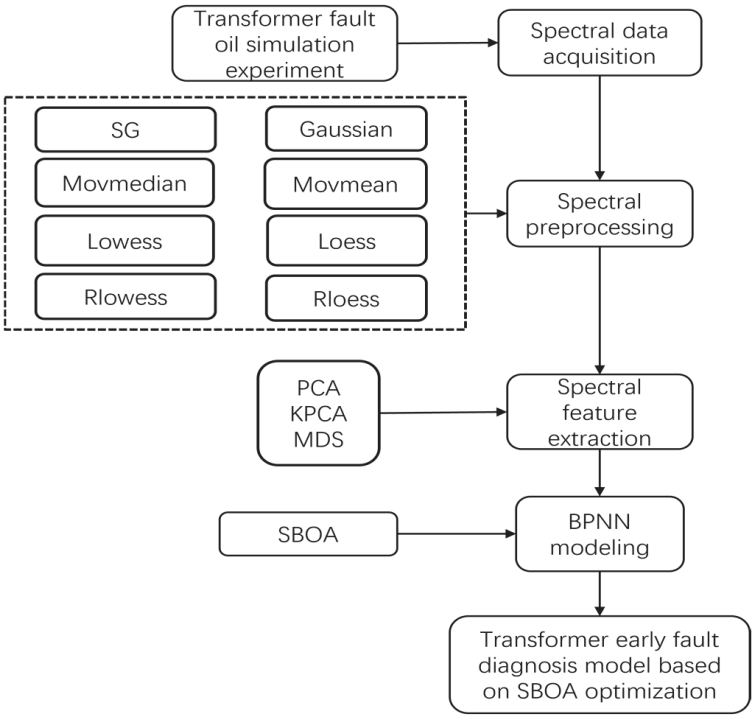


Figure 1. Technical flow chart.

2.1. Materials

2.1.1. Experimental Samples

The new Karamay oil was filtered and discharged to simulate the typical fault of the interval discharge, and the oil samples of early failure were obtained and stored in a sealed container. A total of 6 oil samples with different fault degrees were obtained, and we

collected 10 pieces of spectral data for each faulty oil sample. All faulty oil samples used were prepared by experiments at the Institute of Electrical Engineering, Chinese Academy of Sciences. The storage and spectral acquisition were carried out in a dark room to prevent the influence of light. The faulty oil information is shown in Table 1.

Table 1. The information of the faulty transformer insulating oil samples.

Degree	Acquisition Time	Numbers
0 min	1 March 2024	10
2 min	1 March 2024	10
4 min	1 March 2024	10
6 min	1 March 2024	10
10 min	1 March 2024	10
15 min	1 March 2024	10

2.1.2. Spectral Acquisition

The fluorescence spectroscopy of each faulty oil sample was collected, and the detection instrument was the OmniFluo 900 fluorescence spectrometer produced by Zolix Instrument Co., Ltd (Beijing, China). A 75 W continuous xenon lamp was used as the excitation light source. The collection parameters are shown in Table 2.

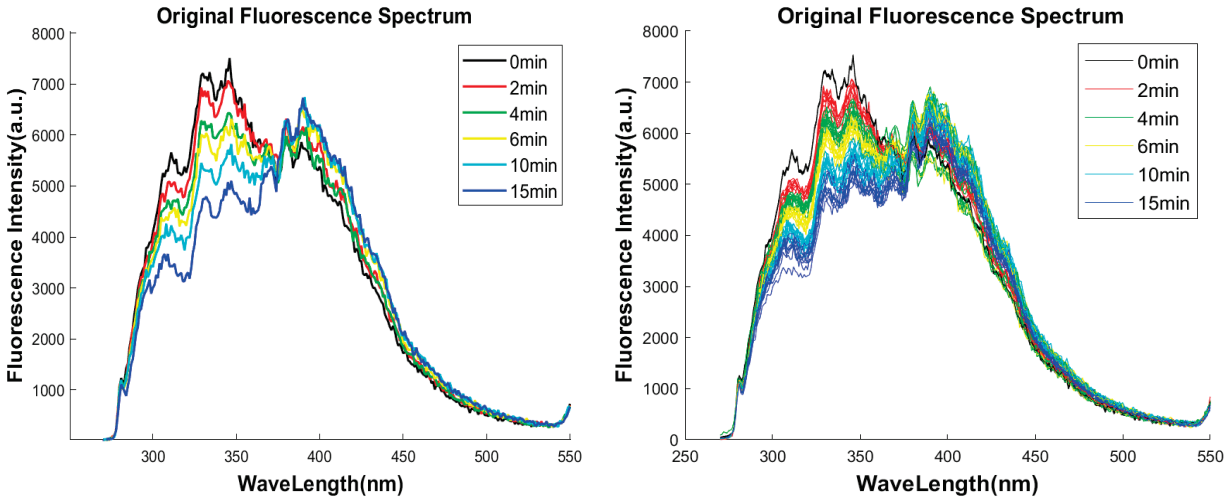
Table 2. Acquisition parameters.

Acquisition Parameters	Value
Excitation wavelength	280 nm
Emission wavelength	270 nm–600 nm
Excitation slit width	1.5 nm
Emission slit width	1.5 nm
Integration time	100 ms
Excitation scan step size	5.0 nm
Emission scan step size	1.0 nm

2.1.3. Original Spectral Analysis

As shown in Figure 2a, the original fluorescence spectra at different degrees of failure were acquired at a fixed excitation wavelength of 280 nm, and a random piece of spectral data was selected at each failure time to plot. The fluorescence characteristic bands of the emission spectrum are mainly in the range of 270 nm–550 nm; the difference in fluorescence intensity is mainly in the range of 300 nm–450 nm. With the increase in the discharge time, the intensity of fluorescence gradually decreased [25]. The typical characteristic peaks of the new oil spectrum are as follows: characteristic peak-1: EM is 311 nm, fluorescence intensity is 5658 a.u., characteristic peak-2: EM is 329 nm, fluorescence intensity is 7167 a.u., characteristic peak-3: EM is 346 nm, fluorescence intensity is 7509 a.u., characteristic peak-4: EM is 390 nm, and fluorescence intensity is 5861 a.u. After the wavelength of 420 nm, the spectrum gradually converges, and no other characteristic peaks appear. The intensity values of characteristic peaks 1, 2, and 3 gradually reduced with the increase in the discharge time, and the variation pattern of the fluorescence intensity of characteristic peak 4 is not obvious. The reason for this phenomenon is that the insulating oil of the power transformer is mainly composed of hydrocarbons, such as alkanes and cycloalkanes, and in the case of arc and local overheating, there is a phenomenon of carbon–hydrogen bond breakage or carbon–carbon bond breakage, and the broken carbon atoms and hydrogen atoms are recombined through complex chemical reactions to generate other chemical substances that do not have fluorescent properties, and the aromatic hydrocarbon concentration decreases,

resulting in the fluorescence intensity of the insulating oil decreasing with the deepening of the degree of failure.



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

Figure 2. Original spectral analysis: (**a**) fluorescence spectra at different fault times; (**b**) all original fluorescence spectral data.

All original spectral data collected by the spectrometer are shown in Figure 2b. The pyrolysis degree of aromatic hydrocarbons of different faulty transformer oils with shorter fault interval times is relatively similar, the peak position and fluorescence intensity of the fluorescence spectral curves overlap, and the degree of overlap between some curves is relatively high. The spectral signal and the aromatic hydrocarbon concentration changes may have a complex nonlinear relationship. In order to correlate the spectral information with the degree of failure and directly detect the degree of failure of unknown oil samples with the information acquired from the known data, the BPNN with a good nonlinear modeling ability was used for the regression analysis.

2.2. Methods

2.2.1. Spectral Preprocessing

Because spectral data acquisition may be performed in a relatively complex environment, the spectrum will inevitably experience interference from factors, such as the noise of high-power instruments, the noise of laser-device heat dissipation, and the noise of the external environment, resulting in the spectrum containing useless interference information, which is superimposed on the spectral image as unsmooth curves with obvious burrs, which affects the data quality and the recognition effect of the subsequent model. Therefore, it is essential to preprocess the original signal to enhance the spectral quality and filter out influences of the characteristic factors of the non-faulty oil itself. In this study, Savitzky–Golay (SG) smoothing, moving median filtering (Movmedian), moving mean filtering (Movmean), Gaussian filtering, locally weighted linear regression smoothing (Lowess), locally weighted quadratic regression smoothing (Loess), and robust (RLowess) and (Rloess) signal smoothing algorithms were used to effectively filter out high-frequency noise and improve spectral smoothness and the signal-to-noise ratio. The SG filtering algorithm uses the least squares polynomial to fit the signal within the filter window, and then uses the fitting function to calculate the smoothed data points, resulting in smooth denoising. The Movmedian, Movmean, and Gaussian filter calculate the median, moving average, and Gaussian-weighted average, respectively, for the data within the window.

Lowess and Loess are algorithms that combine weighted linear least squares with first-order and second-order polynomials, respectively, to fit the polynomial regression curves in the intervals. Rlowess and Rloess are more robust forms and filter better for outliers that deviate farther.

2.2.2. Dimensionality Reduction

The full-band fluorescence spectral data of the sample have many features and contain a large amount of redundant information, which will greatly increase the computational complexity and parameter tuning difficulty, which is not conducive to the rapid and accurate output prediction results [26]. Moreover, there are obvious differences in the importance of different wavelengths for spectral analysis, and non-critical spectral band data affect the model and cannot achieve a good prediction effect. Considering the actual requirements for the speed and accuracy of transformer fault diagnosis, this study uses several dimensionality reduction algorithms for various preprocessed data: principal component analysis (PCA), kernel principal component analysis (KPCA) based on the kernel function, and the multi-dimensional scale (MDS) analysis dimensionality reduction method to simplify the data while retaining the important features of the data, which is helpful to enhance the training speed and performance of the model.

2.2.3. BP Neural Network

The BPNN is a type of multi-layer feedforward neural network trained by backpropagating errors, and has strong nonlinear mapping capabilities [27]. In the training stage, the input samples are transformed nonlinearly to generate an output, and the output error is calculated. Then, the error is backpropagated to continuously adjust the parameters of the network, and the error is gradually reduced through repeated training until the network achieves the desired output results. The structure is divided into an input layer, hidden layer, and output layer, and the neurons are connected by weight and thresholds. The specific number of layers and neurons in each layer can be set arbitrarily. Hidden layer neurons can map the main attributes of the input sample into the output, and the number of neuronal nodes has a crucial impact on the final performance of the model.

In order to select the optimal number of hidden layer neurons and build the optimal network structure, this study used the dimensionality reduction data as the input and the cumulative failure time as the output, and built a network with different numbers of hidden layer neurons to compare the training effect. According to Equation (1), m and n are the number of neurons in the input layer and output layer, respectively, and t is taken as the tuning constant between the integers in the range of 1–10.

$$\text{Hidden Num} = \text{sqrt}(\text{m} + \text{n}) + \text{t}, \tag{1}$$

The optimal number of hidden layer nodes is determined to be in the range of 3–13, and the MSE of the training set corresponding to the number of neurons under each independent experiment is recorded.

As shown in Figure 3a, comparing the magnitude of the MSE of the training set corresponding to the different number of nodes in the hidden layers, the optimal number of nodes is 4, and the training effect of the model is the best. The MSE is only 0.0025891, so the BP neural network model with the corresponding structure is constructed as shown in Figure 3b.

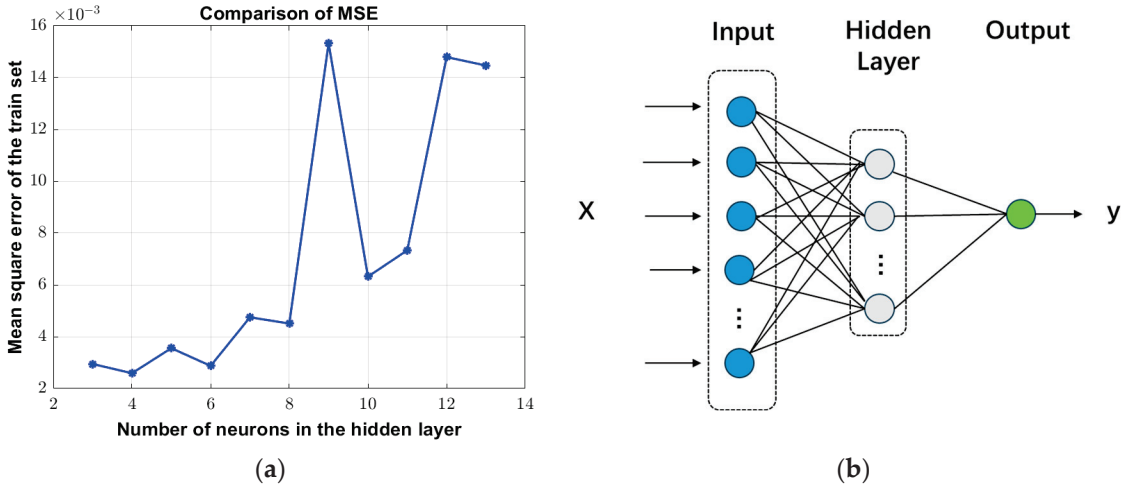


Figure 3. (**a**) MSE comparison under different numbers of hidden layer neurons; (**b**) BP neural network structure.

2.2.4. PSO Optimization

In the training process, the BPNN gradually looks for the optimal parameters, such as the connection weight and threshold, through error backpropagation, but it is likely to fall into the local optimum, which will affect the final effect of the network. Moreover, the initialized network parameters are random and the initial value has a great impact on the network performance. Therefore, this study uses the intelligent optimization algorithm to optimize the weight and threshold parameters of the network model, so as to obtain a better model failure prediction effect.

Particle swarm optimization (PSO) is a classical global optimization algorithm based on swarm intelligence, which is often used for various parameter optimization work. The algorithm simulates the predatory behavior of a population of birds and finds the optimal solution for the search space through the information sharing and cooperation between particles in the population [28].

The motion of the particles is guided by the individual optimal solution (pBest) and the global optimal solution (gBest), and their position and velocity are updated according to Equations (2) and (3):

$$v_i^{t+1} = wv_i^t + c_1 r_1 (pBest_i - x_i^t) + c_2 r_2 (gBest - x_i^t) \quad (2)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (3)$$

where v_i^t and x_i^t represent the velocity and position of particle i in generation t, w is the inertia weight, c_1 and c_2 are learning factors, $pBest_i$ represents the optimal position of particle i in the historical iteration, and gBest is the best position of the whole population.

The flow of the algorithm is as follows:

Step 1: Initialize the position and velocity of the particles randomly, and set the inertia weight, acceleration constant, maximum number of iterations, and other parameters of the algorithm.

Step 2: Calculate the fitness value for each particle based on the objective function.

Step 3: Update the individual optimal solution $pBest_i$ and the global optimal solution gBest using the velocity and position update formulas.

Step 4: Start the cycle from step 2 until the maximum number of iterations or the error meets the requirements, and the result is output.

The flow chart is shown in Figure 4:

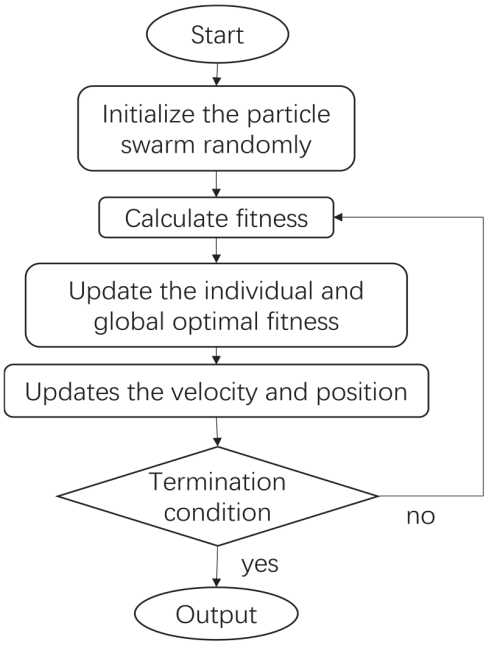


Figure 4. Flow chart for PSO algorithm.

2.2.5. SBOA Optimization

The secretary bird optimization algorithm (SBOA) is a novel swarm bionic intelligent optimization algorithm, which is separated into two stages, the exploration stage and the development stage, which is conceived from the survival and hunting method of the secretary bird to hunt prey and escape from pursuit, with a fast convergence speed and good stability [29]. The flow of the SBOA algorithm is shown in Equations (4)–(11).

Step 1: Initialization stage:

Initialize the population, N; the dimension of the solution space, Dim; the maximum number of iterations, T; initialize the position, $X_{i,j}$ of the secretary bird randomly in the search space; and calculate the fitness function, Fi.

$$X_{i,j} = lb_j + r * (ub_j - lb_j), \tag{4}$$

where ub_j and lb_j are the upper and lower bounds.

Step 2: Exploration stage P1:

1. Hunting for prey:

The differential evolution strategy is used to create new solutions, which boosts the diversity and global search ability of the algorithm and avoids falling into a local optimum.

$$\text{While } t < \frac{1}{3}T,\ x_{i,j}^{\text{new P1}} = x_{i,j} + (x_{\text{random_1}} - x_{\text{random_2}}) * R_1, \tag{5}$$

2. Exhaust Prey:

Brownian Motion (RB) is adopted to simulate random behaviors of secretary birds circling around prey that exhausts their stamina. x_{best} is the present optimal value.

$$RB = \text{randn}(1, \text{Dim}), \tag{6}$$

$$\begin{gathered} \text{while } \tfrac{1}{3}T < t < \tfrac{2}{3}T, \\ x_{i,j}^{\text{new P1}} = x_{best} + \exp\left(\left(\tfrac{t}{T}\right)^4\right) * (RB - 0.5) * (x_{best} - x_{i,j}) \end{gathered} \tag{7}$$

3. Attack prey:

In order to enhance the global search ability, the weighted Levy flight strategy, RL, is introduced. The update method is as follows, where $(1 - t)\hat{}(2 * t)$ is the nonlinear disturbance factor, which is used to achieve a balance between the exploration and development stages.

$$\text{While } t > \frac{2}{3}T,\ x_{i,j}^{\text{new P1}} = x_{\text{best}} + \left(\left(1 - \frac{t}{T}\right)\hat{}\left(2 * \frac{t}{T}\right)\right) * x_{i,j} * RL, \tag{8}$$

In this stage, the position of X_i in the search space is updated as follows:

$$X_i = \begin{cases} X_i^{new,P1}, & \text{if } F_i^{new,P1} < F_i \\ X_i, & \text{else} \end{cases}, \tag{9}$$

Step 3: Development stage P2: there are two strategies for escaping from pursuit.

C1: hide in the environment in disguise;
C2: flee or fly away quickly.

$$X_{i,j}^{new,P2} = \begin{cases} C1 : x_{\text{best}} + (2 * RB - 1) * \left(1 - \frac{t}{T}\right)^2 * x_{i,j}, & \text{if rand} < r_i \\ C2 : x_{i,j} + R_2 * \left(x_{\text{random}} - K * x_{i,j}\right), & \text{else} \end{cases}, \tag{10}$$

$$X_i = \begin{cases} X_i^{new,P2}, & \text{if } F_i^{new,P2} < F_i \\ X_i, & \text{else} \end{cases}, \tag{11}$$

Step 4: until the iterative termination condition is satisfied, the best solution x_{best} is output.

In this study, the SBOA was used to treat the connection weights and thresholds between the layers of the neural network as the subjects in the population, and the global best solution found after iteration stop was assigned to the BPNN. Thus, the optimized SBOA-BP failure diagnosis model was obtained. The flow chart is shown in Figure 5.

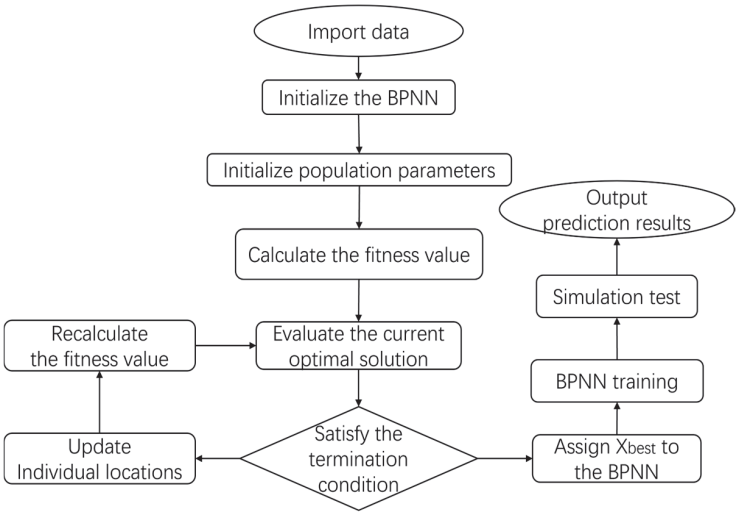


Figure 5. Flow chart for SBOA-BP failure diagnosis model.

3. Results and Discussion

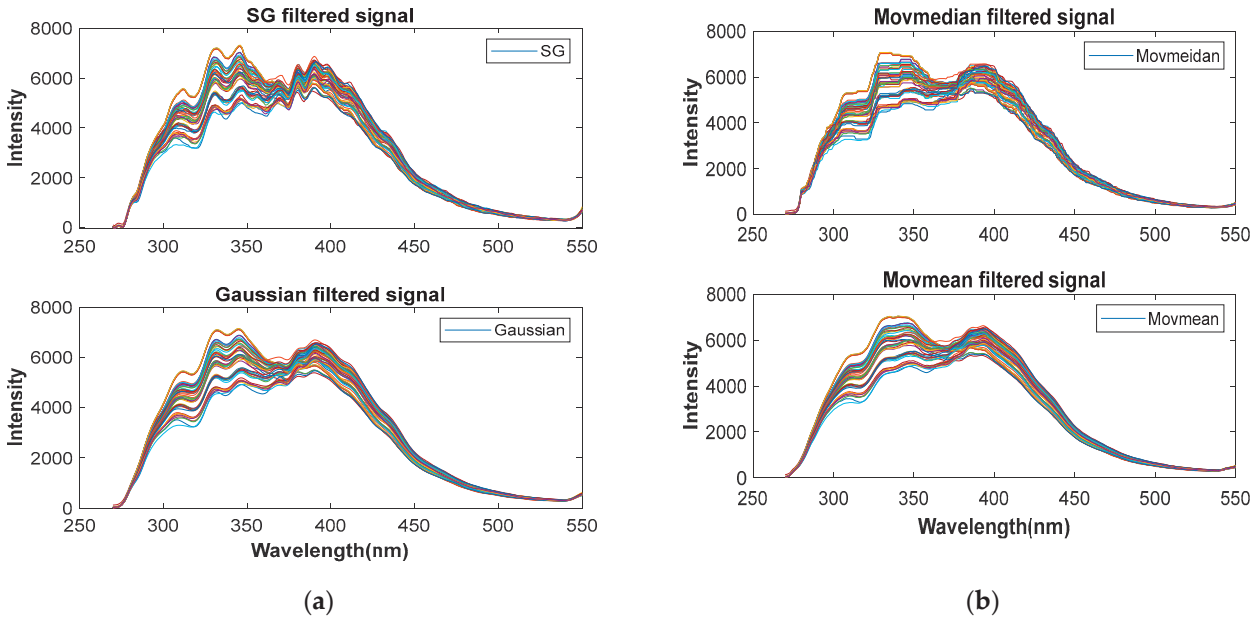
3.1. Spectral Preprocessing

Comparing several preprocessed images in Figure 6, it can be seen that, after SG, Gaussian, Lowess, Loess, and Rloess filtering, the noise of the middle and high frequencies in the signal is effectively suppressed, and the curve becomes smoother. The Movmedian filter calculates the median in each window, the signal is still not smooth enough after

preprocessing, and there are still many interferences in the curve. And Movmean and Rlowess have higher smoothness after pretreatment. However, it is still difficult to select the best filtering method from the image alone. Each oil sample is randomly divided into a training set and a test set according to the proportion, a neural network model is established, and a suitable smooth denoising algorithm is selected according to the regression prediction ability of the model. In order to improve the credibility and avoid the overfitting of the model and affect the generalization ability, an independent test set is used to evaluate the actual performance of the model. In Table 3, the prediction results of the test set after the original spectrum and various preprocessed spectral modeling tests are recorded, among which the prediction value of the unprocessed Original-BPNN model deviates the most from the real value, and the prediction effect is the worst: the RMSE of test set is as high as 2.7423 and the model's R^2 is only 70.477%. However, compared with the Original-BP, the prediction effect of the model after various smooth denoising preprocessing tests improves. The best of these is Rlowess-BP, whose R^2 increases to 96.294% and RMSE decreases to 0.97161; the worst is the Movmedian model, whose R^2 also increases to 87.987% and the RMSE to 1.7493. However, due to the large spectral dimension, the overall prediction error of the model is still large. Considering the practical application, the requirements for the accuracy and precision of the transformer failure time analysis require spectral dimensionality reduction to improve the accuracy of the results.

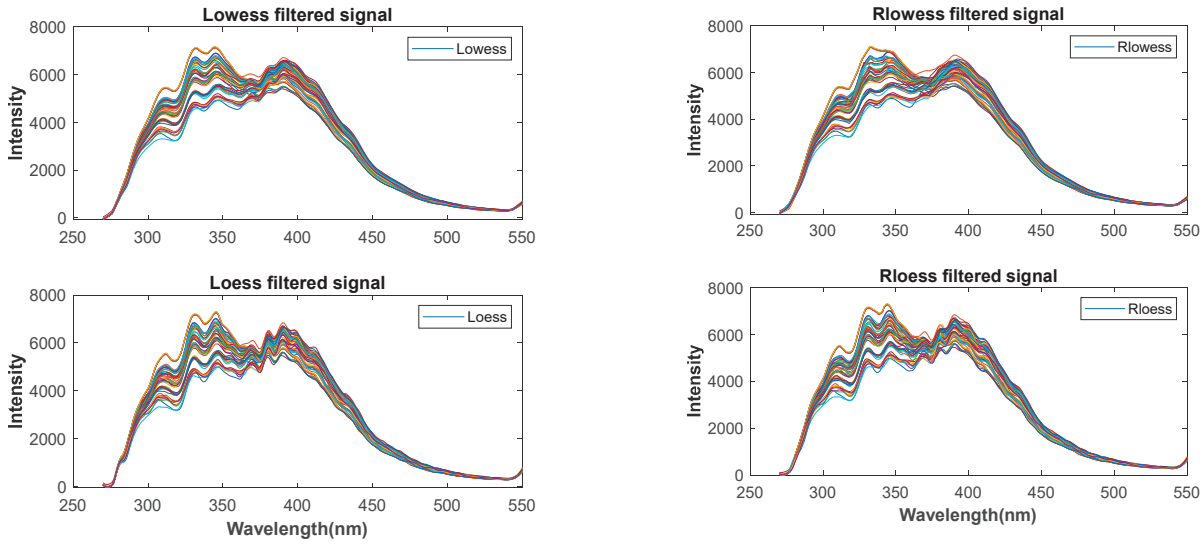
Table 3. Evaluation indicators of models after spectral preprocessing.

Evaluation Indicators	Original	SG	Median	Mean	Gaussian	Lowess	Loess	Rlowess	Rloess
MAE	1.8317	1.1962	1.3435	0.72386	0.80052	0.86717	0.97859	0.73532	1.1972
MSE	7.5201	2.1821	3.0601	1.2868	1.3843	1.5284	1.5227	0.94403	2.2853
RMSE	2.7423	1.4772	1.7493	1.1344	1.1766	1.2363	1.234	0.97161	1.5117
R^2	0.70477	0.91433	0.87987	0.94948	0.94566	0.94	0.94022	0.96294	0.91028



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

Figure 6. *Cont.*



(**c**) (**d**)

Figure 6. Preprocessed images. (**a**) SG and Gaussian filtered signals; (**b**) moving median and moving mean filtered signals; (**c**) Lowess and Loess filtered signals; (**d**) Rlowess and Rloess filtered signals.

3.2. Dimensionality Reduction

Figure 7 shows the number of principal components and the corresponding cumulative contributions of the data under various preprocessing steps. When there is one principal component, the cumulative contribution of all models does not reach 85%, and then the cumulative contribution rate gradually increases with the increase in the number of selected principal components. Taking KPCA dimensionality reduction as an example, when there is one principal component, the cumulative contribution rate of the Mean-KPCA model is the highest, reaching 73.3159%, and the other models are lower than this. The cumulative contribution rates of all preprocessed models are higher than those of the original data; the cumulative contribution rate of the Original-KPCA is the lowest, only 65.6076%. When the number of principal components increases to 5, the cumulative contribution gradually stabilizes, and the cumulative contribution of all data preprocessed models are higher than 95%. At this time, the cumulative contribution of Mean-KPCA is still the highest, which is 98.4293%; the lowest is Rloess-KPCA, which is 95.1081%; and the cumulative contribution rate of untreated Original-KPCA is only 88.9571%. So, the number of principal components is selected as 5 for subsequent modeling and identification.

Comparing the cumulative contribution of different models, under the same dimensionality reduction method, the cumulative contribution of all preprocessed models improved compared with the original data without pretreatment. And in several dimension reduction methods, all the cumulative contributions after Movmean pretreated reached the highest value, and all the cumulative contributions after Rloess pretreatment were the lowest. In general, Movmean-MDS has the highest cumulative contribution rate among all models, with the first five principal components reaching 99.8327%, and the lowest is Original-KPCA, only 88.9571%. Figure 8 shows the cumulative contribution rate of Movmean-MDS and the data distribution of the first two principal components after dimensionality reduction.

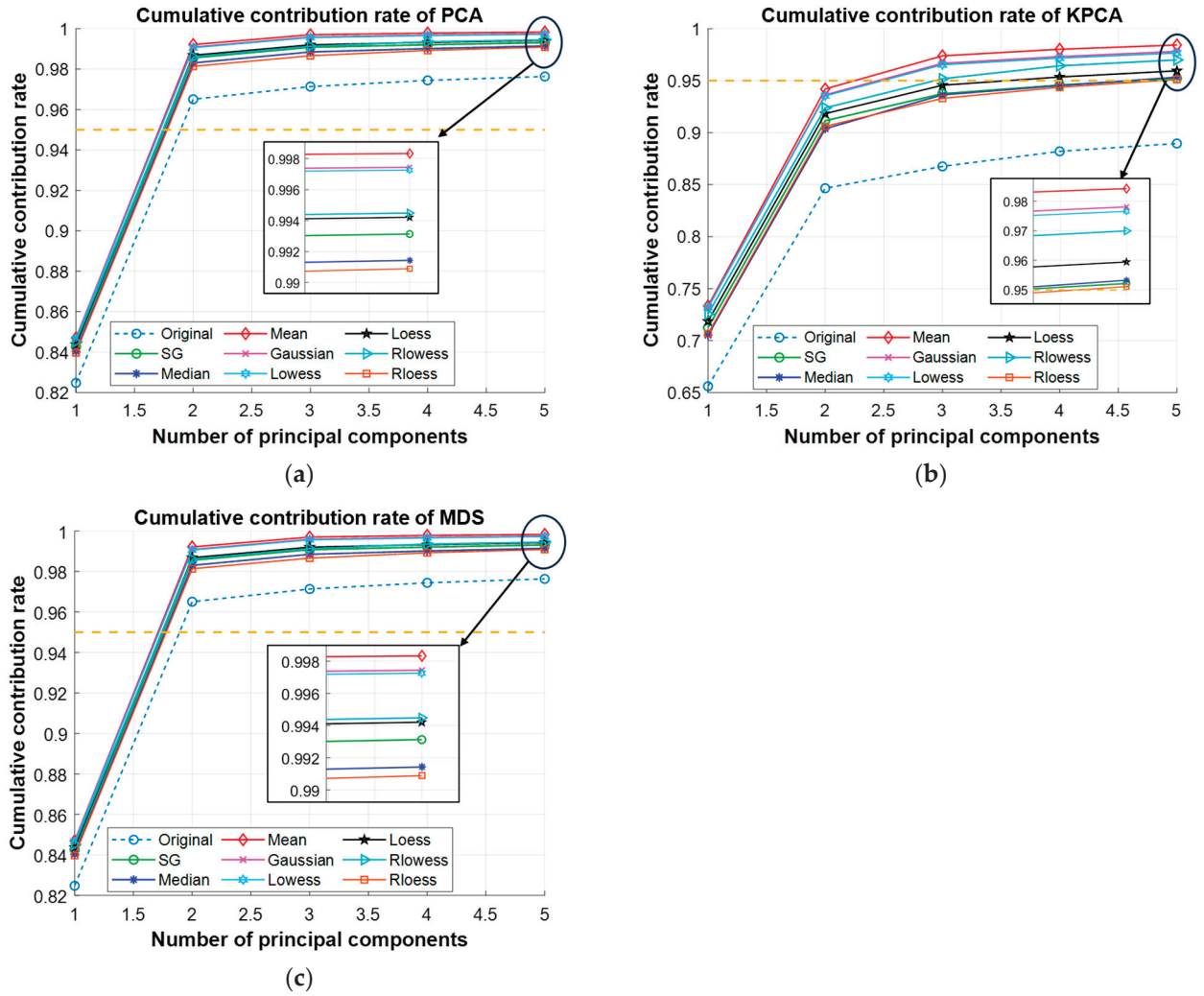


Figure 7. Cumulative contribution of principal components under PCA, KPCA, and MDS: (**a**) PCA; (**b**) KPCA; (**c**) MDS.

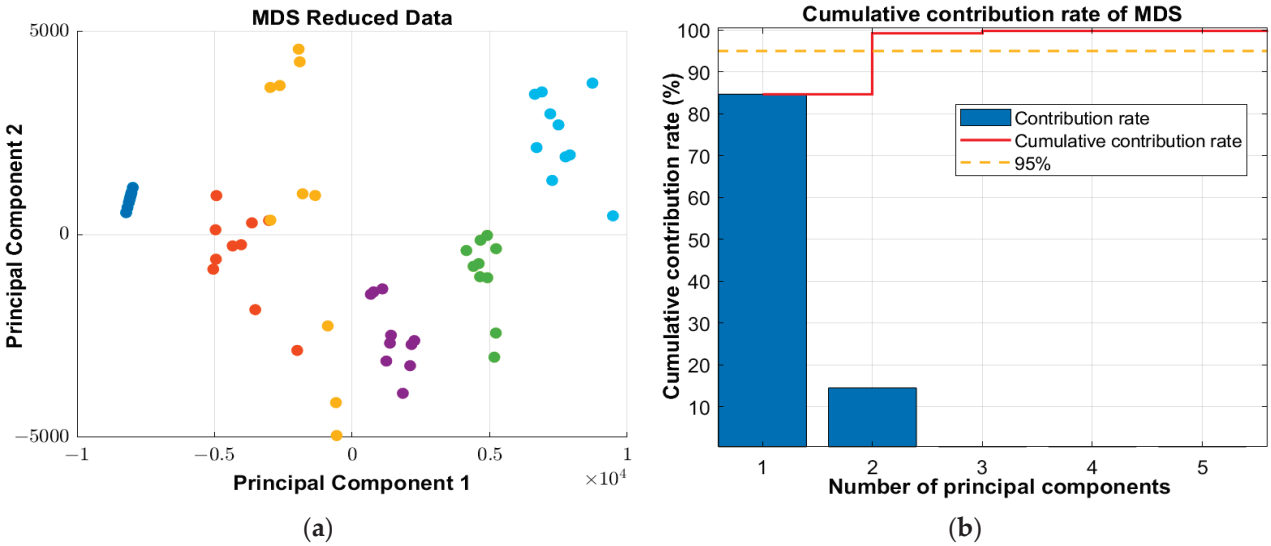


Figure 8. (**a**) First and second component scores after MDS dimensionality reduction; (**b**) cumulative contribution rate after MDS dimensionality reduction.

3.3. BP Neural Network

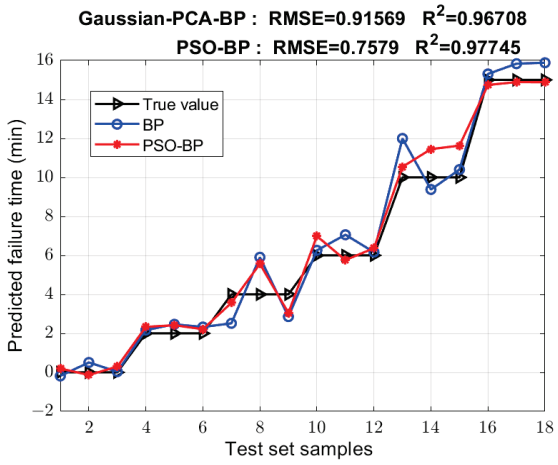
Then, the BP neural network model is established to fit the transformer failure time, and the data visualization and model evaluation parameters are compared, so as to find the most suitable smooth preprocessing and feature extraction methods. The prediction results of each model test set are recorded in Table 4, and the one with the best effect is SG-PCA-BP with an R2 of 99.411% and RMSE of 0.38728 under the same PCA dimensionality reduction method. The worst model is the Rlowess-PCA-BP with an R^2 value of 96.234% and RMSE of 0.97948. The Gaussian-KPCA-BP pretreatment model has the best effect under the same KPCA method, with an R^2 of 97.799% and RMSE of 0.74868. The worst effect model is Lowess-KPCA-BP with an R^2 of 88.559% and RMSE of 1.7071. The best model under MDS reduction is Rlowess-MDS-BP with an R^2 of 98.633% and RMSE of 0.59011, while the worst is the SG-MDS-BP model with an R^2 of 95.575% and RMSE of 1.0616. The model's prediction value fluctuates up and down the real value, the overall error fluctuation is large, and the regression effect of the model needs to be enhanced.

Table 4. Evaluation indicators of models after dimensionality reduction.

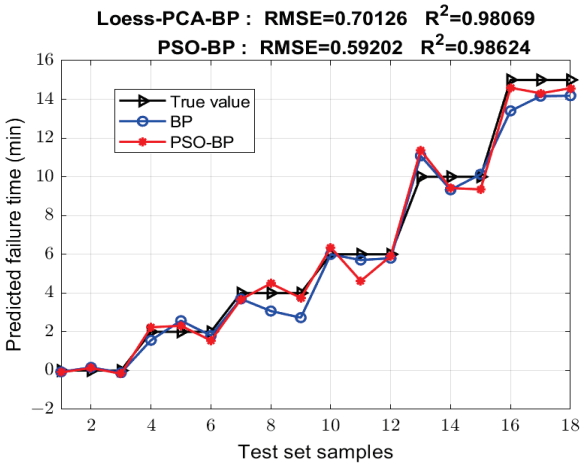
Evaluation Indicators		Original	SG	Mean	Gaussian	Lowess	Loess	Rlowess
PCA	MAE	0.35217	0.31527	0.52965	0.70788	0.41558	0.53789	0.61502
	MSE	0.23524	0.14999	0.35175	0.83849	0.27578	0.49177	0.95938
	RMSE	0.48502	0.38728	0.59309	0.91569	0.52514	0.70126	0.97948
	R^2	0.99076	0.99411	0.98619	0.96708	0.98917	0.98069	0.96234
KPCA	MAE	0.61273	0.86364	0.54661	0.62488	1.1149	0.68433	0.67522
	MSE	1.2029	1.2103	0.74745	0.56053	2.9142	0.64992	0.72844
	RMSE	1.0967	1.1001	0.86455	0.74868	1.7071	0.80618	0.85349
	R^2	0.95278	0.95249	0.97066	0.97799	0.88559	0.97449	0.9714
MDS	MAE	0.60583	0.8366	0.43934	0.65861	0.5407	0.69559	0.43913
	MSE	0.57823	1.1271	0.35401	0.8975	0.54504	0.86949	0.34823
	RMSE	0.76042	1.0616	0.59499	0.94736	0.73827	0.93246	0.59011
	R^2	0.9773	0.95575	0.9861	0.96477	0.9786	0.96587	0.98633

3.4. PSO Optimization

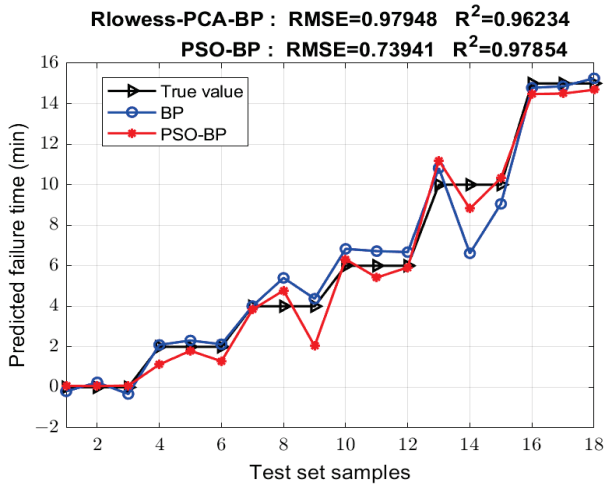
The transformer fault diagnosis model based on the PSO optimized BP neural network is established to quantitatively analyze the state of insulating oil. The prediction results of the following seven neural network models established under different preprocessing dimensionality reductions are improved compared to those before optimization, while the results of other models are not improved. The comparison of the prediction results before and after optimization is shown in Figure 9, and the performance metrics of each optimized model are shown in Table 5. Among them, the Mean-MDS-PSO-BP model has the best performance: R^2 is 99.202% and RMSE is 0.45091. As a mainstream intelligent optimization algorithm, PSO has been widely used for various optimization work for many years, and its optimization of the model needs to be replaced by other optimization algorithms with better results.



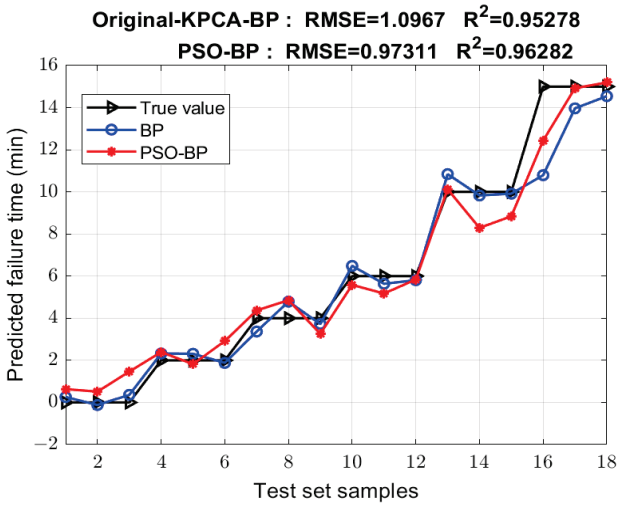
(**a**)



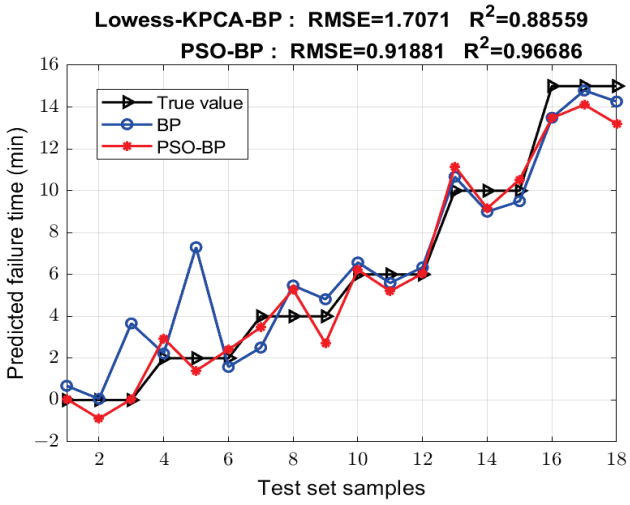
(**b**)



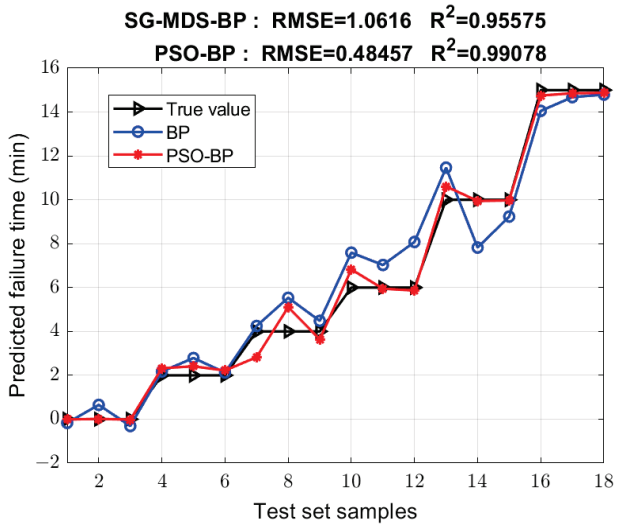
(**c**)



(**d**)

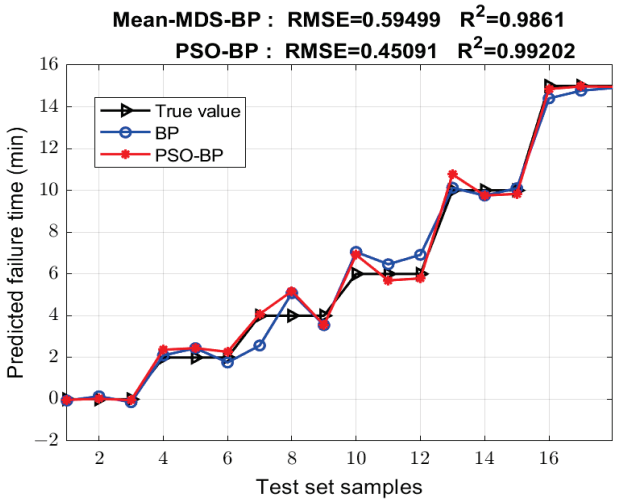


(**e**)



(**f**)

Figure 9. *Cont.*



(**g**)

Figure 9. Comparison of models before and after PSO optimization under different preprocessing methods. (**a**) Gaussian-PCA; (**b**) Loess-PCA; (**c**) Rlowess-PCA; (**d**) Original-KPCA; (**e**) Lowess-KPCA; (**f**) SG-MDS; (**g**) Mean-MDS.

Table 5. Evaluation indicators of models after PSO optimized BPNN.

	Gaussian-PCA	Loess-PCA	Rlowess-PCA	Original-KPCA	Lowess-KPCA	SG-MDS	Mean-MDS
MAE	0.56972	0.46691	0.5522	0.73889	0.76997	0.32626	0.31647
MSE	0.57442	0.35049	0.54673	0.94695	0.84421	0.23481	0.20332
RMSE	0.7579	0.59202	0.73941	0.97311	0.91881	0.48457	0.45091
R^2	0.97745	0.98624	0.97854	0.96282	0.96686	0.99078	0.99202

3.5. SBOA Optimization

The transformer failure detection model based on the SBOA optimized BPNN is established to quantitatively analyze the state of insulating oil. The prediction results of the following neural network models established under different preprocessing dimension reduction methods are improved greatly to varying degrees after SBOA optimization compared with those before optimization. The comparison of the prediction results before and after optimization is shown in Figure 10, and the performance indicators of each model are shown in the Table 6. Among them, the Loess-MDS-SBOA-BP model has the best performance: R^2 is increased to 99.711%, the RMSE is reduced to 0.27144, and the other evaluation indexes are also the best, and the predicted failure time is the closest to the true value. Compared with the best Mean-MDS-PSO-BP model after POS optimization, this model has a higher determination coefficient and a better regression fitting effect, compared to the untreated Original-BP, whose R^2 is only 70.477% and RMSE is 2.7423, which maximizes the prediction ability. The RMSE is reduced to 1/10, as shown in Figure 11.

The prediction error curves of all models are shown in Figure 12. In general, the errors of samples 2–6 and 14–16 fluctuate greatly, while in the remainder of the samples, the errors deviate from the true value slightly. Compared with the BP fault diagnosis model, the error fluctuation between the predicted value and the real value of the optimized SBOA-BP failure detection model is small, and the regression prediction effect of the model is greatly improved. Among them, the Loess-MDS-SBOA-BP model still has the smallest fluctuation in the error curve; the error curves of the training set and the test set before and after optimization are shown in Figure 13. The experimental results show that the Loess-MDS-

SBOA-BP model is the best model suitable for the quantitative diagnosis of early faults in the transformer, which can provide an accurate diagnosis of the fault time and provide a basis for the further development of the domain of transformer failure diagnosis.

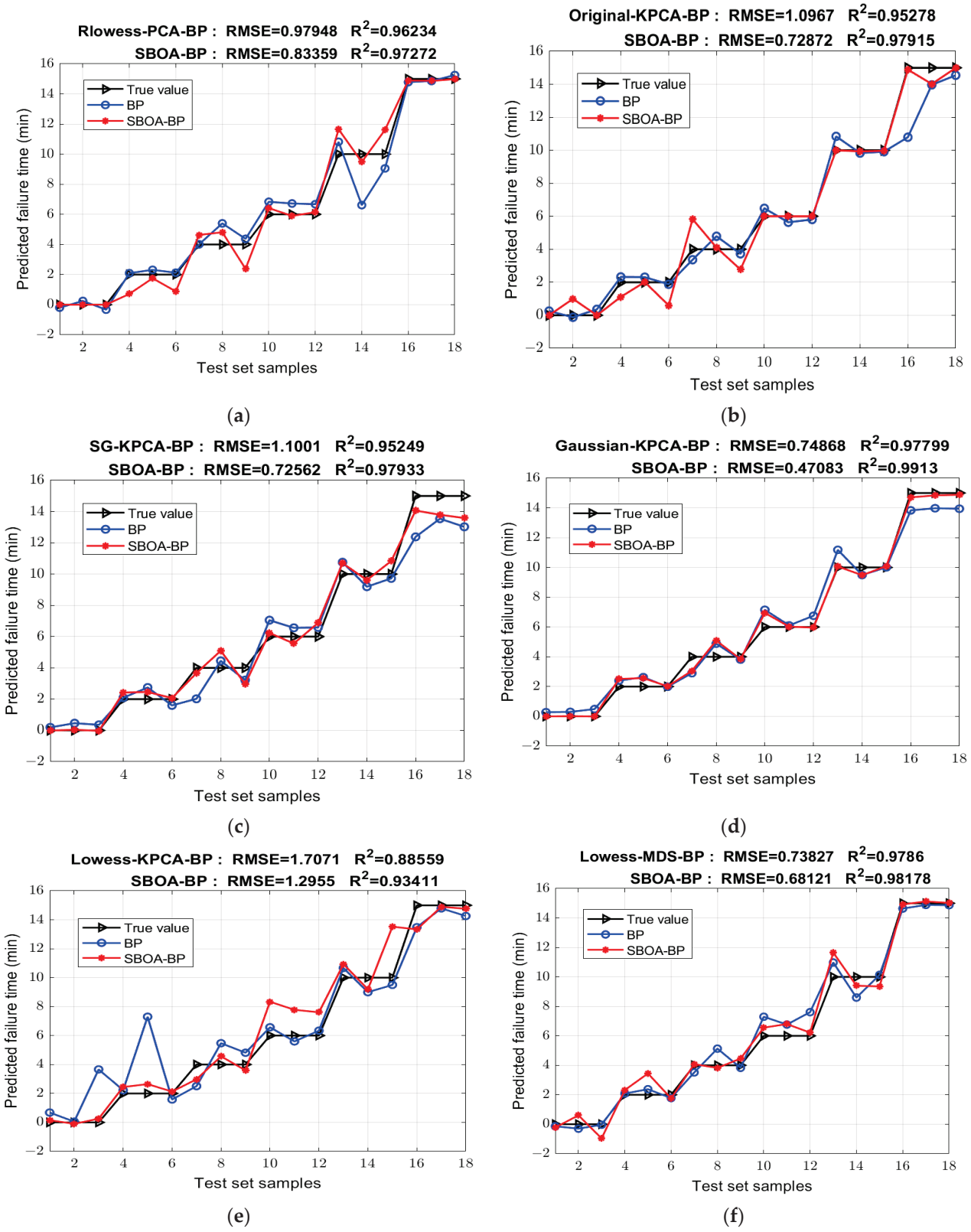


Figure 10. *Cont.*

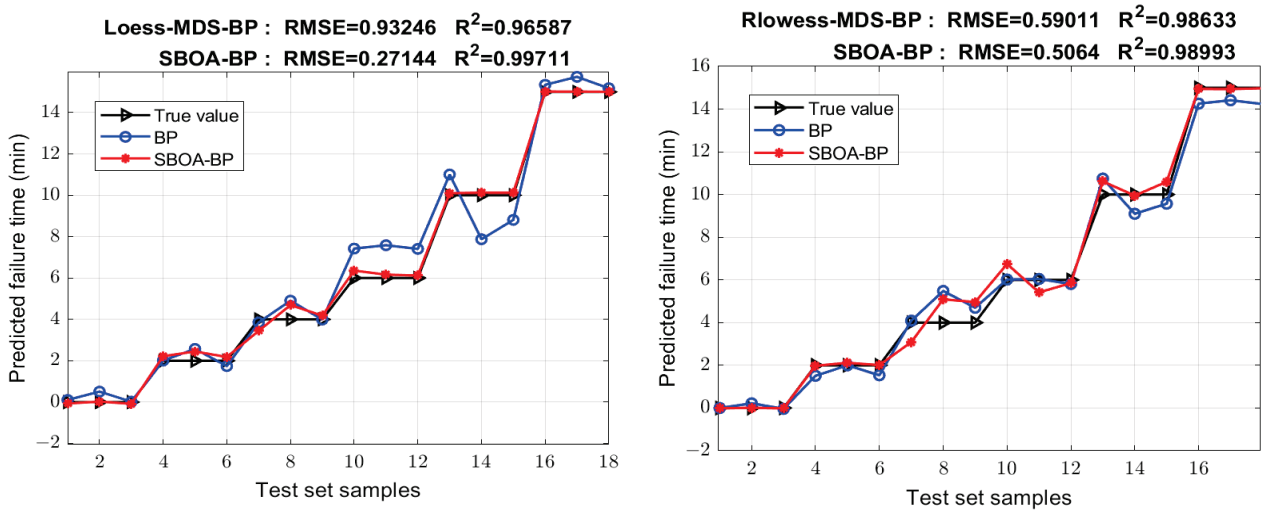


Figure 10. Comparison of models before and after SBOA optimization under different preprocessing methods. (**a**) Rlowess-PCA; (**b**) Original-KPCA; (**c**) SG-KPCA; (**d**) Gaussian-KPCA; (**e**) Low-ess-KPCA; (**f**) Lowess-MDS; (**g**) Loess-MDS; (**h**) Rlowess-MDS.

Table 6. Evaluation indicators of models after SBOA optimized BPNN.

	Rlowess-PCA	Original-KPCA	SG-KPCA	Gaussian-KPCA	Lowess-KPCA	Lowess-MDS	Loess-MDS	Rlowess-MDS
MAE	0.57971	0.42633	0.58447	0.30992	0.92728	0.51085	0.19345	0.33531
MSE	0.69486	0.53103	0.52652	0.22168	1.6783	0.46405	0.073678	0.25644
RMSE	0.83359	0.72872	0.72562	0.47083	1.2955	0.68121	0.27144	0.5064
R^2	0.97272	0.97915	0.97933	0.9913	0.93411	0.98178	0.99711	0.98993

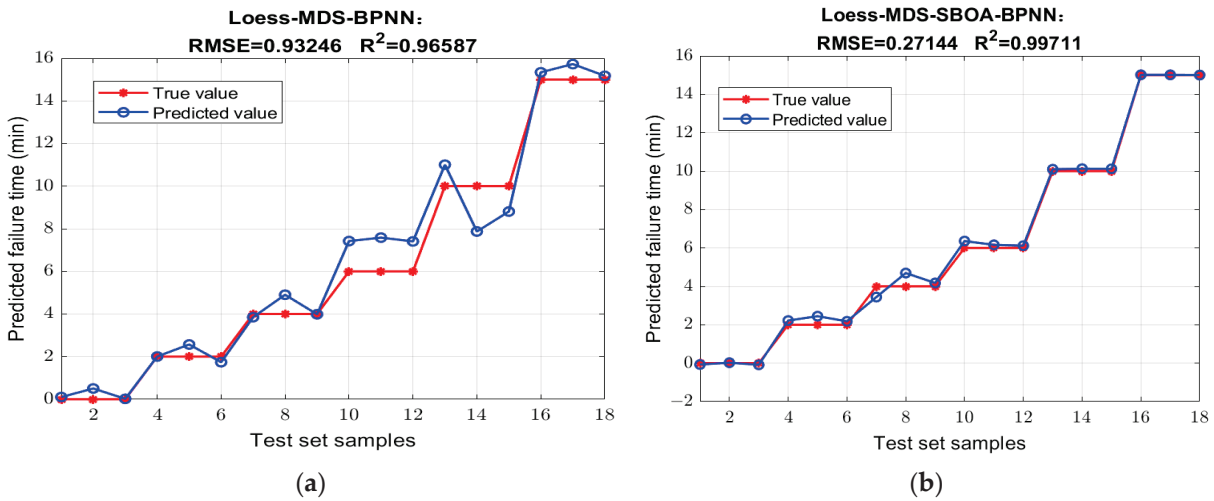


Figure 11. Comparison of the best Loess-MDS-SBOA-BP model before and after optimization. (**a**) Lo-ess-MDS-BPNN; (**b**) Loess-MDS-SBOA-BPNN.

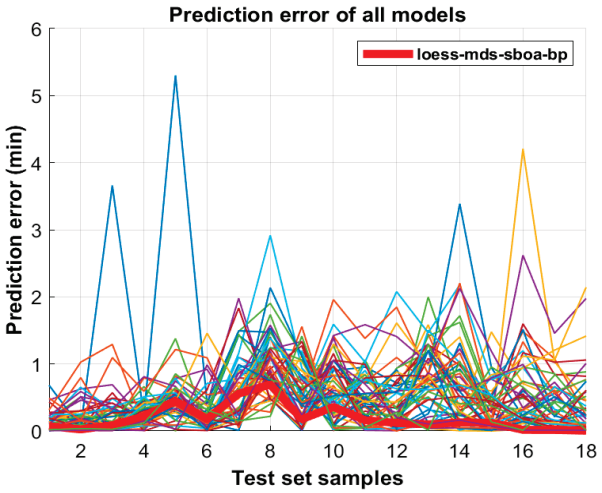


Figure 12. The prediction error curves of all models in the test set.

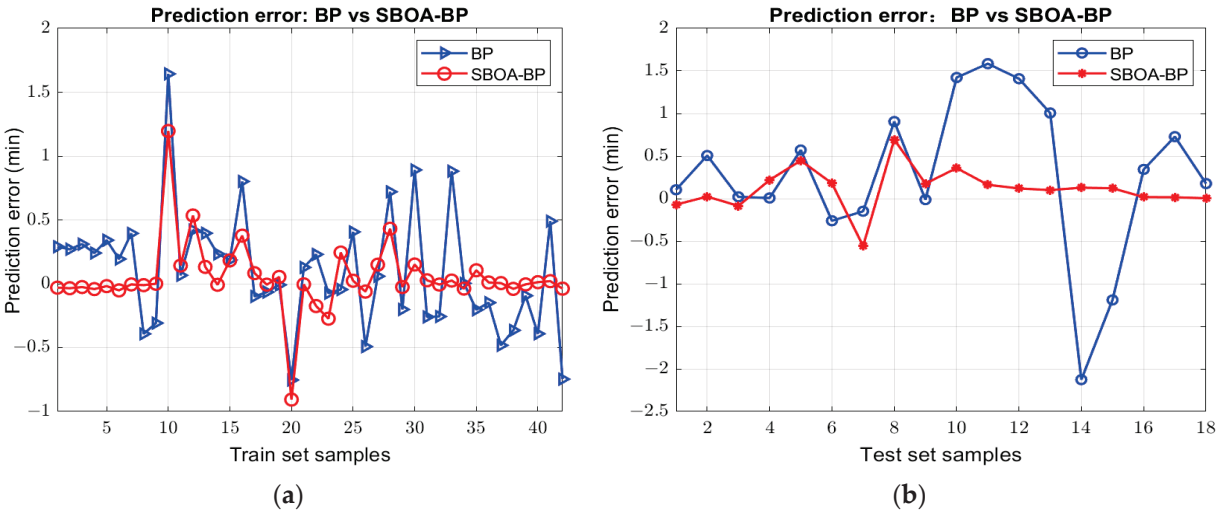


Figure 13. Prediction error of Loess-MDS-SBOA-BPNN before and after optimization. (**a**) Training set; (**b**) test set.

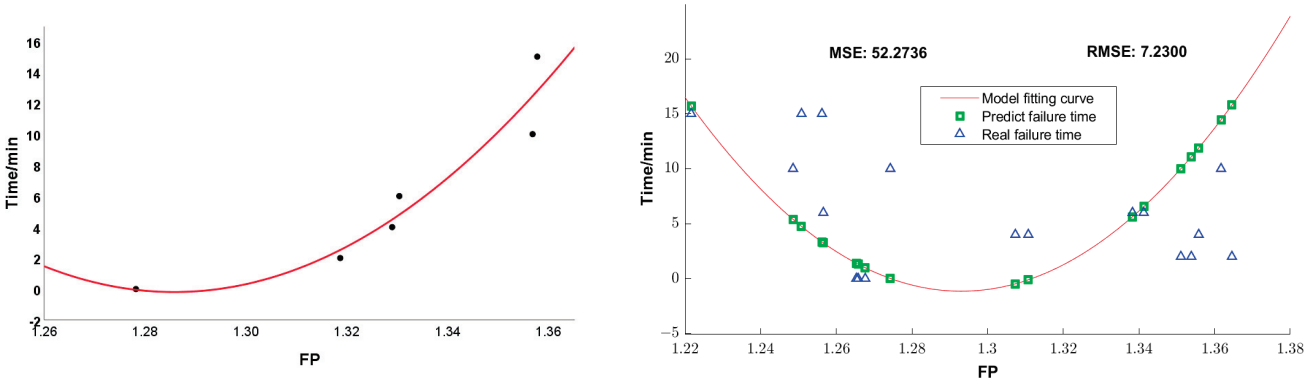
3.6. Fluorescence Double-Color Ratio

The fluorescence double-color ratio method is a method that has been proposed in recent years for quantitative fault diagnosis based on the analysis of fluorescence characteristic peak intensity data. Based on the analysis of the fluorescence signals in this experiment, two characteristic peaks in the ranges of 308–311 nm and 329 nm–332 nm are extracted as the double-color information extraction bands, and the feature proportions (FP) are shown in Table 7. The fluorescence two-color ratio fault diagnosis model established by least squares fitting is T = 2500.64*(FP)^2 − 6430.51*(FP) + 4113.90, the correlation coefficient is 0.9157, and the fitting curve is shown in Figure 14.

Using the established double-color ratio diagnostic model, the failure time of the oil sample is deduced according to the fluorescence ratio value calculated by the measurement for the same test set. The result is shown by the green square mark in the figure, and the blue triangle mark is the real fault time. The MSE of the prediction results of the double-color ratio model is 52.2736 and RMSE is 7.2300, and this model's predicted values differ significantly from the true values. The reason is that the changes in the fluorescence multi-peak spectrum is complex, and the fluorescence ratio value cannot express the fault characteristics of the oil samples well. The result shows that the double-color ratio model is less effective in the application of multi-peak complex fluorescence spectra.

Table 7. Fluorescence spectral feature ratio value.

	0 min	2 min	4 min	6 min	10 min	15 min
FP	1.2782	1.3187	1.3290	1.3305	1.3569	1.3578



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

Figure 14. Double-color proportional failure diagnosis model of insulating oil. (**a**) Model fitting curve; (**b**) failure time diagnosis effect.

4. Conclusions

In this paper, fluorescence analysis and fault diagnosis are carried out for the interval discharge fault of transformer insulating oil, and the spectral data are directly correlated with the quantitative analysis results of the transformer fault. Comparing the diagnostic effects of the models under different preprocessing methods, the experimental results show that the Loess-MDS-SBOA-BP failure diagnosis model optimized by the SBAO algorithm has the best prediction effect, which can accurately predict the fault duration time. The R2 value of the model reaches 99.711% and the MSE of the model's test set is only 0.073678. The proposed model enables maintenance personnel to find potential safety hazards in time at the early stage of a fault and carry out fault repairs, and also provides a research basis for more adequate transformer oil fluorescence data collection methods and further development in the field of transformer fault diagnosis in the future.

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

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this study can be accessed by contacting the corresponding authors with a valid request.

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

References

1. Zhou, X.; Liu, Z.; Shi, Z.; Ma, L.; Du, H.; Han, D. Fault Diagnosis of the Power Transformer Based on PSO-SVM. In Proceedings of the 11th International Conference on Communications, Signal Processing, and Systems, Changbaishan, China, 23–24 July 2022. [CrossRef]

2. Dhiman, A.; Kumar, R. Fault Diagnosis of a Transformer using Fuzzy Model and PSO optimized SVM. In Proceedings of the 2023 IEEE 8th International Conference for Convergence in Technology (I2CT 2023), Pune, India, 7–9 April 2023. [CrossRef]
3. Zeng, B.; Guo, J.; Zhu, W.; Xiao, Z.; Yuan, F.; Huang, S. A Transformer Fault Diagnosis Model Based on Hybrid Grey Wolf Optimizer and LS-SVM. *Energies* **2019**, *12*, 4170. [CrossRef]
4. Yang, X.L.; Yue, M.; Jia, T.B.; Sun, A.Q. Transformer Fault Diagnosis Based on DGA Information Fusion. In Proceedings of the 2023 4th International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE 2023), Hangzhou, China, 25–27 August 2023; pp. 241–244. [CrossRef]
5. Ghoneim, S.S.M.; Taha, I.B.M. A new approach of DGA interpretation technique for transformer fault diagnosis. *Int. J. Electr. Power. Energy Syst.* **2016**, *81*, 265–274. [CrossRef]
6. Sahoo, S.; Chowdary, K.V.V.S.R.; Das, S. DGA and AI Technique for Fault Diagnosis in Distribution Transformer. In Proceedings of the Lecture Notes in Electrical Engineering, Bhubaneswar, India, 5–6 March 2020; pp. 35–46. [CrossRef]
7. Shi, Y.H.; Yang, Y.; Liao, Y.; Hong, L.Z.; Liao, M.Y.; Pang, S. Transformer fault diagnosis based on improved particle swarm optimization SVM. *Eng. J. Wuhan Univ.* **2023**, *56*, 1238–1244. [CrossRef]
8. Kong, W.; Wan, F.; Lei, Y.; Wang, C.; Sun, H.; Wang, R.; Chen, W. Dynamic Detection of Decomposition Gases in Eco-Friendly $C_5F_{10}O$ Gas-Insulated Power Equipment by Fiber-Enhanced Raman Spectroscopy. *Anal Chem.* **2024**, *96*, 15313–15321. [CrossRef]
9. Si, S.H.; Zhao, J.Z.; Zou, G.L.; Liu, F.; Liu, C.; Yan, G.Y. Application of Fluorescence Spectroscopy in Identification of Aromatic Components in Single Oil Inclusions. *Spectrosc. Spectral Anal.* **2020**, *40*, 1736–1740. [CrossRef]
10. Deepa, S.; Sarathi, R.; Mishra, A.K. Synchronous fluorescence and excitation emission characteristics of transformer oil ageing. *Talanta* **2006**, *70*, 811–817. [CrossRef] [PubMed]
11. Suwarno; Pasaribu, R. Effects of thermal aging on paper characteristics in paper-mineral oil composite insulation. In Proceedings of the 2017 International Symposium on Electrical Insulating Materials, Toyohashi, Japan, 11–15 September 2017; pp. 705–708. [CrossRef]
12. Boyekong, G.O.; Mengounou, G.M.; Nkouetcha, E.T.; Imano, A.M. Analysis of the Dielectric and Physicochemical Performances of Thermally Aged Natural Monoester Insulating Liquids. *IEEE Trans. Dielectr. Electr. Insul.* **2023**, *30*, 2498–2506. [CrossRef]
13. Markova, L.V.; Myshkin, N.K.; Makarenko, V.M. Fluorescence Method for Quick Transformer Oil Monitoring. *Chem. Technol. Fuels Oils* **2016**, *52*, 194–202. [CrossRef]
14. Zhao, Y.; Leng, W.; Lin, Q.; Song, Y.M.; Zhang, Y.K.; Li, K.P.; Liu, J.Y.; Xu, F. Study on Insulation States Monitoring Technology of Transformer Insulating Oils Based on Three-dimensional Fluorescence Spectroscopy. *Insul. Mater.* **2022**, *55*, 87–91. [CrossRef]
15. Yan, P.C.; Zhang, C.Y.; Sun, Q.S.; Shang, S.H.; Yin, N.N.; Zhang, X.F. LIF Technology and ELM Algorithm Power Transformer Fault Diagnosis Research. *Spectrosc. Spectral Anal.* **2022**, *42*, 1459–1464. [CrossRef]
16. Dai, R.Y. Research on Rapid Fault Diagnosis of Transformers Based on Fluorescence Spectroscopy. Master's Thesis, Anhui University of Science & Technology, Huainan, China, 2022. [CrossRef]
17. Yan, P.C.; Zhang, C.Y.; Mei, K.F.; Chen, F.X.; Wang, Y.H. Research on fault diagnosis of transformer based on laser induced fluorescence technology. *J. Mol. Struct.* **2022**, *1258*, 132645. [CrossRef]
18. Hu, F.; Hu, J.; Dai, R.; Guan, Y.; Shen, X.; Gao, B.; Wang, K.; Liu, Y.; Yao, X. Selection of characteristic wavelengths using SMA for laser induced fluorescence spectroscopy of power transformer oil. *Spectrochim. Acta Part A* **2023**, *288*, 122140. [CrossRef]
19. Zou, D.; Li, Z.; Quan, H.; Peng, Q.; Wang, S.; Hong, Z.; Dai, W.; Zhou, T.; Yin, J. Transformer fault classification for diagnosis based on DGA and deep belief network. *Energy Rep.* **2023**, *9*, 250–256. [CrossRef]
20. Zhang, Y.; Wang, Y.; Fan, X.; Zhang, W.; Zhuo, R.; Hao, J.; Shi, Z. An Integrated Model for Transformer Fault Diagnosis to Improve Sample Classification near Decision Boundary of Support Vector Machine. *Energies* **2020**, *13*, 6678. [CrossRef]
21. Hu, L.; Yin, L.; Mo, F.; Liang, Z.; Ruan, Z.; Wang, Y. A classification and prediction model with the sparrow search-probabilistic neural network algorithm for transformer fault diagnosis. *Int. J. Sens. Netw.* **2024**, *44*, 249–257. [CrossRef]
22. Zhao, Y.; Ma, F.X.; Wang, A.J.; Li, D.C.; Song, Y.M.; Wu, J.; Cui, F.X.; Li, Y.Y.; Cao, Z.C. Research on Electric Breakdown Fault Diagnosis Model of Transformer Insulated Oil Based on Fluorescent Double-Color Ratio. *Spectrosc. Spectral Anal.* **2022**, *42*, 1134–1138. [CrossRef]
23. Mu, L.; Li, D.C.; Wang, A.J.; Xie, J.; Zhao, Y.; Sun, X.B. Fluorescence Spectroscopy-based very Sensitive Detection Technique for Insulating Oil Conditions. In Proceedings of the 2023 International Conference on Optoelectronic Information and Functional Materials (OIFM 2023), Guangzhou, China, 14–16 April 2023. [CrossRef]
24. Wang, Y.; Guo, X.; Liu, X.; Liu, X. Research on a Fault Diagnosis Method of an A-Class Thermal Insulation Panel Production Line Based on Multi-Sensor Data Fusion. *Appl. Sci.* **2022**, *12*, 9642. [CrossRef]
25. Xie, Y.C.; Wang, R.F.; Cao, Y.; Liu, K.; Gao, X.M. Research on Detecting CO_2 with off-Beam Quartz-Enhanced Photoacoustic Spectroscopy at 2.004 μm. *Spectrosc. Spectral Anal.* **2000**, *40*, 2664–2669. [CrossRef]
26. Zhang, X.; Yang, K.; Zheng, L. Transformer Fault Diagnosis Method Based on TimesNet and Informer. *Actuators* **2024**, *13*, 74. [CrossRef]

27. Song, Z.; Xiao, D.; Wei, Y.; Zhao, R.; Wang, X.; Tang, J. The Research on Complex Lithology Identification Based on Well Logs: A Case Study of Lower 1st Member of the Shahejie Formation in Raoyang Sag. *Energies* **2023**, *16*, 1748. [CrossRef]
28. Kennedy, J.; Eberhart, R. Particle Swarm Optimization. In Proceedings of the ICNN'95—International Conference on Neural Networks, Perth, Australia, 27 November–1 December 1995; Volume 4, pp. 1942–1948. [CrossRef]
29. Fu, Y.; Liu, D.; Chen, J.; He, L. Secretary bird optimization algorithm: A new metaheuristic for solving global optimization problems. *Artif. Intell. Rev.* **2024**, *57*, 1–102. [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.

sensors

MDPI

Article

Deep Learning Based Pile-Up Correction Algorithm for Spectrometric Data Under High-Count-Rate Measurements

Yiwei Huang [1,2], Xiaoying Zheng [1,2,*], Yongxin Zhu [1,2,*], Tom Trigano [3], Dima Bykhovsky [3] and Zikang Chen [1,2]

[1] Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China; huangyw@sari.ac.cn (Y.H.); chenzk@sari.ac.cn (Z.C.)
[2] University of Chinese Academy of Sciences, Beijing 101408, China
[3] Electrical and Electronics Engineering Department, Shamoon College of Engineering, Be'er Sheva 8410802, Israel
* Correspondence: zhengxy@sari.ac.cn (X.Z.); zhuyongxin@sari.ac.cn (Y.Z.)

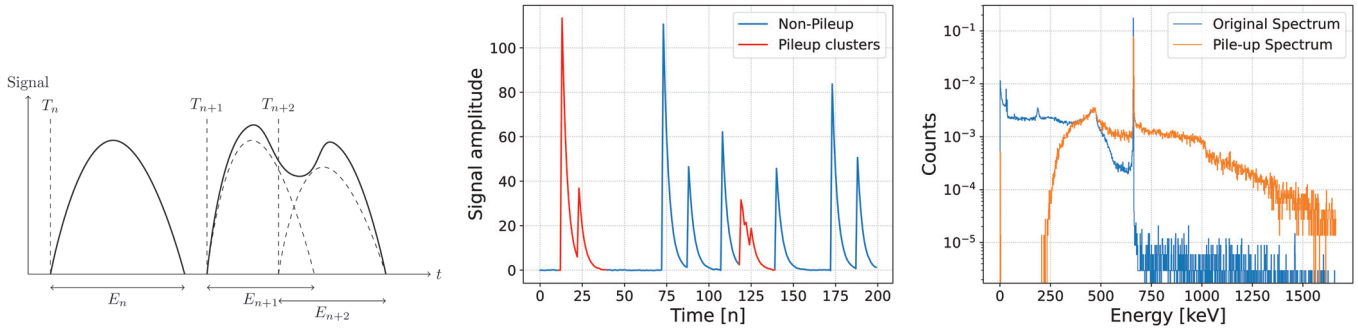
Abstract: Gamma-ray spectroscopy is essential in nuclear science, enabling the identification of radioactive materials through energy spectrum analysis. However, high count rates lead to pile-up effects, resulting in spectral distortions that hinder accurate isotope identification and activity estimation. This phenomenon highlights the need for automated and precise approaches to pile-up correction. We propose a novel deep learning (DL) framework plugging count rate information of pile-up signals with a 2D attention U-Net for energy spectrum recovery. The input to the model is an Energy–Duration matrix constructed from preprocessed pulse signals. Temporal and spatial features are jointly extracted, with count rate information embedded to enhance robustness under high count rate conditions. Training data were generated using an open-source simulator based on a public gamma spectrum database. The model's performance was evaluated using Kullback–Leibler (KL) divergence, Mean Squared Error (MSE) Energy Resolution (ER), and Full Width at Half Maximum (FWHM). Results indicate that the proposed framework effectively predicts accurate spectra, minimizing errors even under severe pile-up effects. This work provides a robust framework for addressing pile-up effects in gamma-ray spectroscopy, presenting a practical solution for automated, high-accuracy spectrum estimation. The integration of temporal and spatial learning techniques offers promising prospects for advancing high-activity nuclear analysis applications.

Keywords: high count rate; pulse pile-up; deep learning; nuclear spectroscopy

1. Introduction

Gamma-ray spectroscopy is an important tool in the field of nuclear science, enabling the identification and characterization of radioactive materials by analyzing the energy spectra of emitted gamma rays. This technique is widely used in various applications, including nuclear security, environmental monitoring, and geological exploration. Analyzing gamma-ray spectra, particularly at high count rates, is often complicated by the "pile-up" effect, as illustrated in Figure 1. The overlap of electrical pulses, particularly present at high activities, results in spectral distortions and inaccuracies in the estimation of activity [1]. Various correction strategies have been developed to address the pile-up issue. The most common approach is pile-up rejection, which detects pile-up events and discards them to retain individual pulses only [2]. Although effective, this method inevitably increases

the duration of experiments and can lead to a loss of valuable data. Some methods attempt to compensate for rejected pulses in the time domain by correcting the start point of the pile-up pulse, as referenced in [3,4]. Other techniques include pulse clipping [5] and template fitting [6]. These methods are typically most effective in systems with low pile-up probabilities and minimal pulse-shape dictionaries. Other research has considered the output signal as a linear combination of known shapes, relating the pile-up effect to a sparse regression problem [7].



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

Figure 1. An example of signal pile-up phenomenon and its effect on analyzing spectrum. At a count rate of 0.05, that is 0.05×10^7cps, the duty cycle is given by 0.82 with a theoretical pile-up probability of 0.558 (numerical details are introduced in Section 4). (**a**) Illustration of electrical pulses and their stacking phenomenon. (**b**) Raw pile-up signal and pile-up clusters (red). (**c**) The true spectrum of ^{137}Cs (blue) and the distorted spectrum (orange) were obtained by directly calculating the pile-up clusters.

Rather than recognize individual pulses in the time domain, recent efforts seek to analyze pulse collections; in [8], the authors formulated a non-linear inversion problem and developed a non-parametric pile-up correction method. In [9], the authors reformulated the pile-up problem as a decompounding problem, constructing a kernel-based estimator to model pulse pile-up. Similarly, ref. [10] used the characteristic function of the energy spectrum to infer the distribution of gamma photon energies from indirect measurements.

In addition to pile-up problems, traditional methods for isotope identification rely on the expertise of nuclear scientists who manually discriminate between background and source photopeaks, adjust calibrations, and account for shielding effects [11]. The subjective nature of this analysis and the increasing demand for automated, real-time solutions have driven the exploration of machine learning techniques to address these challenges. In particular, deep neural networks (DNNs) have emerged as a powerful approach to automated gamma-ray spectroscopy. These algorithms can learn complex patterns from the data, mimicking the intuitive decision-making process of experienced professionals. They have shown promising results in the identification of gamma-ray spectra under various conditions, including large calibration drifts and unknown background radiation fields [12]. The application of DNNs to spectroscopy has been extensive, with uses in peak fitting, isotope identification, and activity estimation [13]. Jeon et al. [14] developed a convolutional neural network (CNN) that attempts to separate and estimate the true pulse height of pile-up events from high-count-rate radiation. A DL-based technique for detecting and classifying pulses under high count rates was proposed in [15]. In [16], a CNN-based automated isotope identification method was proposed. In [17], another DL-based architecture was proposed for activity estimation. A pulse height analysis technique based on deep learning using U-net was introduced in [18].

This study proposes a DL-based method for processing pulse pile-up clusters, enabling accurate spectrum estimation for different sources at very high count rates. Our method

focuses on the pile-up correction algorithm. Specifically, we construct the Energy–Duration histogram of the pile-ups signals observed over a certain time period as the model input and employ a hybrid estimation model combining attention U-Net and known signal activity information, embedding activity information into the neural network to predict the corrected energy spectrum. Our training data are generated from an open-source simulator [19,20]. Our experiments show that under very high activity conditions of 2.95×10^6 count per second (cps), and of pile-up probability 0.996, the proposed hybrid DL estimation model is capable of recovering severe pile-up spectrum of various sources and achieving accurate prediction results.

Our contribution can be summarized as follows:

1. Spectrum Estimation Model for High Count Rates: A novel model is proposed for pile-up correction and spectrum estimation for high-count-rate scenarios. It achieves accurate predictions even in heavily distorted conditions, addressing the challenges in high-activity gamma source analysis;
2. Innovative Input Design with Energy–Duration Matrix: The Energy–Duration matrix, constructed via zero-crossing segmentation, is introduced as the input for 2D-UNet. This design effectively represents the mandatory information for pile-up correction;
3. Hybrid Model Combining Temporal and Spatial Features: This work integrates activity information and 2D-UNet architectures to extract temporal dependencies and spatial features from pulse signals. The embedding of count rate information enhances the robustness and accuracy of spectrum recovery under high count rate conditions.

The paper is organized as follows. We introduce the signal formulation and problem model in Section 2. The proposed DL algorithm is outlined in Section 3 and the experiments are carried out in Section 4, including the simulation details, and the model's performance evaluations are detailed in Section 5.

2. Problem Formulation

In this section, we introduce the problems that this paper aims to solve, including the theory of signal composition, the explanation of the pile-up effect, the generation method of training and test data, and the construction method of the Energy–Duration matrix.

2.1. Pile-Up Description and Correction Theory

In gamma-ray spectroscopy, "pile-up" refers to the phenomenon where multiple events, such as pulse signals, overlap within a short time window due to a high count rate, causing the signals to be hard to separate. This phenomenon is commonly observed when the detector's response time is limited, or its signal processing capacity is insufficient. Pile-up affects the energy spectrum's accuracy and can lead to spectrum distortion or loss of signals.

Assume that the detector records multiple events $x_i(t)$ within a given time window. The detector's overall response $y(t)$ will be the sum of the signals stemming from all the events. This can be expressed as:

$$y(t) = \sum_{i=1}^{N} x_i(t - t_i), \tag{1}$$

where $x_i(t)$ represents the signal from the i-th event, t_i is the time stamp of that event, and N is the number of events that occur within the current time window. To understand the impact of pile-up on energy measurements, let E_i denote the energy of the i-th event. In the case of pile-up, the total energy E_{total} will be the sum of the energies of N overlapping events, which can be written as:

$$E_{total} = \sum_{i=1}^{N} E_i. \tag{2}$$

Let $E_{\text{true}} = \{E_1, E_2, ..., E_N\}$, if the true energies in set E_{true} belongs to particular bins $b(E_i)$, the pile-up energy E_{total} may instead fall into an incorrect bin $b(E_{total})$, leading to a distortion in the histogram:

$$H(E) = \sum_{i=1}^{N} \delta(b(E_i)), \tag{3}$$

where δ is the Dirac delta function that places each signal's energy in its corresponding bin. In the case of pile-up, the distribution of E_{total} across the bins will be incorrect, leading to errors in the energy statistics and, ultimately, in the interpretation of the spectrum, such as the generation of artificial peaks or the suppression of true peaks.

Our goal is to correct the E_{true} from $E_{distorted}$, where $E_{distorted} = \{E_{total_1}, E_{total_2}, \dots E_{total_n}\}$. To address this issue, similar to [21], we regard this as a nonlinear mapping problem:

$$E_{true} = f(E_{distorted}), \tag{4}$$

where f is non-linear because the energy distribution of pile-up events is influenced by various factors, such as the detector's response time, the time interval between overlapping events, and the energy DL model to solve the pile-up correction problem.

2.2. Signal Formula

In order to create our dataset, we rely on synthetic signals, simulated in the time domain as follows. The detector output can be modeled as a linear combination of pulse shapes. A signal consisting of N events can be modeled as follows:

$$X(n) \triangleq \sum_{k \geq 1}^{N} \mathcal{E}_k \phi_k(n - T_k) + \varepsilon(n), \tag{5}$$

where $\mathcal{E}_k$ represents each photon's where energy which is drawn randomly from an energy distribution based on known spectrum data H, T_k is the sample path of a homogeneous Poisson process with constant known intensity λ, $\varepsilon(n)$ denotes the additive Gaussian noise, and pulse shape ϕ_k, is randomly chosen from a dictionary A. The signal construction rules are described as follows

1. Shape dictionary: One of the common shape models is the double exponential [1] which can be modeled as:

$$h(t) = a\left(\exp\left(-\frac{t}{\tau_2}\right) - \exp\left(-\frac{t}{\tau_1}\right)\right), \tag{6}$$

where a is the normalization factor and $\tau_1 \ll \tau_2$ are characteristic decay times. For practical purposes, any pulse shape is assumed to have a finite time length $\mathcal{T}$. A key characteristic of the pulse shape is the peak time, t_r, also known as the rise time, which is complementary to the fall time $t_f = \mathcal{T} - t_p$. For instance, the rise time for the double exponential model is given by:

$$t_r = \frac{\tau_1 \tau_2}{\tau_2 - \tau_1} \ln \frac{\tau_2}{\tau_1}, \tag{7}$$

where $\tau_1 \ll \tau_2$ are the decay times. In practice, a single precise parameter set cannot perfectly fit the pulse shape due to the physical inhomogeneities, temperature fluctu-

ation, and other factors, leading us to generate our parameter set randomly from a Gaussian distribution with reasonable expectations and variances [22].

2. Arrival times: as mentioned above, the arrival of gamma particles is modeled using a Poisson counting process with a constant intensity of λ. The inter-arrival times follow an exponential distribution with expectation $1/\lambda$.
3. Source: The spectrum of the signal can be either single source or multi-sources, the interactions among a mixture of multiple elements can be formed as:

$$P(E = e) = \sum_{i}^{n} \frac{w_i}{\sum_i w_i} P_{e_i}(E = e), \tag{8}$$

where P_{e_i} represents the probability that e_i emits energy e, E is the signal source's spectrum, and w_i is the proportion of source e_i. Thus, the spectrum of the mixture sources can be modeled as the following:

$$H_{mix} = \sum_{i}^{n} \frac{w_i}{\sum_i w_i} E_i. \tag{9}$$

By doing so, the probability density function of the mixture histogram can be derived.

4. Energy–Probability: The common representation of the probability of arrival particles' energies per energy bin H_i (typically measured by keV) is by a histogram that represents the energies of a series of signal events. This energy histogram can be easily modeled by a discrete random distribution where the probability of energy is proportional to b_i:

$$\Pr(H = H_i) = \frac{b_i}{\sum_i^B b_i}, \tag{10}$$

where b_i is the number of events with i_{th} energy, B represents the bin index, and the pairs of values for H_i and b_i of a single source or mixed source can be determined by the accept–reject method [23]. An example of a mixed spectrum is shown in Figure 2.

Pile-up events occur when multiple pulses appear within a single trace. To assess the probability of pile-up, the duty cycle (DC) value is used as a convenient parameter. The DC is calculated by multiplying the arrival count rate by the pulse shape length:

$$DC = \lambda \mathcal{T}. \tag{11}$$

Based on the probability relation,

$$\Pr(N(t) = n) = \frac{(\lambda t)^n}{n!} \exp -(\lambda t), \tag{12}$$

the resulting pile-up probability $\mathcal{P}$ is given by:

$$\mathcal{P} = \Pr(N(\mathcal{T}) > 0) = 1 - \Pr(N(\mathcal{T}) = 0) = 1 - \exp(-DC). \tag{13}$$

Well-developed standard methods can be used to generate sequences that fit a particular exponential distribution [23].

In this study, we consider DC values above 0.1 to have a high pile-up probability.

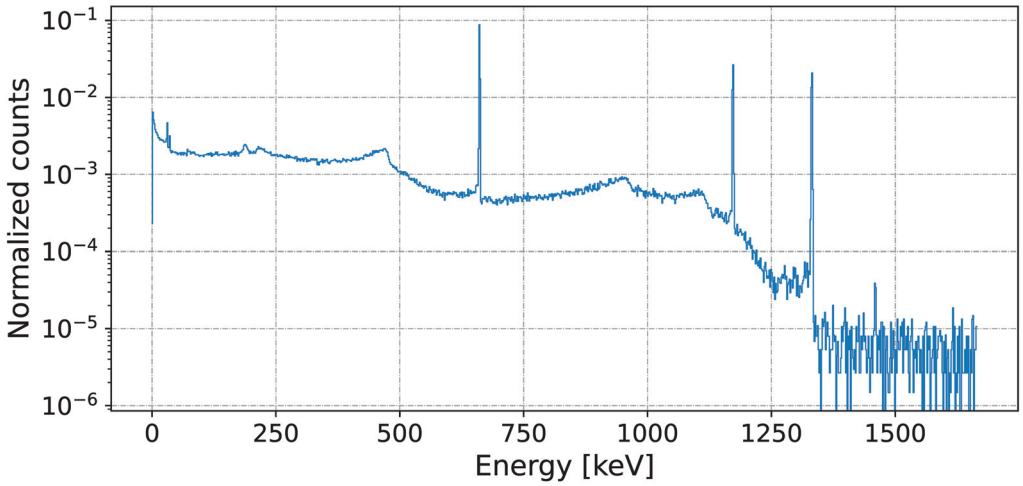


Figure 2. An example of a 1:1 mixture of ^{137}Cs and ^{60}Co.

2.3. Dataset Generation

We used different sources and their mixtures and different activities to generate pile-up signals based on a Gamma spectroscopy database and an open-source simulator. The sources are chosen from [^{60}Co, ^{137}Cs, ^{133}Ba, ^{57}Co, ^{152}Eu, ^{22}Na, ^{225}Ac, ^{241}Am, ^{131}Ba, ^{51}Cr, ^{169}Er, ^{197m}Hg, ^{124}I, ^{125}I, ^{131}I, ^{40}K, ^{138}La, ^{176}Lu,^{210}Pb, ^{226}Ra, ^{186}Re, ^{188}Re, ^{153}Sm]. The raw spectrum from the database is 8192 bins, and the maximum energy is 1666 keV; to simplify the representation and calculations, we processed the 8192 bins of data into 1024 bins by merging adjacent bins together, and we described the spectra using bin index rather than the energy values. The detailed parameters are listed in Table 1.

Table 1. Simulator parameter settings during data generation.

Parameter	Value
Count rate (λ)	0.05~0.3
Energy bins (B)	1024
Noise level (σ)	10^{-1}
Signal length per sample ($\mathcal{L}$)	0.2048
Sampling rate (f_S)	10^7
Shape type	double exponential
Shape factor	τ_1: $\mu = 3 * 10^{-7}, \sigma = 10^{-8}$ τ_2: $\mu = 10^{-8}, \sigma = 10^{-10}$

2.4. Pulse Signal Segmentation

The time signals are first processed by detecting zero-crossing points, which are used to segment the signals into pulse clusters. Zero-crossing points are identified as the points where the signal transitions from positive to negative (or vice versa). The signal is then divided into segments $\{s_1, s_2, \ldots, s_N\}$, where each segment s_i is bounded by two consecutive zero-crossing points.

The energy E_i of each segment s_i is computed as the sum of the absolute values of the signal within that segment:

$$E_i = \sum_{n \in s_i} |X(n)|$$

The duration D_i of each segment is simply the length of the segment in terms of time:

$$D_i = \text{length}(s_i)$$

Thus, each segment is characterized by its energy E_i and duration D_i that contain either a pile-up of individual pulses or a single individual pulse.

2.5. Energy–Duration Matrix Construction

Based on [21], the minimal information required for pile-up correction is the couple durations/energies, and there exists a non-linear relation linking the 2D histogram of the pile-ups to the 2D histogram of the individual pulses. Since DNN architectures can estimate non-linear relations efficiently, it seems logical to use this 2D representation. To construct the input matrix for the neural network, the energy and duration values of all pulse clusters are accumulated into an Energy–Duration matrix M. The matrix is indexed by energy and duration, with each element $M(e, d)$ representing the count of pulse segments that have energy e and duration d.

Each segment s_i is assigned a position (e_i, d_i) in the matrix corresponding to its energy E_i and duration D_i, and the matrix is updated through a period time $\mathcal{L}$ of raw signal as follows:

$$\text{for each } i\text{: } M(e_i, d_i) \leftarrow M(e_i, d_i) + 1$$

This results in a sparse matrix M, where the value at position (e, d) represents the frequency of pulse segments with energy e and duration d.

Overall, our goal can be summarized as follows: recover the true energy spectrum H from a short period of distorted pulse signals $\mathcal{L}$ by processing the Energy–Duration matrix M along with known count rate λ using a DL model. The optimization objective is to ensure that the output energy spectrum $\hat{H}$ predicted by the model closely matches H by minimizing the discrepancy between the predicted and actual spectra. We define the pile-up spectrum recovery problem as:

$$\min_{\theta} L(\mathcal{D}(M(e, d); \lambda, \theta), H), \tag{14}$$

where $\hat{H}_i = \mathcal{D}(M(e, d); \lambda, \theta)$ is the predicted energy spectrum at the i-th energy bin by model $\mathcal{D}$, H is the true energy spectrum, $L(\cdot)$ represents the loss function, typically Mean Squared Error (MSE) measuring discrepancy between the predicted and true spectra, and we aim to find the optimal θ parameters of the model by optimizing the function.

3. Method

In this section, we introduce the workflow of the proposed novel approach, as shown in Figure 3, to recover the true energy spectrum from distorted pulse signals using a DL model that integrates the Energy–Duration matrix with real-time count rate information directly embedded into a 2D U-Net architecture. The overall methodology can be broken down into the following key steps:

1. Pulse Signal Preprocessing: The raw pulse signals are first processed to generate an Energy–Duration matrix, which is achieved by segmenting the signal based on zero-crossing points. The resulting Energy–Duration matrix contains the spatial characteristics of the signal and serves as a feature representation for the 2D U-Net model.
2. Embedding Count Rate Information into 2D U-Net: The true count rate information is directly embedded into the 2D U-Net model, representing the temporal feature, intuitively giving the model prior information about the intensity of signal pile-up. This allows the 2D U-Net to process both spatial and temporal features simultaneously.
3. Energy Spectrum Recovery: The 2D U-Net, now augmented with the Energy–Duration matrix and the count rate information, processes these inputs to generate the predicted energy spectrum. The output of the network is compared with the

true energy spectrum during training, and the model is optimized to minimize the reconstruction error using Mean Squared Error (MSE).

By directly embedding count rate information into the bottleneck, as well as utilizing the pixel-level feature extract ability of 2D U-Net to process the fine-grained Energy–Duration matrix data, our approach streamlines the process of energy spectrum recovery, improving robustness and accuracy under high-count-rate conditions.

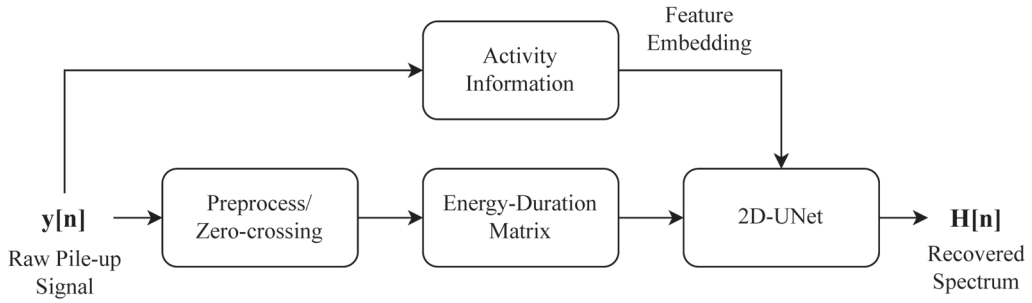


Figure 3. Workflow of our proposed method.

3.1. Count Rate Embedding Module

We directly embed the true count rate information into the 2D U-Net architecture to enhance its capability to recover the true energy spectrum from distorted pulse signals, as shown in Figure 4. The count rate embedding module is implemented as a fully connected neural network followed by a convolutional layer, which transforms the scalar count rate value into a high-dimensional feature map that is compatible with the 2D U-Net bottleneck. The structure of the embedding module is as follows: A fully connected linear layer that maps the scalar input λ to a 16-dimensional space, enabling the model to learn an intermediate representation of the count rate. A ReLU activation function is applied after the first linear transformation, introducing non-linearity to facilitate complex feature learning. A second linear layer projects the 16-dimensional representation into a higher-dimensional space corresponding to the energy dimension of the input data, effectively encoding the count rate information. A 1D convolutional layer that processes the count rate feature map, producing an output with a size consistent with the duration of the signal, thus enabling spatial processing of the count rate features.

The processed count rate feature from the embedding module is integrated into the 2D U-Net bottleneck via tensor addition. Specifically, after the count rate embedding module generates a high-dimensional feature map matching the dimensions of the bottleneck tensor, the two tensors are added element-wise. This operation ensures that the count rate information is directly infused into the 2D U-Net's intermediate representation, enabling the model to jointly consider both temporal dynamics (captured by the Energy–Duration matrix) and signal intensity (represented by the count rate).

By integrating the count rate information at the bottleneck, the model can jointly process the temporal dynamics (through the Energy–Duration matrix) and the signal intensity (via the count rate), leading to improved performance in estimating the true energy spectrum, particularly under high-count-rate conditions.

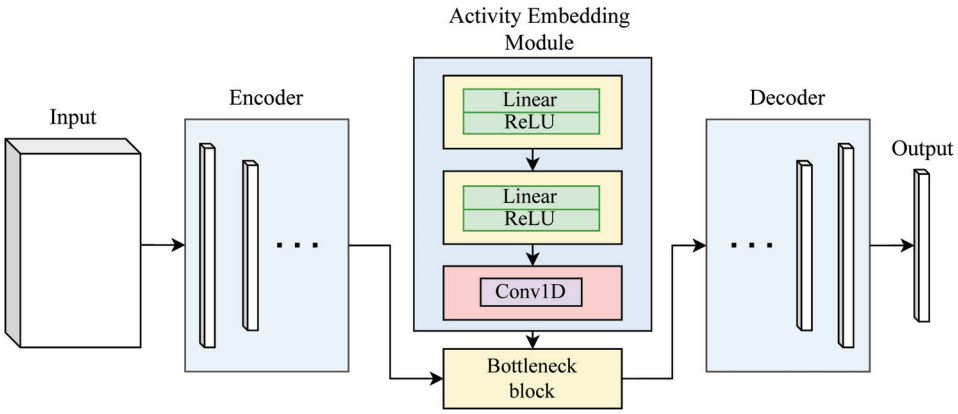


Figure 4. Schematic diagram of the neural network for embedding count rate features. The encoding and decoding of data are simply represented by white cubes.

3.2. 2D-UNet

3.2.1. Attention U-Net Model Architecture

The proposed model employs an Attention U-Net architecture [24], which consists of an encoder, a decoder, and attention gates applied to the skip connections between the encoder and decoder sub-networks, as shown in Figure 5. The attention mechanism in this architecture allows the network to focus on the most relevant parts of the input signal, improving its ability to learn meaningful features while discarding irrelevant information. This results in more accurate energy spectrum recovery, especially for high-dimensional data.

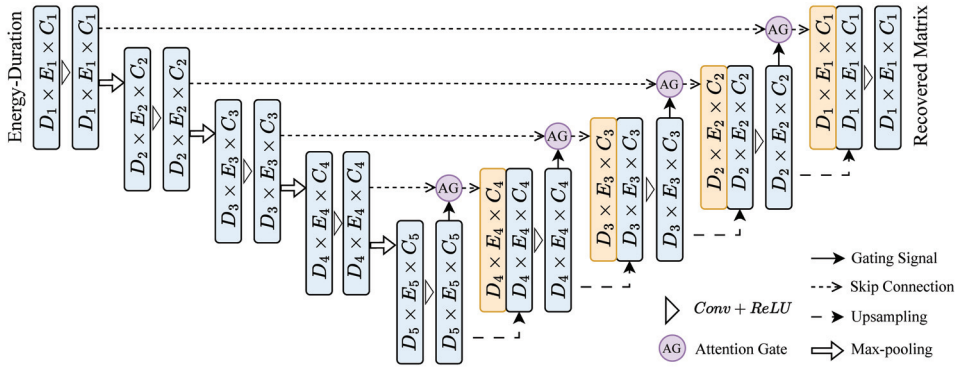


Figure 5. Architecture of the Attention U-Net model used for Energy–Duration matrix recovery. The model processes the input Energy–Duration matrix (size of $D_1 \times E_1 \times C_1$) through an encoder–decoder structure with skip connections and attention gates (AGs). The encoder extracts hierarchical features through convolution and max-pooling layers, reducing spatial dimensions (D, E) while increasing channel depth (C). The decoder progressively upsamples the feature maps, integrating encoder features via skip connections and attention mechanisms to selectively focus on relevant regions. The decoder output is the recovered matrix with the same spatial dimensions as the input.

The encoder–decoder architecture is similar to that of the original U-Net [25] but with the key modification of incorporating attention gates. These gates allow the decoder to focus on specific regions of the feature maps received from the encoder, thus promoting the learning of multi-scale information and improving feature propagation.

Let $\mathbf{X} \in \mathbb{R}^{B \times E \times D \times C}$ represent the input matrix to the model, where B is the batch size, E is the energy size, D is the duration, and C is the number of channels. For each input instance $\mathbf{x} \in \mathbb{R}^{E \times D}$, the goal of the Attention U-Net model is to predict the energy spectrum vector $\hat{\mathbf{y}} \in \mathbb{R}^{\mathcal{L} \times 1}$, where $\hat{\mathbf{y}}$ is a vector of length $\mathcal{L}$ representing the corrected energy spectrum. The network is trained end-to-end to minimize the reconstruction error between the predicted and true energy spectra.

3.2.2. Encoder Architecture

The encoder is responsible for extracting hierarchical feature representations from the input matrix **X**. It consists of several convolutional blocks, each followed by a ReLU activation and a max-pooling operation to downsample the feature maps. The feature maps are progressively downsampled in spatial resolution, but their depth (number of channels) is increased at each layer to capture more abstract features.

Let $f_{\text{input}} = \mathbf{X}$ be the input to the first convolutional layer of the encoder. At each subsequent layer, the output f_i is computed as:

$$f_i = \text{Conv}(\text{ReLU}(\text{Conv}(f_{i-1})))$$

where $\text{Conv}(\cdot)$ denotes a convolution operation and $\text{ReLU}(\cdot)$ is the rectified linear unit activation. The encoder reduces the spatial dimensions of the input matrix while increasing the feature depth, allowing it to learn increasingly complex patterns in the data.

3.2.3. Decoder Architecture

The decoder reconstructs the output energy spectrum from the features extracted by the encoder. It consists of several transposed convolutional layers (also known as up-sampling or deconvolution layers), which progressively increase the spatial resolution of the feature maps. At each layer of the decoder, skip connections from the corresponding encoder layers are incorporated to retain high-resolution features.

The decoder's input is the feature map from the last layer of the encoder, denoted as f_{encoded}. The decoder progressively upsamples this feature map to recover the input's spatial dimensions. The feature map at layer i of the decoder, denoted f_{dec}^{i}, is computed as

$$f_{\text{dec}}^{i} = \text{Upsample}(\text{Conv}(f_{\text{dec}}^{i-1}))$$

where $\text{Upsample}(\cdot)$ denotes an up-sampling operation, and $\text{Conv}(\cdot)$ is a convolution operation applied to the upsampled feature map.

3.2.4. Attention Gates

In our case, the Energy–Duration matrix we constructed is typically sparse, in which the peak feature is fine-grained and spatial-related. The AG implementation enables the model to filter out irrelevant background regions and focus on the target sum-up peaks.

The key feature of the Attention U-Net architecture is the use of attention gates that control the flow of information through the skip connections. These gates allow the model to focus on important spatial regions by dynamically weighting the importance of the features received from the encoder.

The attention map A_i is computed dynamically for each layer based on the feature maps passed through the attention gate. At each skip connection between the encoder and decoder, the feature map f_{encoded}^{i} is processed by an attention gate, which is defined as:

$$A_i = \sigma(f_{\text{skip}} \cdot W_{\text{skip}} + f_{\text{encoded}} \cdot W_{\text{enc}}),$$

where $\sigma(\cdot)$ is the sigmoid activation function, W_{skip} and W_{enc} are learnable weight matrices, and f_{skip} is the feature map from the decoder and f_{encoded} is the feature map from the encoder at layer i. In practice, the weight matrices W_{skip} and W_{enc} are initialized randomly and updated during training using backpropagation.

The output of AGs is the element-wise multiplication of input feature maps and attention maps:

$$f_{\text{att}}^{i} = A_i \odot f_{\text{skip}},$$

where $\odot$ denotes element-wise multiplication. This operation selectively focuses on the important features while suppressing irrelevant ones.

Applying the attention gates allows the network to adjust its focus on different regions of the input, based on the spatial relevance of the features, which improves the recovery of important energy regions.

3.2.5. Final Output Layer

The output of Attention U-Net's decoder is a reconstructed de-pile-up energy spectrum feature. The final output $\hat{\mathbf{y}} \in \mathbb{R}^{\mathcal{L}\times 1}$, is obtained by firstly applying a 1D convolutional layer to the output of the decoder which refines the feature representations, and followed by an adaptive average pooling operation for extracting features from a sparse matrix. Finally, the pooled features are flattened and passed through a multi-layer perceptron (MLP) to map the output to a vector of dimensions identical to the target energy spectrum

$$\hat{\mathbf{y}}_{hidden} = \text{Conv1D}(f_{\text{dec}}^{N}),$$

$$\hat{\mathbf{y}} = \text{MLP}(\text{AvgPool2d}(\hat{\mathbf{y}}_{hidden}))$$

where f_{dec}^{N} is the final feature map from the decoder, AvgPool2d is the adaptive average 2d pooling, Conv1D$(\cdot)$ is a 1D convolution operation, and $\hat{\mathbf{y}}_{hidden}$ is the output spectrum's hidden state. The output $\hat{\mathbf{y}}$ represents the predicted energy spectrum, which is compared with the true spectrum during training.

3.3. Loss Function and Optimization

The optimization objective is to minimize the Mean Squared Error (MSE) between the predicted energy spectrum $\hat{y}$ and the true energy spectrum y:

$$\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2, \tag{15}$$

where N is the number of energy bins, $\hat{y}_i$ and y_i are the predicted and true energy values at the i-th energy bin.

4. Experiments

In this section, we describe the experimental setup, including the training and inference processes, and the datasets used for training the 2D U-Net model with count rate embedding.

4.1. Training Parameters

The total size of the dataset used for the training is 1800, we split the dataset into training (80%) and validation sets (20%), 1440 samples, and 360 samples, respectively. The training set consists of signals with activities in the range of [0.05, 0.3], with a step size of 0.01, while the dataset used for the test contains activities in the range of [0.05, 0.3] with a step size of 0.005, excluding the activities used for training. This setup ensures that the model's generalization ability is evaluated on unseen activity levels.

The model is trained for 200 epochs using the Adam optimizer with an initial learning rate of 1×10^{-4}. The batch size is set to 16, and training is performed using two GPUs in parallel. The learning rate is dynamically adjusted: if the validation loss does not improve, the learning rate is halved every 10 epochs. The training process is monitored to ensure convergence and prevent overfitting. Table 2 lists the detailed training parameters.

Table 2. Experimental settings.

Training Parameters	Value
Batch size	16
Epoch	200
Learning rate	0.0001
Optimizer	Adam
Train sample	1440
Test sample	360
Model depth	5
Hidden size	64
Kernel size	2
Stride	2

4.2. Hardware and Software Setup

All experiments were implemented using the PyTorch framework (https://pytorch.org/). Training was performed on two NVIDIA A30 GPUs in parallel. The model training and inference were carried out on a distributed system to speed up computation. The training environment included the use of mixed precision for efficient memory usage.

4.3. Evaluation Metrics

The performance of the 2D U-Net model was evaluated using four metrics: the Kullback–Leibler (KL) divergence, the Mean Squared Error (MSE), the Energy Resolution (ER), and the Full Width at Half Maximum (FWHM). These metrics were used to assess the accuracy of the predicted energy spectra on test datasets.

Kullback–Leibler (KL) divergence: The KL divergence is used to measure the difference between two distributions. We use the symmetric KL divergence (KL_s) to evaluate the probability distribution distance between the predicted spectra and the reference spectrum:

$$D_{KL}(P||Q) = \sum_{i=1}^{B} P(i) \cdot \log\left(\frac{P(i)}{Q(i)}\right), \tag{16}$$

and

$$KL_s = \left[\frac{1}{2}\sum_{i=1}^{B}(\mathrm{KL}(\mathcal{P}_i, \| \mathcal{Q}_i) + \mathrm{KL}(\mathcal{Q}_i \| \mathcal{P}_i))\right]. \tag{17}$$

Mean Squared Error (MSE):

$$\mathrm{MSE} = \frac{1}{B}\sum_{i=1}^{B}(S_i - \hat{S}_i)^2 \tag{18}$$

Energy Resolution (ER): We define ER to evaluate the accuracy of the model in estimating the position of key peaks in the energy spectrum, which is defined as the relative difference between the estimated peak energy and the true peak energy:

$$\mathrm{ER} = \frac{|E_{\text{estimated}} - E_{\text{true}}|}{E_{\text{true}}}$$

Additionally, we compare the performance of our method against two conventional methods:

1. Method 1: A baseline method that does not correct for pulse pile-up.
2. Method 2: A fast correction algorithm, as described in [21].

The model was trained using the training set and validated on the validation set to monitor performance. After training, the model was evaluated on the test set, and the results were compared with those of the two baseline methods.

5. Results and Discussion

The proposed method shows significant improvement in the energy spectrum recovery under high-count-rate conditions when compared to conventional techniques. As shown in Table 3 and Figure 6, the KL divergence and MSE values for our method are consistently lower than those for the baseline methods across all count rate levels. Also, our method achieves an accurate estimation effect when calculating the ER and FWHM metric, as shown in Table 4.

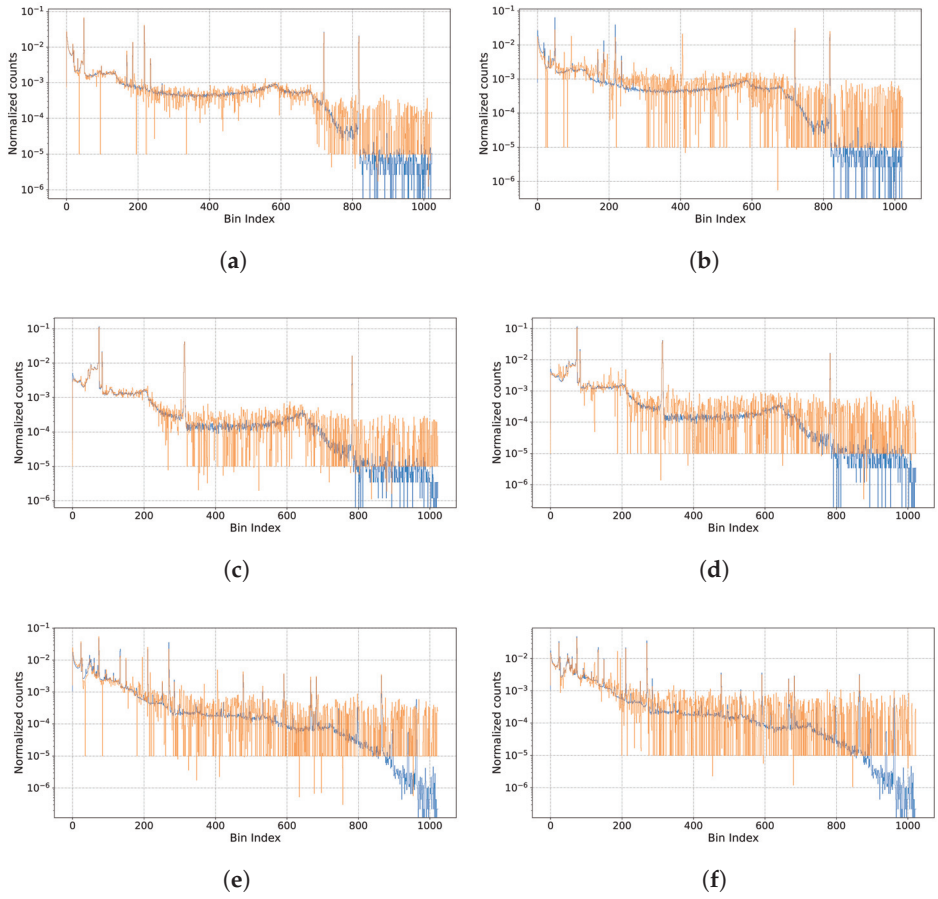


Figure 6. Pile-up spectrum recovery results of our methods, under different high count rates and multiple sources measurements. (**a**) $\lambda = 0.055, DC = 1.12, \mathcal{P} = 0.674,$ ^{60}Co,^{133}Ba, KL = 0.07, MSE = 0.7×10^{-7}. (**b**) $\lambda = 0.265, DC = 5.41, \mathcal{P} = 0.996,$ ^{60}Co,^{133}Ba, KL = 0.35, MSE = 32.14×10^{-7}. (**c**) $\lambda = 0.09$, $DC = 1.84, \mathcal{P} = 0.841,$ ^{57}Co,^{22}Na, KL = 0.06, MSE = 0.80×10^{-7}. (**d**) $\lambda = 0.23$, $DC = 4.69, \mathcal{P} = 0.991,$ ^{57}Co,^{22}Na, KL = 0.17, MSE = 3.37×10^{-7}. (**e**) $\lambda = 0.195, DC = 3.98$, $\mathcal{P} = 0.981,$ ^{152}Eu,^{225}Ac, KL = 0.20, MSE = 9.02×10^{-7}. (**f**) $\lambda = 0.265, DC = 5.41, \mathcal{P} = 0.996$, ^{152}Eu,^{225}Ac, KL = 0.24, MSE = 7.43×10^{-7}. The results demonstrate that our method achieves highly accurate peak estimation for increasing levels of pile-up and spectral distortion. The blue curve represents the real energy spectrum, and the orange curve represents the estimated result of our proposed method.

Table 3. The table shows the KL and MSE score calculated by three algorithms under different λ, from 0.055 to 0.295, $\mathcal{P}$ from 0.674 to 0.996. Multi-sources are mixed with a weight of 1:1, and the symbol '-' indicates that the algorithm fails under the current conditions. The results show that the proposed method outperforms both Method 1 and Method 2.

Activity	Source Name	Multi-Source												Single-Source							
		^{60}Co,^{133}Ba		^{137}Cs,^{57}Co		^{133}Ba,^{152}Eu		^{57}Co,^{22}Na		^{152}Eu,^{225}Ac		^{22}Na,^{241}Am		^{51}Cr		^{169}Er		^{197m}Hg		^{125}I	
λ_n	Metrics(/1 $\times$ 10^{-5})	KL	MSE	KL	MSE	KL	MSE	KL	MSE	KL	MSE	KL	MSE	KL	MSE	KL	MSE	KL	MSE	KL	MSE
0.055	Method 1	2.69	96.03	3.20	330.44	2.76	147.35	2.54	295.69	2.82	91.95	2.29	730.43	3.40	885.63	2.75	223.97	3.39	201.97	0.50	419.23
	Method 2	2.97	89.68	2.98	307.29	2.68	144.24	2.52	285.18	2.79	90.98	2.19	713.27	3.51	892.72	2.45	218.25	3.43	182.00	0.28	352.35
	Method 3 (Ours)	**0.07**	**0.70**	**0.08**	**0.44**	**0.07**	**0.46**	**0.08**	**0.86**	**0.06**	**0.68**	**0.09**	**4.80**	**0.05**	**0.27**	**0.07**	**19.44**	**0.08**	**0.40**	**0.08**	**0.60**
0.095	Method 1	3.05	101.82	4.35	347.86	3.35	156.91	3.55	305.31	3.32	97.89	3.62	748.12	3.20	939.75	3.97	253.52	3.87	222.89	1.48	663.49
	Method 2	4.10	96.47	5.25	329.31	3.77	158.07	4.30	294.98	3.89	98.04	4.19	739.54	3.81	850.03	4.13	260.33	4.76	187.65	1.14	615.32
	Method 3 (Ours)	**0.06**	**213.81**	**0.07**	**0.40**	**0.09**	**1.25**	**0.06**	**0.63**	**0.11**	**1.48**	**0.06**	**0.45**	**0.08**	**2.96**	**0.07**	**15.03**	**0.09**	**0.38**	**0.04**	**0.11**
0.135	Method 1	3.38	103.04	4.57	352.06	3.44	158.91	3.45	309.62	3.32	99.82	3.62	751.75	2.72	968.47	4.37	259.26	4.17	236.14	3.10	774.69
	Method 2	4.79	95.18	6.21	326.73	4.83	157.68	5.07	297.46	4.97	99.84	5.14	740.70	3.98	884.49	6.03	288.89	5.89	233.80	4.301	931.57
	Method 3 (Ours)	**0.06**	**0.74**	**0.07**	**0.49**	**0.07**	**1.26**	**0.11**	**1.27**	**0.12**	**1.43**	**0.11**	**0.75**	**0.260**	**72.87**	**0.09**	**0.64**	**0.1**	**0.40**	**0.04**	**0.07**
0.175	Method 1	3.30	103.82	4.25	354.77	3.28	159.87	3.65	311.28	3.16	100.40	3.51	753.43	2.86	978.56	4.21	260.36	3.90	236.81	3.91	808.39
	Method 2	5.20	99.69	6.41	321.99	5.60	163.61	5.76	311.84	5.53	102.83	5.57	752.71	4.59	974.64	7.29	408.54	6.19	231.41	9.40	-
	Method 3 (Ours)	**0.09**	**1.21**	**0.10**	**0.88**	**0.22**	**31.72**	**0.14**	**3.84**	**0.12**	**3.14**	**0.08**	**0.47**	**0.24**	**11.64**	**0.10**	**0.50**	**0.23**	**2.03**	**0.06**	**0.15**
0.215	Method 1	3.01	104.30	3.65	355.84	3.08	159.90	3.64	311.34	2.86	100.41	3.45	753.55	3.17	978.93	3.83	260.69	3.42	237.40	4.04	822.34
	Method 2	5.46	105.58	6.30	321.11	6.00	161.84	6.20	312.13	5.85	101.50	6.20	753.70	5.95	980.13	8.260	117.78	6.27	233.68	11.77	-
	Method 3 (Ours)	**0.12**	**1.18**	**0.12**	**0.58**	**0.32**	**46.21**	**0.22**	**9.80**	**0.22**	**4.74**	**0.22**	**7.55**	**0.53**	**124.44**	**0.16**	**1.90**	**0.37**	**7.94**	**0.05**	**0.09**
0.255	Method 1	2.47	104.50	2.94	356.43	2.95	159.90	3.38	311.37	2.59	100.40	3.25	753.56	2.83	979.03	3.42	260.86	3.35	237.53	3.86	826.66
	Method 2	5.69	101.06	6.46	333.69	6.22	161.14	6.47	311.65	6.21	101.33	6.48	753.36	6.35	980.00	8.97	5021.00	6.97	239.46	12.65	-
	Method 3 (Ours)	**0.35**	**8.02**	**0.25**	**5.01**	**0.44**	**63.32**	**0.24**	**4.62**	**0.59**	**47.49**	**0.28**	**12.11**	**0.77**	**529.59**	**0.16**	**1.76**	**0.62**	**25.68**	**0.05**	**0.19**
0.295	Method 1	2.09	104.61	3.77	356.58	6.74	159.90	3.92	311.41	2.09	100.45	2.53	753.60	2.52	979.02	3.40	260.87	4.30	237.54	3.80	828.18
	Method 2	6.68	107.30	6.61	340.03	2.88	161.51	6.67	311.98	6.66	101.52	6.84	753.53	6.59	980.29	7.59	291.10	7.48	239.89	13.20	-
	Method 3 (Ours)	**0.86**	**107.95**	**0.60**	**30.45**	**0.51**	**43.16**	**0.23**	**5.49**	**0.66**	**67.48**	**0.30**	**6.01**	**0.96**	**419.36**	**0.35**	**22.53**	**0.76**	**58.33**	**0.08**	**0.71**

The MSE metric is in units of 10^{-7}.

Table 4. The table shows the ER and FWHM metrics of the energy spectrum estimated by the model, calculated from the peak values of the sources. The experiments were carried out under different count rates under different λ, from 0.055 to 0.295, $\mathcal{P}$ from 0.674 to 0.996 and different sources: ^{60}Co, ^{137}Cs, and ^{241}Am. The results show that our model can accurately estimate the peak value at high count rates in most cases.

	Single Sources									
	^{60}Co				**^{137}Cs**				**^{241}Am**	
Activity	**1173.24**		**1332.50**		**31.80**		**661.66**		**59.54**	
λ_n	**ER**	**FWHM**	**ER**	**FWHM**	**ER**	**FWHM**	**ER**	**FWHM**	**ER**	**FWHM**
0.055	0.073	2.521	0.076	2.520	0.004	0.805	0.032	2.620	0.042	2.503
0.105	0.019	2.520	0.016	2.516	0.009	2.802	0.002	2.620	0.014	2.503
0.135	0.014	2.522	0.013	2.515	0.028	0.818	0.029	2.620	0.004	2.503
0.175	0.023	2.520	0.018	2.518	0.023	2.836	0.040	2.620	0.004	2.503
0.245	0.159	2.519	0.163	2.529	0.046	0.818	0.099	2.623	0.001	2.504
0.295	0.587	2.519	0.521	2.568	0.158	1.698	0.355	2.620	0.077	2.504

The peak position and FWHM are represented by energy (keV) and bins, respectively. One bin represents 1.6 keV.

5.1. Visualization

The visualized results are shown in Figure 6, where the blue curve represents the ground truth spectrum (labels), and the orange curve denotes the estimated spectrum.

The experiment was conducted under different signal pile-up levels, quantified by λ, and varying distortion complexities (DC). The values of λ range from 0.055 to 0.265, while DC ranges from 1.12 to 5.41, and pile-up probability($\mathcal{P}$) from 0.674 to 0.996. These parameters simulate increasing levels of pile-up and spectral distortion.

The results demonstrate that our method achieves highly accurate peak estimation across all conditions:

- In low pile-up conditions (e.g., Figure 6a, $\lambda = 0.055, DC = 1.12, \mathcal{P} = 0.674$), the estimated peaks closely align with the ground truth spectrum, showcasing minimal deviations.
- As the pile-up level increases (e.g., Figure 6f, $\lambda = 0.265, DC = 5.41, \mathcal{P} = 0.996$), slight deviations are observed, particularly in higher-energy regions. Nevertheless, the primary peaks remain well-estimated, and the overall spectral trends are preserved.
- Across all cases, the proposed method effectively reconstructs the peaks, even under severe pile-up conditions, highlighting its robustness and accuracy.

The experimental results validate the effectiveness of our method in accurately estimating spectral peaks under high-count rate scenarios. Despite increasing levels of pile-up and distortion, the method maintains reliable performance, demonstrating its robustness for applications in high-activity environments.

5.2. Quantification

In Table 3, we compare KL divergence and MSE metric of energy spectrum estimations using our proposed method against two conventional algorithms: one without pulse correction (Method 1) and the fast correction algorithm (Method 2) [21]. These comparisons are made using different count rates and various sources. In both multi-source and single-source scenarios, the KL divergence and MSE are significantly lower compare to Method 1 and Method 2, indicating more accurate spectrum estimation with smaller errors.

At a high activity ($\lambda = 0.055$), Method 3 shows the best performance (KL$\sim$0.05, MSE$\sim$0.27), while Methods 1 and 2 perform poorly (KL > 2.5, MSE > 90). When activity reaches ($\lambda = 0.095, 0.135$), Method 3 remains robust, while Methods 1 and 2 worsen. Finally, when the pile-up effect becomes more severe ($\lambda = 0.175$), Method 3 still performs well, while traditional methods have high KL (>5) and MSE (>900), highlighting its ability to recover the energy spectrum even in ultra high-count scenarios.

The results show that the model performs well across a range of activities, demonstrating its robustness and ability to generalize to unseen data. A critical challenge for traditional algorithms is their performance degradation as the count rate increases. Our approach incorporates count rate information as a conditioning factor, allowing it to maintain stable and accurate performance even under extremely high count rates. This robustness ensures reliable energy spectrum estimation, where traditional methods would typically fail or yield significant errors.

To better demonstrate the estimation effect of our model on the energy spectrum in more detail, we selected three sources to conduct experiments at different count rates to evaluate the model's prediction results for the peak value. Specifically, the ER metric and the Full Width at Half Maximum (FWHM) are evaluated and shown in Table 4. It can be seen that our model can accurately estimate both the peaks in the low-energy region and the peaks in the high-energy region, and the ER in the low-energy region is slightly better than that in the high-energy region. When the count rate gradually increases, the ER increases; however, until the pile-up probability increases to 0.996, the estimation error of our method for the high-energy peak increases significantly. Under high counting rate conditions, the model can effectively suppress the pile-up effect, and the peak width of the estimated energy spectrum is in a very small range (with FWHM < 3 bins), which shows that our method has effective and stable pile-up correction capability in high activity environments.

In order to explore the impact of Energy–Duration matrix inputs of different dimensions on model performance and the accuracy of estimation results, we conducted experiments using matrices of four different sizes, as shown in Table 5. Note that since the input dimension of $[256 \times 4096]$ is large, the training batch size is unified to 4 to ensure the consistency of training. The data show that in general, the input dimension of $[64 \times 2048]$ will obtain the best results while the input dimension of $[128 \times 2048]$ will obtain the worst results. The reason why the effect deteriorates when the duration value increases to 128 but the energy remains unchanged at 2048 is that the matrix becomes sparse, that is, it does not contain more information about the stacking clusters. For example, when a stacking cluster of length 128 appears, its energy value is likely to exceed 2048, so it is not recorded. This also explains why when the input dimension increases to $[128 \times 4096]$, the estimation effect is better than that of $[128 \times 2048]$. We did not show the result of $[64 \times 4096]$ because the probability of the element appearing in high-energy photons is extremely low, the stacking clusters with a length of 64 must all be high-energy particle events, and the main peaks of the elements also appear in the low-energy region.

A smaller matrix size can avoid the problem of information sparsity, which means that the model can capture more relevant features without being disturbed by redundant information, while larger matrix sizes may lead to over-fitting, especially when there is insufficient training data. Smaller matrices force the model to focus on more essential features by limiting the input dimension, which helps improve its generalization ability on the validation set. Furthermore, a larger matrix will require many more samples to provide a histogram that is statistically relevant, challenging the signal preprocessing. Increasing the input matrix dimension will sharply increase the model training cost with more training time and occupy larger memory of GPUs; however, it gives worse results. Overall, $[64 \times 2048]$ is proved to be the optimal training input size in our experiments.

To better illustrate the pile-up spectrum recovery performance of different methods, we plot the predicted spectrum of different methods in Figure 7, where Figure 7a shows the spectrum estimated by the traditional methods, and Figure 7b shows the predicted spectrum by our method. The result shows that the spectrum predicted by the pile-up correction algorithm has a smaller peak error; however, it fails at estimating low-energy

region area. Our proposed DL algorithm, focuses on these two flaws, both low-energy peaks and high-energy peaks have less prediction error.

Table 5. The model estimation results were obtained by using the Energy–Duration matrix of different scales as the input training data of the neural network, using KL and MSE as evaluation indicators. The experiments were carried out under different count rates from 0.055 to 0.255, $\mathcal{P}$ from 0.674 to 0.994, and different sources: ^{51}Cr, ^{169}Er, and ^{125}I. The smallest Energy–Duration matrix has the best performance.

Energy–Duration	Activity	Single Sources					
		^{51}Cr		^{169}Er		^{125}I	
	λ_n	KL	MSE	KL	MSE	KL	MSE
2048×64	0.055	0.03	0.22	0.03	0.18	0.02	0.03
	0.135	0.14	34.18	0.04	0.50	0.02	0.11
	0.255	0.50	243.71	0.18	9.42	0.03	0.11
2048×128	0.055	0.11	1.47	0.08	1.08	0.08	7.97
	0.135	0.07	0.74	0.08	0.55	0.06	2.13
	0.255	0.81	607.47	0.26	11.52	0.04	0.26
4096×128	0.055	0.03	0.66	0.02	0.08	0.03	0.87
	0.135	0.04	1.71	0.04	0.53	0.02	0.10
	0.255	0.66	538.34	0.20	24.70	0.02	0.11
4096×256	0.055	0.07	0.80	0.03	0.04	0.05	0.13
	0.135	0.07	1.99	0.05	0.14	0.03	0.14
	0.255	0.80	604.16	0.23	14.45	0.03	0.16

The MSE metric is in units of 10^{-7}.

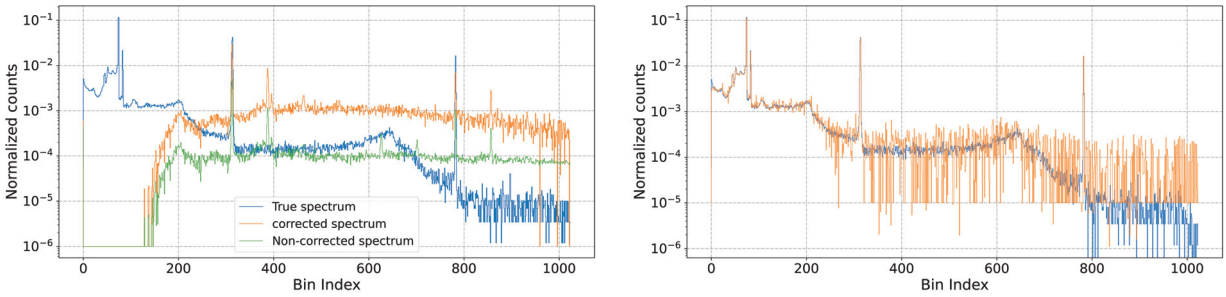


Figure 7. Comparison of spectra estimated by different traditional methods and our DL method shows that our approach provides more accurate estimations. The provided conditions are of $\lambda = 0.095, DC = 1.94, \mathcal{P} = 0.856,$ ^{57}Co,^{22}Na. (**a**) The blue curve is the true spectrum, the green curve is Method 1's result, and the orange curve represents Method 2's result. (**b**) The blue curve is the true spectrum, the orange curve is our proposed method's result.

6. Conclusions

In this study, we proposed a DL-based algorithm for pile-up correction in spectrometric data under high-count-rate measurements. Our approach leverages the power of a 2D-UNet architecture, combined with a count rate embedding module, utilizing both spatial and temporal features to effectively recover the energy spectrum distorted by pile-up effects. The method utilizes attention gates to focus on the relevant portions of the input distorted Energy–Duration matrix, significantly improving the accuracy of energy recovery, even under challenging high-count-rate conditions.

Through comprehensive experimentation, we demonstrated that our method outperforms traditional pile-up correction techniques in both low and high-count-rate scenarios. Specifically, our model consistently showed lower KL divergence and MSE, indicating superior recovery of the energy spectrum.

Furthermore, our model exhibits strong robustness and generalization performance across varying unseen count rate levels, with minimal performance degradation as the pile-

up degree increases. This indicates the adaptability of our method to different measurement conditions, making it a reliable tool for spectrometric applications in fields such as nuclear physics, environmental monitoring, and medical diagnostics.

The proposed deep learning-based pile-up correction method offers a novel and effective solution to the challenges posed by high-count-rate measurements. It significantly improves the accuracy and reliability of energy spectrum recovery, with strong potential for further enhancement and deployment in real-world spectrometric systems. In future work, we plan to extend this work by integrating a pre-trained count rate estimator neural network into the model, which can predict the count rate directly from raw signals, eliminating the reliance on known count rate information. To further address the real-world detector's nonlinear response as well as the dead time effect, a more comprehensive approach is needed to account for the challenges; our future work will extend the current model as well as refine the choice of simulation parameters to incorporate these factors.

Author Contributions: Conceptualization, Y.H.; Methodology, Y.H.; Software, Y.H.; Investigation, Z.C.; Writing—original draft, Y.H.; Supervision, X.Z., Y.Z., T.T. and D.B.; Funding acquisition, Y.Z. 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 numbers 12373113 and 12475196.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are restricted due to ongoing intellectual property development.

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

References

1. Knoll, G.F. *Radiation Detection and Measurement*, 4th ed.; Wiley: Hoboken, NJ, USA, 2010.
2. Siddavatam, A.P.I.; Vaidya, P.P.; Nair, J.M. Pileup rejection using estimation technique for high resolution nuclear pulse spectroscopy. In Proceedings of the 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, Kereala, 6–7 July 2017; pp. 531–535. [CrossRef]
3. Lee, M.; Lee, D.; Ko, E.; Park, K.; Kim, J.; Ko, K.; Sharma, M.; Cho, G. Pulse pileup correction method for gamma-ray spectroscopy in high radiation fields. *Nucl. Eng. Technol.* **2020**, *52*, 1029–1035. [CrossRef]
4. Liu, B.; Liu, M.; He, M.; Ma, Y.; Tuo, X. Model-Based Pileup Events Correction via Kalman-Filter Tunnels. *IEEE Trans. Nucl. Sci.* **2019**, *66*, 528–535. [CrossRef]
5. Bolic, M.; Drndarevic, V. Processing Architecture for High Count Rate Spectrometry with NaI(Tl) Detector. In Proceedings of the 2008 IEEE Instrumentation and Measurement Technology Conference, Victoria, BC, Canada, 12–15 May 2008; pp. 274–278. [CrossRef]
6. Scoullar, P.A.B.; Evans, R.J. Maximum likelihood estimation techniques for high rate, high throughput digital pulse processing. In Proceedings of the 2008 IEEE Nuclear Science Symposium Conference Record, Dresden, Germany, 19–25 October 2008; pp. 1668–1672. [CrossRef]
7. Trigano, T.; Gildin, I.; Sepulcre, Y. Pileup Correction Algorithm using an Iterated Sparse Reconstruction Method. *IEEE Signal Process. Lett.* **2015**, *22*, 1392–1395. [CrossRef]
8. Trigano, T.; Souloumiac, A.; Montagu, T.; Roueff, F.; Moulines, E. Statistical Pileup Correction Method for HPGe Detectors. *IEEE Trans. Signal Process.* **2007**, *55*, 4871–4881. [CrossRef]
9. Mclean, C.; Pauley, M.; Manton, J.H. Non-Parametric Decompounding of Pulse Pile-Up Under Gaussian Noise with Finite Data Sets. *IEEE Trans. Signal Process.* **2020**, *68*, 2114–2127. [CrossRef]
10. Ilhe, P.; Moulines, É.; Roueff, F.; Souloumiac, A. Nonparametric estimation of mark's distribution of an exponential shot-noise process. *Electron. J. Stat.* **2015**, *9*, 3098–3123. [CrossRef]
11. Khatiwada, A.; Klasky, M.; Lombardi, M.; Matheny, J.; Mohan, A. Machine Learning technique for isotopic determination of radioisotopes using HPGe γ-ray spectra. *Nucl. Instrum. Methods A* **2023**, *1054*, 168409. [CrossRef]
12. Kamuda, M.; Stinnett, J.; Sullivan, C.J. Automated Isotope Identification Algorithm Using Artificial Neural Networks. *IEEE Trans. Nucl. Sci.* **2017**, *64*, 1858–1864. [CrossRef]
13. Luo, R.; Popp, J.; Bocklitz, T. Deep Learning for Raman Spectroscopy: A Review. *Analytica* **2022**, *3*, 287–301. [CrossRef]

14. Jeon, B.; Lim, S.; Lee, E.; Hwang, Y.S.; Chung, K.J.; Moon, M. Deep Learning-Based Pulse Height Estimation for Separation of Pile-Up Pulses From NaI(Tl) Detector. *IEEE Trans. Nucl. Sci.* **2022**, *69*, 1344–1351. [CrossRef]
15. Morad, I.; Ghelman, M.; Ginzburg, D.; Osovizky, A.; Shlezinger, N. Model-Based Deep Learning Algorithm for Detection and Classification at High Event Rates. *IEEE Trans. Nucl. Sci.* **2024**, *71*, 970–980. [CrossRef]
16. Kamuda, M.; Zhao, J.; Huff, K. A comparison of machine learning methods for automated gamma-ray spectroscopy. *Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.* **2020**, *954*, 161385. [CrossRef]
17. Trigano, T.; Bykhovsky, D. Deep Learning-Based Method for Activity Estimation From Short-Duration Gamma Spectroscopy Recordings. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 6504811. [CrossRef]
18. Regadío, A.; Esteban, L.; Sánchez-Prieto, S. Unfolding using deep learning and its application on pulse height analysis and pile-up management. *Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.* **2021**, *1005*, 165403. [CrossRef]
19. Chen, Z.; Bykhovsky, D.; Zheng, X.; Trigano, T.; Zhu, Y. GaSim: A python class to generate simulated time signals for gamma spectroscopy. *SoftwareX* **2025**, *29*, 102037. [CrossRef]
20. Bykhovsky, D.; Chen, Z.; Huang, Y.; Zheng, X.; Trigano, T. Advanced Spectroscopy Time-Domain Signal Simulator for the Development of Machine and Deep Learning Algorithms. *IEEE Sens. Lett.* **2025**, 1–4. [CrossRef]
21. Trigano, T.; Barat, E.; Dautremer, T.; Montagu, T. Fast Digital Filtering of Spectrometric Data for Pile-up Correction. *IEEE Signal Process. Lett.* **2015**, *22*, 973–977. [CrossRef]
22. Ortec. Gamma-Ray Spectroscopy Using NaI(Tl). Available online: https://www.ortec-online.com.cn/-/media/ametekortec/third-edition-experiments/3-gamma-ray-spectroscopy-using-nai-tl.png?la=zh-cn&revision=001dbc1d-9559-49c0-b57d-5567e28d1b96&hash=0D4E147F8F47A4485C42D55776D9E075 (accessed on 16 November 2010).
23. Casella, G.; Robert, C.P. *Monte-Carlo Statistical Methods*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2004.
24. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**, arXiv:1804.03999.
25. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597.

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.

sensors

MDPI

Article

Deep Learning Model Compression and Hardware Acceleration for High-Performance Foreign Material Detection on Poultry Meat Using NIR Hyperspectral Imaging

Zirak Khan [1], Seung-Chul Yoon [2,*] and Suchendra M. Bhandarkar [1]

[1] School of Computing, University of Georgia, Athens, GA 30602, USA; zirak.khan@uga.edu (Z.K.); suchi@uga.edu (S.M.B.)
[2] Quality and Safety Research Unit, U.S. National Poultry Research Center, U.S. Department of Agriculture—Agricultural Research Service, Athens, GA 30605, USA
* Correspondence: seungchul.yoon@usda.gov

Abstract: Ensuring the safety and quality of poultry products requires efficient detection and removal of foreign materials during processing. Hyperspectral imaging (HSI) offers a non-invasive mechanism to capture detailed spatial and spectral information, enabling the discrimination of different types of contaminants from poultry muscle and non-muscle external tissues. When integrated with advanced deep learning (DL) models, HSI systems can achieve high accuracy in detecting foreign materials. However, the high dimensionality of HSI data, the computational complexity of DL models, and the high-paced nature of poultry processing environments pose challenges for real-time implementation in industrial settings, where the speed of imaging and decision-making is critical. In this study, we address these challenges by optimizing DL inference for HSI-based foreign material detection through a combination of post-training quantization and hardware acceleration techniques. We leveraged hardware acceleration utilizing the TensorRT module for NVIDIA GPU to enhance inference speed. Additionally, we applied half-precision (called FP16) post-training quantization to reduce the precision of model parameters, decreasing memory usage and computational requirements without any loss in model accuracy. We conducted simulations using two hypothetical hyperspectral line-scan cameras to evaluate the feasibility of real-time detection in industrial conditions. The simulation results demonstrated that our optimized models could achieve inference times compatible with the line speeds of poultry processing lines between 140 and 250 birds per minute, indicating the potential for real-time deployment. Specifically, the proposed inference method, optimized through hardware acceleration and model compression, achieved reductions in inference time of up to five times compared to unoptimized, traditional GPU-based inference. In addition, it resulted in a 50% decrease in model size while maintaining high detection accuracy that was also comparable to the original model. Our findings suggest that the integration of post-training quantization and hardware acceleration is an effective strategy for overcoming the computational bottlenecks associated with DL inference on HSI data.

Keywords: hyperspectral imaging; deep learning; inference optimization; post-training quantization; hardware acceleration; foreign material detection; poultry processing; industrial applications; real-time processing; sensor technology

1. Introduction

Ensuring the safety and quality of poultry products is of high importance in the poultry industry, as contamination with foreign materials (FMs) poses severe health risks

to consumers and can lead to high economic losses to the industry due to product recalls [1]. Non-biological FMs, such as rubber, plastic, and metal fragments, are common contaminants introduced inadvertently during the deboning, trimming, portioning, and packaging processes in poultry processing plants [2,3]. The efficient detection and removal of these FMs are critical to maintaining food safety standards and complying with regulatory requirements.

While many traditional detection techniques, including X-ray imaging, metal detection, thermal imaging, spectroscopy, and conventional color imaging, have been employed to identify FMs in food products [4], each presents inherent limitations: X-ray imaging, though highly accurate for dense objects, struggles with low-density materials like plastic and rubber; thermal imaging is sensitive to environmental fluctuations; terahertz spectroscopy is hindered by the high water content in poultry products; and color imaging often fails to distinguish contaminants from meat due to similar appearances [5–8]. Hyperspectral imaging (HSI) has emerged as a powerful non-destructive sensing technology that combines imaging and spectroscopy to capture detailed spatial and spectral information from materials [9]. When integrated with advanced deep learning (DL) techniques, such as convolutional neural network (CNN) and generative adversarial network (GAN), these HSI-based DL solutions further enhanced the detection capabilities in the remote sensing and food safety sectors [10,11]. This fusion allows for the extraction of deep features from high-dimensional HSI data, leading to improved accuracy in classification and detection tasks. For instance, Chung et al. [12] demonstrated that a fusion model combining two HSI modalities—visible near-infrared (VNIR) HSI and short-wave infrared (SWIR) HSI—achieved superior precision and recall in detecting FMs on broiler meat compared to using standalone modalities. Similarly, Seo et al. [13] showed that VNIR HSI coupled with a 1D CNN model outperformed traditional support vector machine (SVM) methods in detecting organic residues on food machinery surfaces. Additionally, Yang et al. [14] highlighted the effectiveness of CNNs and sparse autoencoders in identifying mold in stored maize kernels using HSI data. Furthermore, Campos et al. [10] presented a semi-supervised DL approach leveraging near-infrared (NIR) HSI and GANs to detect FMs on poultry meat. These studies underscore that the integration of HSI and DL techniques presents a more robust approach for enhancing foreign material (FM) detection in food processing, offering great improvements over traditional methods.

However, modern poultry processing lines operate at high speeds, processing between 140 (a typical speed in the U.S.) and 250 birds per minute (BPM) [15], which presents deployment challenges for HSI-based DL models in such high-throughput environments. This is mainly due to the high-dimensional nature of HSI data, the over-parameterized DL models, and the resource-constrained environment. The high volume of input HSI data for these advanced DL models increases the required intermediate activation memory and computation demands during inference. Similarly, the large number of pre-trained models' weights also increases the required memory footprint and computational needs [16]. To avoid such memory and computation bottlenecks in high-throughput and resource-constrained environments like food processing lines, efficient DL- and HSI-based solutions are needed. In recent years, several studies have been dedicated to improving DL model efficiency in the computer vision domain. Model compression and hardware acceleration techniques provide efficient DL models without compromising accuracy.

Model compression aims to reduce the size and computational demands of DL models, facilitating faster inference and lower memory usage [17–21]. Two primary model compression methods are pruning and quantization [22,23]. Pruning involves reducing the size of pre-trained DL models by eliminating redundant or less significant weights and activations while maintaining accuracy [24]. Pruning can be both fine-grained and coarse-grained.

Both approaches have certain merits and drawbacks. The fine-grained approach allows for larger compression ratios but is challenging to accelerate hardware due to the complexity of sparse matrix multiplication. In contrast, the coarse-grained approach offers moderate compression rates but is easier to accelerate because it is similar to running a smaller model [25,26]. Moreover, pruning typically requires retraining or fine-tuning the pruned model to recover any accuracy loss [25]. The size of a DL model depends on the number of weights and the bit width, also known as precision in graphics processing unit (GPU) computing, required to store each weight. Quantization involves reducing the bit width to store each weight. Because low-bit operations not only reduce memory footprint but are also cheaper both from arithmetic operations and data movement perspectives [27–29]. Some commonly supported low-bit data types by different hardware architectures are 32-bit floating-point (FP32), 16-bit floating-point (FP16), and 8-bit integer (INT8). Quantization can be applied during training, allowing the model to adjust weights to reduced precision, called Quantization Aware Training (QAT). QAT generally achieves better accuracy with low bit widths but requires access to the original data and is best-suitable when training a model from scratch [30]. On the other hand, post-training quantization (PTQ) applies quantization to a pre-trained DL model without the need for retraining [31]. This process primarily involves mapping floating-point weights and activations into discrete levels using scaling factors (S) and zero points (Z). PTQ can quantize an entire tensor with a single scale (per-tensor) or assign unique scales per channel (per-channel), creating a balance between computational overhead and accuracy [28,32]. More advanced techniques, such as group quantization, refine the scaling factors at an intermediate level, optimizing both memory usage and performance [33]. Additionally, a small calibration set is often used to determine the dynamic range of activations (rmin, rmax), which helps reduce quantization errors near the centers of distribution [34–36]. However, when quantizing weights and activations directly from FP32 to FP16, it is not typically necessary to calculate optimal quantization parameters like zero points and scales using a calibration dataset, making this process more convenient [35,36]. Calibration is mainly required when quantizing activations and weights to integer formats, such as INT8. In this study, we adopted FP16 PTQ to eliminate the need for retraining and avoid the requirement for a calibration dataset.

Hardware acceleration exploits specialized processors to enhance the computational efficiency of DL models. GPUs offer high parallelism, making them well-suited for the matrix operations prevalent in DL inference. Field-programmable gate arrays (FPGAs) allow for custom hardware configurations, enabling tailored acceleration for specific models or operations. In this study, we accelerated the HSI-based DL model on a GPU that is based on Ampere Architecture supporting only FP16 and INT8 quantization.

Although the effectiveness of the aforementioned optimization techniques is recognized in various fields, their use in HSI-based DL models remains underexplored. This is particularly evident when it comes to achieving real-time inference with HSI-based DL models for FM detection in poultry meat. Research was needed to address this challenge. In this study, we explored PTQ and hardware acceleration to optimize a semi-supervised FM detection DL model with two pre-trained GAN discriminators, for high-performance FM detection using HSI data. Our contributions are as follows: (1) Hardware acceleration with TensorRT: We developed two inference engines for the two discriminators utilizing NVIDIA's TensorRT to accelerate model inference, achieving a fivefold reduction in inference time compared to unoptimized GPU-based inference and exceeding the inference-time requirements of poultry processing. (2) Model compression using PTQ on TensorRT without fine-tuning: We applied FP16 PTQ to quantize the weights and activations of the same pre-trained semi-supervised FM detection DL model with two GAN discriminators. This process did not require calibration data and any fine-tuning, while still maintaining the

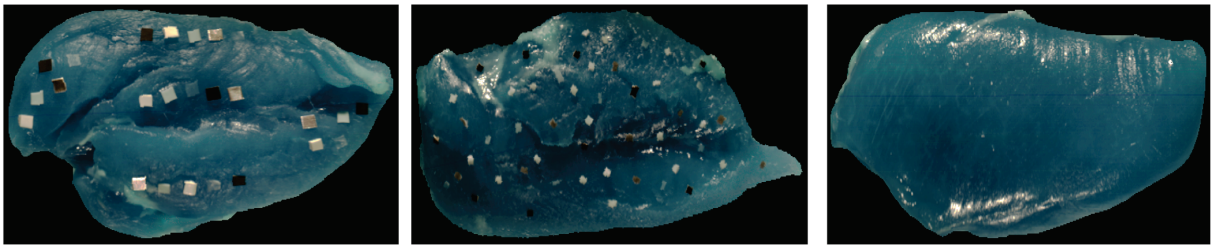
accuracy of the pre-trained model. As a result, we achieved a 50% reduction in model size. (3) We conducted simulations using two hypothetical hyperspectral line-scan cameras to evaluate the feasibility of deploying an optimized HSI-based DL model for the real-time detection of FMs.

2. Materials and Methods

2.1. Dataset and Imaging System

The dataset employed in this study consisted of hyperspectral images captured from 12 boneless, skinless broiler breast fillets (pectoralis major muscle) using a push-broom line-scanning SWIR HSI system (Hyperspec® SWIR, Headwall Photonics, Fitchburg, MA, USA) with a wavelength range from 1000 to 2500 nm. The HSI system was equipped with a mercury–cadmium–telluride (MCT) camera and a spectrograph and operated via custom C++ software with tungsten halogen illuminators. Intensity calibration and denoising techniques were applied to optimize image quality for further analysis, with additional procedural details also available in our previous work [10,12].

Utilizing the SWIR HSI system, each fillet was imaged under two conditions: clean without FMs first and contaminated with FMs for later imaging. The contamination involved 30 distinct types of FMs commonly found in poultry processing environments, such as polymers, metals, and wood, detailed in the prior work [10,12]. The target sizes of FMs were approximately 2 × 2 mm (small) and 5 × 5 mm (large). Each fillet for imaging was placed on a blue food-grade polyurethane conveyor belt background. In total, 48 SWIR hyperspectral images were collected: 24 clean and 24 contaminated (12 with each size of FM). These SWIR hyperspectral images (1000–2500 nm) were subsequently transformed into NIR hyperspectral images with 95 spectral bands from 1000 to 1700 nm. Figure 1 shows three exemplar RGB pseudo-color images derived from clean and contaminated NIR hyperspectral images utilized in this study.



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

Figure 1. RGB pseudo-color images derived from NIR hyperspectral images with (**a**) 5 × 5 mm FMs; (**b**) 2 × 2 mm FMs; (**c**) no FMs.

These NIR hyperspectral images were employed in this study to evaluate the performance of a pre-trained semi-supervised DL model with two GAN discriminators for FM detection on the surface of poultry meat [10]. This DL model was originally trained on part of the dataset; however, in this study, all images in the dataset were used to compare and evaluate inference performance—such as accuracy, inference time, and throughput—among different models, including the original DL model on CPU and GPU platforms and the optimized DL model through the proposed PTQ and hardware acceleration techniques. This analysis aimed to assess the effects of optimization techniques on model performance in real-world scenarios, specifically simulating the typical imaging conditions for modern poultry processing lines.

2.2. Inference Pipeline for FM Detection

The inference pipeline for FM detection includes several key components: pre-processing steps applied to the input hyperspectral data, an inference model based on a semi-supervised DL model with two GAN discriminators [10], an inference engine comprising two sub-engines that executes inputs using the model, and post-processing steps to generate the prediction output. The pipeline and data flow are illustrated in Figure 2. Pre-processing steps involve calibrating raw hyperspectral data to relative reflectance values and reshaping the calibrated hyperspectral image (a 3D data cube) into a 2D array of spectral vectors. The DL model with two GAN modules was uniquely designed to address the spectral heterogeneity found in two major tissue types of skinless poultry meat, specifically in distinguishing the hyperspectral signatures of muscle tissue and non-muscle external tissues, such as external fat and connective tissues, observed on the surface. The model operates on the hypothesis that training with clean chicken muscle and non-muscle external tissues' spectral responses. Thus, two GANs are sufficient for the model to recognize typical spectral characteristics of normal chicken breast meat. Once the model is trained, any anomalous spectral signatures lead to significantly higher prediction errors during inference, indicating the presence of FMs. The two GAN discriminators in the pre-trained inference model were frozen and saved as HDF5 files. These HDF5 files were then used for unoptimized inference engines, such as Tensorflow-CPU and Tensorflow-GPU, and optimized inference engines with TensorRT GPU. The saved two GAN discriminator files were converted to ONNX format for compatibility with TensorRT. Given a hyperspectral image, both inference sub-engines received the same pre-processed spectral vectors for inferencing and outputting the probabilities of FM presence at each pixel independently. Post-processing steps include thresholding to generate binary output vectors, reshaping the output binary vectors to 2D images representing the original spatial resolution, median filtering for denoising, creating individual detection maps for each engine, and combining the individual detection maps with a data fusion rule for the final pixel-level FM detection map. The post-processing steps were performed offline on the CPU.

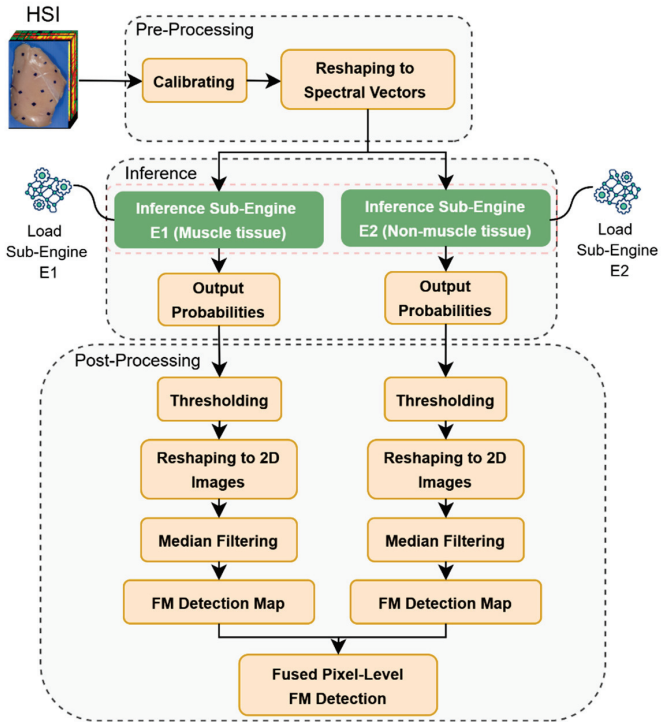


Figure 2. Inference pipeline for FM detection.

2.2.1. Inference Model

As mentioned above, the inference model employs two GAN discriminators, designed to process distinct spectral responses associated with muscle and non-muscle external tissues in chicken breast fillets separately. This specialized approach enhances the precision of FM detection within hyperspectral data. Each GAN module for training consists of two competing components: the generator and the discriminator.

- **Generator:** The generator includes several 1D convolutional layers activated by ReLu, leading to a sigmoid activation layer at the end. The generators in two GAN models create synthetic spectral data that closely resemble clean (normal) hyperspectral data for each tissue type, respectively. The generator was only used during model training in our case.
- **Discriminator:** The discriminator adopts a 1D U-Net architecture, consisting of an encoder and a decoder, to process spectral vectors at individual pixel locations (x,y). Before being sent to two GAN modules, the input spectral vectors are initially subjected to clustering into two groups using a 1D Gaussian Mixture Model (GMM). This clustering segregates the spectral data into distinct clusters representing muscle and non-muscle external tissues, which are then processed by their respective discriminators for further analysis. Note that this GMM module for spectral clustering is only necessary during the model training phase. During the inference, only the encoder part of the two discriminators is utilized. Therefore, the inference model consists of two pre-trained GAN discriminator encoders. The overall architecture of the discriminator is shown in Figure 3. The discriminator outputs probabilities that are thresholded and further processed for binary classification at each pixel, indicating the presence (1) or absence (0) of FM, as previously described. For the GAN discriminators, a dummy band was added to each spectrum, resulting in a final spectrum size of 96.

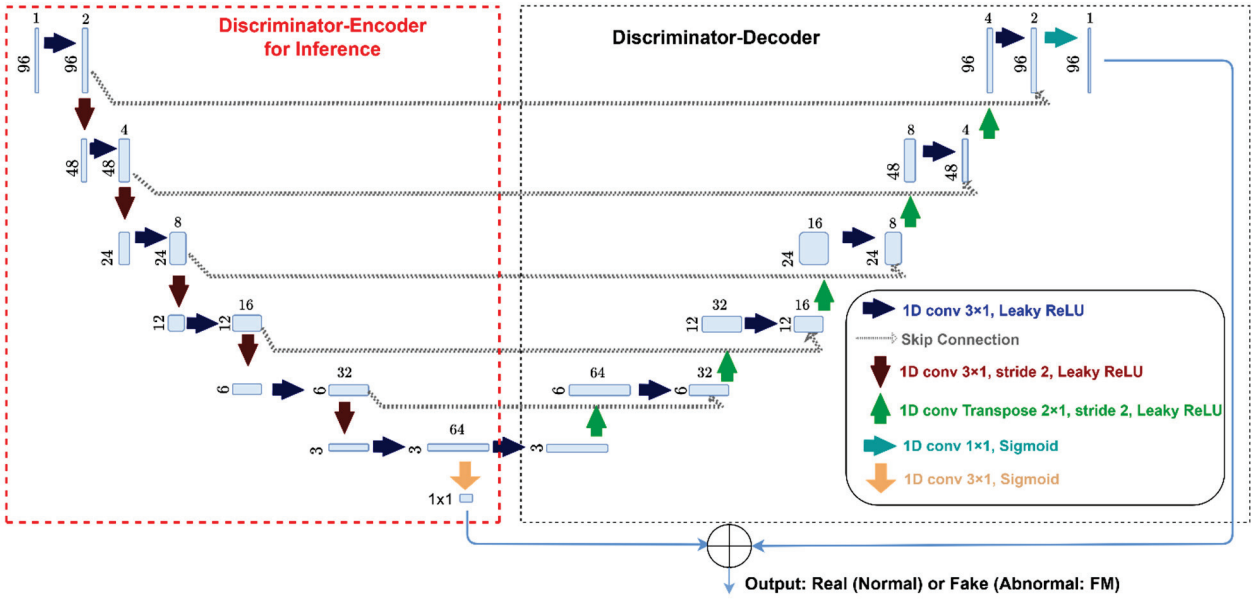


Figure 3. Architecture of the GAN discriminator encoder (shown in a red dotted box) for inference, as part of the overall discriminator, which includes both an encoder and a decoder used during training. Two GAN discriminators are trained to effectively differentiate between the spectral signatures of muscle and non-muscle external tissues.

2.2.2. Creation of Pre-Trained Inference Model

The process of creating a pre-trained inference model [10] for the study is briefly described next. The DL model with two GAN modules was trained using approximately

850,000 spectral responses from a subset of 12 clean hyperspectral images in the dataset. This training phase focused on enabling the two GAN models to discern between normal (muscle and non-muscle) and anomalous spectral patterns of any FMs effectively. Following training, the inference model with two GAN discriminators was tested using approximately 2.6 million spectral responses, as independent test samples. This testing phase assessed the model's accuracy. The model was developed using TensorFlow 2.12.0 and Python 3.10.10. The trained two GAN discriminators were saved as HDF5 files.

2.3. Optimizatized Inference

The optimization process involved hardware acceleration and model compression (PTQ) through NVIDIA's TensorRT, a high-performance DL framework designed to improve model inference on NVIDIA GPUs. The inference model was optimized using hardware acceleration and PTQ to improve performance for real-time HSI operations.

2.3.1. Hardware Acceleration

The optimization process began by freezing the pre-trained inference model and converting its format from TensorFlow to ONNX suitable for NVIDIA's TensorRT. The DL inference optimization framework within TensorRT is specifically designed to enhance performance on NVIDIA GPUs by creating GPU-optimized inference engines. The models benefit from various features, including dynamic tensor memory and multi-stream execution. Dynamic tensor memory efficiently manages memory allocation, reducing energy consumption and minimizing operational overhead by allocating memory to tensors only when needed. In contrast, multi-stream execution facilitates the parallel processing of multiple data inputs, significantly enhancing performance by delivering low latency and high throughput.

The construction of a TensorRT inference engine involved several important steps. These steps include creating a network definition, importing the inference model with two GAN discriminators using the ONNX parser, and creating a TensorRT engine with two sub-engines, one for each discriminator, utilizing a builder mechanism. The process of inference in a TensorRT runtime consists of six key steps: (1) initiating an inference execution context; (2) allocating memory on the CUDA device for both input and output data; (3) transferring input data from the host to the allocated input memory on the CUDA device; (4) executing inference through the TensorRT engines using an API that supports the asynchronous execute for improved performance; (5) moving the output data back to host memory; (6) synchronizing the stream used for data transfers and inference execution to ensure the completion of all operations.

In our initial hardware acceleration setup via TensorRT, the model remains in FP32 precision. This primarily engages CUDA cores rather than leveraging half-precision Tensor cores. Although we achieve speedups over CPU and unoptimized GPU inference, due to optimized kernel fusion, dynamic memory allocation, and multi-stream parallelism, the full benefits of hardware acceleration are only realized after compressing the model to FP16.

2.3.2. Model Compression

As a model compression strategy, the inference model with two pre-trained GAN discriminators underwent PTQ as its first step. This process reduced the precision of the model's weights and activations from FP32 to FP16. The goal of this reduction was to decrease computational demands and enhance memory efficiency while maintaining model accuracy without any fine-tuning. As a result of PTQ, the model size became smaller. After quantization, the inference model with two GAN discriminators was accelerated using TensorRT.

2.3.3. Computational Complexity

Each discriminator model for inference consists of 1D convolutional layers that progressively increase the number of channels (2, 4, 8, 16, 32) with stride-2 down-sampling in selected layers. The inference model comprises approximately 6359 parameters and 36,000 multiply-accumulate (MAC) operations per forward pass (see Section 2.2.1 for detailed layer specifications). When quantizing a model from FP32 to FP16, the number of MAC operations in the inference model stays the same. However, the arithmetic cost per MAC is reduced because fewer bits are used. In addition, the storage required for parameters is effectively halved, decreasing from approximately 24.8 KB to around 12.4 KB. This reduction helps us to alleviate memory bandwidth demands.

Once quantized, Tensor cores on NVIDIA GPUs process FP16 data at higher throughput than FP32 computations using CUDA cores, often achieving 2 to 4 times the theoretical speedups in terms of floating-point operations per second (FLOPS) [37,38]. By reducing memory usage, decreasing the bit-width per operation, and utilizing specialized hardware blocks (Tensor cores) for half-precision calculations, FP16 PTQ allows for faster end-to-end inference without the need for model retraining. This combination of model compression and hardware acceleration is central to achieving real-time speeds in our hyperspectral detection pipeline.

2.4. Experimental Setup and Conditions

2.4.1. Experimental Setup

The experimental setup for evaluating the performance of unoptimized and optimized inference models was simulated by replicating real-world, real-time HSI conditions for FM detection in poultry processing. The dataset used in this study was derived from a comprehensive dataset of approximately 3.4 million spectral responses (96 bands per spectrum in the range of 1000–1700 nm) obtained from 48 hyperspectral images used in the previous study [10]. This dataset was sampled or expanded to create the necessary datasets for various experimental conditions, requiring only a set of input spectral vectors as input and a set of output vectors.

For the simulation, two hypothetical hyperspectral cameras were modeled after two commercial products [39]: an SWIR hyperspectral camera (Hyperspec® SWIR, Headwall Photonics, Fitchburg, MA, USA) and a NIR hyperspectral camera (Specim FX17, Specim, Oulu, Finland). The SWIR camera has an MCT detector with 320 spatial pixels × 256 spectral bands in the wavelength range of 1000–2500 nm, while the NIR camera has an InGaAs detector with 640 spatial pixels × 224 spectral bands in the wavelength range of 900–1700 nm. Halved broiler breast fillets vary in size, measuring up to 20.33 cm (length) × 12.47 cm (width) in published studies [40,41]. In this study, we hypothesized that the fillet dimensions would be 9 inches (22.86 cm) in length and 5 inches (12.7 cm) in width, representing an upper bound. Two different imaging conditions were established: one for the SWIR camera where the longitudinal direction of a fillet was perpendicular to the direction of the image scan line, and another for the NIR camera where it was parallel to the image scan line. The NIR camera was assumed to produce a hyperspectral image of 640 pixels (image width, pixel resolution of 0.4 mm) by 168 scanlines (image height, pixel resolution of 0.8 mm), while the SWIR camera was assumed to produce a hyperspectral image of 320 pixels (image width, pixel resolution of 0.4 mm) by 800 scanlines (image height, pixel resolution of 0.3 mm). It was assumed that both cameras would produce 96 wavebands including a dummy band ranging from 1000 to 1700 nm to align with the input dimensions required for the pre-trained inference model. Therefore, the dimensions of the hypercubes studied were 320 width × 800 height × 96 bands for the SWIR camera (256,000 spectra) and 640 width × 168 height × 96 bands for the NIR camera (107,520 spectra). Notably, the

maximum frame rates of the two cameras were 45 frames per second (FPS) and 670 FPS, corresponding to lines per second in hyperspectral imaging.

2.4.2. Experimental Conditions

Key experimental conditions included four inference engines, each with two sub-engines, E1 and E2, corresponding to the two GAN discriminators D1 and D2, varying batch sizes, and diverse simulation setups. Four different inference engines—TensorFlow-CPU, TensorFlow-GPU, TensorRT-FP32, and TensorRT-FP16—were compared. A batch size refers to the number of input spectral vectors processed simultaneously during the inference runtime. As aforementioned, two hypothetical hyperspectral cameras were used in the study.

Inference Engines: Four different inference engines, each with two sub-engines, were established to measure the variations in model performance under different simulation conditions. This study compares the performance of these inference engines to find which engine performs best in different scenarios.

1. **TensorFlow-CPU:** This inference engine (two sub-engines) loads and executes the pre-trained inference model (two GAN discriminators) on the CPU without further optimizations using the TensorFlow CPU version. This inference engine provides a baseline benchmark for performance under highly resource-constrained conditions, such as an embedded industrial computer. It serves as a reference point for subsequent optimizations. The model's structural or computational specifics did not change.
2. **TensorFlow-GPU:** The inference model was loaded and executed on a single GPU board. This inference engine was utilized to directly evaluate the enhancements achievable through GPU acceleration alone, using the TensorFlow GPU version without further optimizations or modifications similar to TensorFlow-CPU.
3. **TensorRT-GPU (FP32):** The two discriminators for inference, originally as files in the HDF5 format, were converted to the ONNX format to ensure compatibility with NVIDIA's TensorRT. The ONNX files were then loaded and executed on the GPU through TensorRT. This inference engine took full advantage of TensorRT's capabilities to enhance the execution efficiency of the models while maintaining their original FP32 precision for weights and activations. Thus, the focus was on improving performance through hardware acceleration.
4. **TensorRT-GPU-PTQ (FP16):** This inference engine is essentially the same as TensorRT-GPU except for the added PTQ with TensorRT optimization. Initially, the model's weights and activations were quantized from FP32 to FP16 (half-precision), aiming to reduce computational demands and enhance memory efficiency. Subsequently, these quantized models were accelerated using TensorRT to maximize the computational performance, assessing the overall impact of both quantization and advanced hardware acceleration on the models' effectiveness.

Batch Size: A batch size in the number of spectral vectors (i.e., pixels) cannot be generalized for all scenarios, as different hyperspectral cameras and imaging conditions can result in varying wavelengths per pixel and thus require different memory sizes. For example, a batch size of one spectrum with 100 wavebands requires the same amount of memory as a batch size of 10 spectra with 10 wavebands per spectrum. However, batch sizes differ in these two cases (1 versus 10). To standardize the batch size and aid the understanding of readers, we also provide a batch size in memory size, measured in bytes, such as kilobytes (KB), megabytes (MB), and gigabytes (GB). The batch sizes in bytes were first designed to increase approximately by powers of 10, specifically: 1 KB, 10 KB, 100 KB, 1 MB, 10 MB, 100 MB, and 1 GB. The batch size in the number of pixels was then obtained by calculating the nearest integer using the formula, (batch size in KB)/(96 $\times$ 4 bytes). For

example, the batch size of 1 KB contained 3 spectral vectors with 96 bands of 32-bit floating numbers. Note that the number of spectral vectors was provided to an inference engine as a batch size.

Static and Dynamic Batch Size: When setting the input shape in TensorRT, a batch size affects performance. Static and dynamic batch sizes are compared in this study. A static batch size refers to a fixed one that does not change at runtime, whereas a dynamic batch size is a variable batch size that can change at runtime, providing flexibility to handle varying batch sizes. The static batch sizes used are 1 KB, 10 KB, 100 KB, 1 MB, 10 MB, 100 MB, and 1 GB. The dynamic batch size was defined using three parameters: minimum, optimal, and maximum values. An inference engine configured with a dynamic batch size works for all batch sizes within the [minimum, maximum] range and is optimized for performance at an optimal size within the range. The minimum size ensures responsiveness and quick processing for smaller data inputs, the optimal size provides the best performance in throughput and resource utilization, and the maximum sets the engine's upper limit to manage the largest expected data volumes efficiently. In the study, the static batch sizes were set as the optimal sizes, and the ranges were defined as follows, [1, 10], [1, 50], [1, 500], [1, 5000], [1, 50,000], [1, 500,000], and [1, 5,000,000] (the number of pixels). A comparative experiment was designed to evaluate the performance of the static and dynamic batch sizes.

2.5. Computing Environment

This subsection outlines the computing infrastructure used for the experiments, which is critical for optimizing and evaluating the performance of the inference models in real-time scenarios. The hardware configuration included a Lenovo IdeaCenter 5i desktop PC (Lenovo Group Limited, Morrisville, NC, USA) equipped with a 12th generation Intel(R) Core(TM) i7 processor at 2.10 GHz and an NVIDIA GeForce RTX 3060 GPU, which has 12 GB of memory and 3584 CUDA cores. This setup provided sufficient computational power and memory capacity necessary for handling the demanding tasks of DL and real-time data processing required by the study. For software, Python was chosen due to its robust ecosystem and flexibility, crucial for high-performance computational tasks in HSI. The TensorRT Python API played a pivotal role in optimizing DL models for inference on NVIDIA GPUs, enhancing computational efficiency significantly. The specific versions of the software tools employed included CUDA 12.1.1, which maximizes GPU resource utilization, TensorFlow 2.13.1 for conducting baseline tests, and TensorRT 8.6.1 for the optimization and acceleration of the models.

2.6. Performance Evaluation

This subsection elaborates on the metrics utilized to evaluate the performance of four inference engines, focusing on accuracy, latency time, and throughputs to quantify the enhancements in real-time FM detection capabilities. These metrics allow for a thorough comparison of performance before and after optimization. The insights gained from these evaluations are crucial in validating the effectiveness of hardware acceleration and model compression.

2.6.1. Accuracy

The accuracy of FM detection at the pixel level was assessed using precision, recall, F1 score, balanced accuracy (BACC), and overall accuracy (ACC). A binary abnormal pixel prediction (indicative of FM or contamination) was considered a positive outcome, while a binary normal pixel prediction (clean) served as a negative outcome. True positives (TPs) represented the number of correctly predicted FM pixels, and true negatives (TNs) indicated the number of correctly identified normal pixels. False positives (FPs) were the number of normal pixels misclassified as FM, and false negatives (FNs) were the

number of FM pixels misclassified as normal. Precision, defined as the ratio of correct FM predictions to all predictions labeled as FM, and recall, or sensitivity, measured as the ratio of correct FM predictions to the actual number of FM pixels, were calculated. The F1 score, a harmonic mean of precision and recall, ranged from 0 (indicating poor performance) to 1 (indicating perfect accuracy), calculated without considering the TN due to the significant class imbalance in the dataset. Balanced accuracy (BACC) took into account this imbalance, averaging the sensitivity and specificity (true negative rate), with values ranging from 0 to 1.

2.6.2. Latency Time

Inference latency time was measured as the end-to-end execution time of the engine's forward inference cycle, excluding data retrieval, engine initialization, and input preprocessing. To ensure an accurate measurement of inference time, five preliminary "warm-up" operations were conducted. Following these operations, the execution time was recorded over 10 repetitions to compute a robust mean inference time. Time measurements were made using the "time" function from Python's "time" module, capturing timestamps before and after inference execution. Latency time is also called inference time interchangeably in this paper.

2.6.3. Throughput (MB/s)

Inference throughput, measured in MB/s, indicates the amount of data being inferred per second. This measure is calculated by dividing the total amount of inferred data in bytes by the total latency time. As previously mentioned, in our experiments, the batch size began at 1 KB of hyperspectral data, which is approximately equivalent to three pixels, and increased tenfold, until reaching the maximum limit of 1 GB that the GPU could handle without memory issues. It is important to note that the inference engines using TensorRT were specifically built for each batch size—whether static or dynamic—by utilizing CUDA kernels optimized for those sizes. In contrast, TensorFlow-CPU and TensorFlow-GPU engines operated with a static batch size at runtime. The effects of batch size and dynamic batch size over static batch size were analyzed with mean inference time per batch, standard deviation of inference times, and inference throughput.

2.6.4. Hypercube Rate

To better align with the operational demands of poultry processing lines, we evaluated another throughput in terms of the number of hyperspectral images processed per second, called a hypercube rate. This metric also implies how many broiler breast fillets can be inspected per second using the proposed methods. In typical large U.S. poultry processing plants, the evisceration lines operate at 140 BPM, which produce 280 halved breast fillets per minute and equivalently 4.7 halved fillets per second at the downstream lines. The maximum line speed for the poultry industry can reach up to 250 BPM, equating to 8.3 halved fillets per second, in other countries that allow higher line speeds. Thus, it is critical for a vision system to keep up with this speed requirement.

The calculation of the hypercube rate was based on the throughput rate (measured in batches per second or MB/s) and the batch size (measured in pixels per batch or KB). The hyperspectral imaging conditions configured with the two hypothetical NIR and SWIR hyperspectral cameras (see Section 2.4.1) required 107,520 and 256,000 pixels, respectively, to capture a hyperspectral image (i.e., a hypercube) with the camera's maximum spatial dimensions for scanning a halved broiler breast fillet entirely. The SWIR camera can be described as a 256k-pixel SWIR camera with 96 bands, while the NIR camera can be referred to as a 108k-pixel NIR camera with 96 bands. By relating the throughput to these pixel

requirements, we established a metric that provides a realistic gauge of performance across different operational scales and batch sizes for each inference engine examined.

3. Results and Discussion

3.1. Inference Accuracy

Table 1 presents a detailed breakdown of the pixel-level detection accuracies of four inference engines, each with two sub-engines used in this study. The table presents each sub-engine's results as well as the final fused results. In the table, the muscle channel and non-muscle channel refer to sub-engines E1 and E2, which utilize discriminators D1 and D2, respectively. The "fused channels" rows represent the pixel classification results of the final fused inference engine's results, predicted from both muscle and non-muscle channels and appropriate post-processing steps described in Section 2.2 and Figure 2. The "fused channels" rows provide a comprehensive assessment of the performance of inference engines with two sub-engines. Figure 4 shows the detection results of both sub-engines along with the final pixel-level detection compared to the ground truth for a single NIR hyperspectral image with 5 × 5 mm FMs.

Table 1. FM detection accuracies of different inference engines are measured in precision, recall, F1 score, balanced accuracy (BACC), and overall accuracy (ACC).

Inference Engine	Sub-Engine	Precision	Recall	F1 Score	BACC	ACC
TensorFlow-CPU	Muscle channel only	0.881	0.806	0.841	0.902	0.9948
	Non-muscle channel only	0.964	0.796	0.872	0.897	0.9959
	Fused channels	0.981	0.776	0.867	0.888	0.9960
TensorFlow-GPU	Muscle channel only	0.881	0.806	0.841	0.902	0.9948
	Non-muscle channel only	0.964	0.796	0.872	0.897	0.9959
	Fused channels	0.981	0.776	0.867	0.888	0.9960
TensorRT-GPU	Muscle channel only	0.881	0.806	0.841	0.902	0.9948
	Non-muscle channel only	0.963	0.796	0.872	0.897	0.9959
	Fused channels	0.981	0.776	0.867	0.888	0.996
TensorRT-GPU-PTQ	Muscle channel only	0.880	0.806	0.841	0.902	0.9948
	Non-muscle channel only	0.964	0.795	0.872	0.897	0.9959
	Fused channels	0.981	0.776	0.867	0.888	0.9960

Overall, the transition from CPU to GPU platforms, along with the introduction of hardware acceleration and model compression, has had little effect on accuracy. Across different inference engines, the precision, recall, F1 scores, BACC, and ACC remained stable and nearly identical, with only three cases with a negligible difference of 0.1%. Notably, F1 scores consistently exceeded 84.1% for the muscle channel sub-engine and 87.2% for the non-muscle channel sub-engine across all inference engines. The precision and recall of the TensorRT-GPU-PTQ engine were reduced by 0.1% for the muscle channel sub-engine (from 88.1% to 88%) and the non-muscle channel sub-engine (from 77.6% to 77.5%), respectively, while the BACC (88.8%) and ACC (99.6%) remained unchanged. This confirms that both hardware acceleration and model compression to FP16 effectively preserve the inference engine's accuracy without additional model training.

The consistency in accuracy across different inference engines highlights the robustness of the inference model with two GAN discriminators and validates the effectiveness of GPU acceleration and PTQ optimization strategies. These findings provide strong evidence that these strategies can be implemented without compromising model accuracy, making them ideal for high-throughput, real-time detection applications in industrial settings.

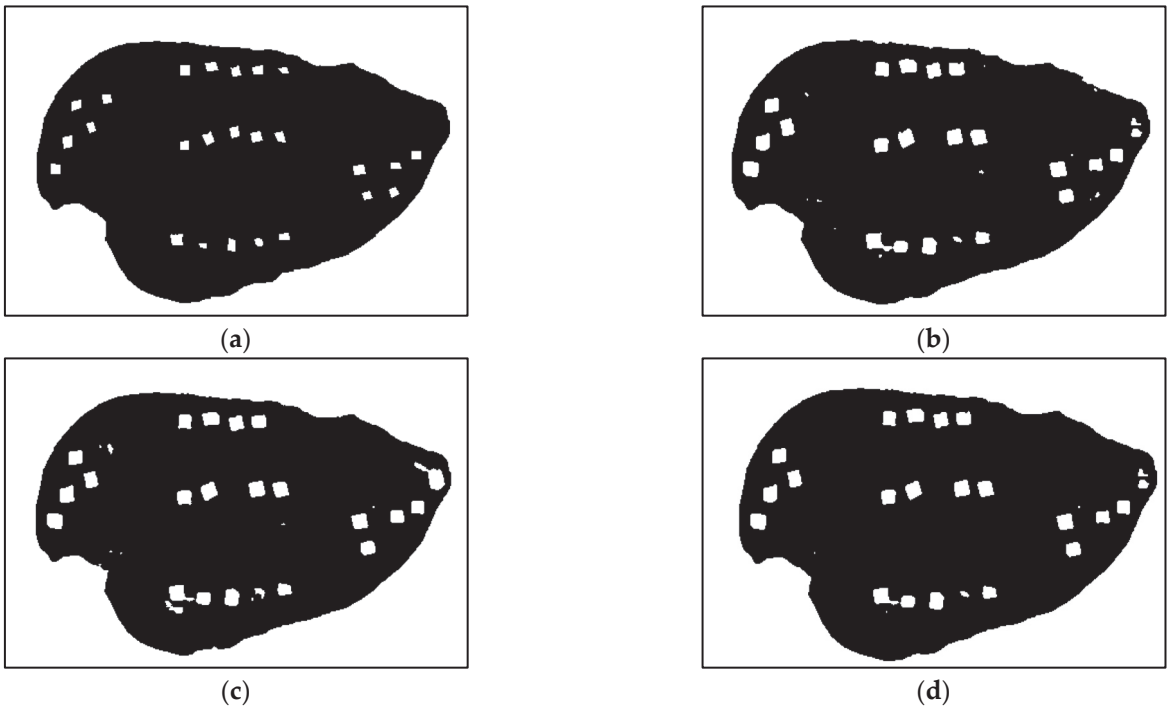


Figure 4. Binary detection result maps from (**a**) ground truth; (**b**) sub-engine E1 (muscle channel); (**c**) sub-engine E2 (non-muscle external tissue channel); (**d**) fused model.

3.2. Static vs. Dynamic Batch Size

We compared the performance of static and dynamic batch size configurations using TensorRT-GPU engines. This analysis is essential because it helps us decide whether we can use a TensorRT inference engine with a single dynamic batch size, which offers great flexibility without building multiple inference engines. This flexibility is particularly beneficial in high-throughput imaging scenarios common in industrial settings without compromising performance. The results, summarized in Table 2, highlight the comparative metrics for both batch-size types across various batch sizes.

Table 2. Comparative performance of batch-size configurations.

Batch Size Type	Batch Size (Pixels)	Size/Batch (KB) [1]	Mean Inference Time/Batch (ms)	SD [2] of Inference Time/Batch (ms)	Throughput (MB/s)
Static	3	1	1.2	0.42	0.81
	26	10	1.09	0.32	8.88
	260	100	1.1	0.31	88.62
	2604	1000	1.4	0.51	696.51
	26,042	10,000	4.84	0.83	2016.02
	260,417	100,000	33.17	0.84	2944.34
	2,604,167	1,000,000	320.63	0.93	3046.88
Dynamic	[1, 3, 10] [3]	1	1.2	0.42	0.81
	[1, 26, 50]	10	1.11	0.32	8.86
	[1, 260, 500]	100	1.2	0.42	81.27
	[1, 2600, 5000]	1000	1.5	0.52	649.81
	[1, 26,042, 50,000]	10,000	4.45	0.49	2194.73
	[1, 260,417, 500,000]	100,000	33.3	0.48	2932.62
	[1, 2,604,167, 5,000,000]	1,000,000	321.95	5.29	3037.11

[1] The values in KB are approximated numbers. [2] SD: Standard deviation [3] [minimum, optimal, maximum] configures a range and an optimal size.

We observed that both static and dynamic batch-size configurations achieved comparable performance across the measured metrics, including mean inference time, standard deviation of inference time, and throughput, at each specified batch size. At the smallest batch size of 1 KB, both configurations maintained an inference time of 1.2 ms and a throughput rate of 0.81 MB/s. As the batch size increased exponentially by powers of 10, throughput (in MB/s) also increased similarly for both batch-size types, attributed to the utilization of GPU processing capabilities. Even at larger batch sizes, such as 100 MB and 1 GB, dynamic batch-size configurations maintained performance comparable to their static counterparts.

These findings suggest that while both static and dynamic batch sizes deliver similar performance in terms of latency time and throughput, dynamic batch sizing offers significant flexibility. It can accommodate a wider range of input sizes without compromising performance efficiency. This flexibility is particularly beneficial in FM detection applications, where changes in line speeds or imaging conditions necessitate an inference engine capable of handling varying input sizes. The minimal performance trade-off, combined with the added flexibility, makes the dynamic batch size configuration preferable for high-throughput, real-time detection environments. As a result, the dynamic batch size became the selected option for the subsequent efficiency analyses in Section 3.3.

3.3. Optimization Performance Analyses

3.3.1. Latency

Latency, defined as the average inference time per batch, is a critical metric for evaluating the real-time performance of inference engines. We examined the inference latency for each sub-engine within across four different inference engines. Figure 5 illustrates latency trends as a function of batch size for both sub-engines across all four inference engines. Generally, latency increased with larger batch sizes due to increased processing requirements for more data, though the rate of increase varied among different inference engines.

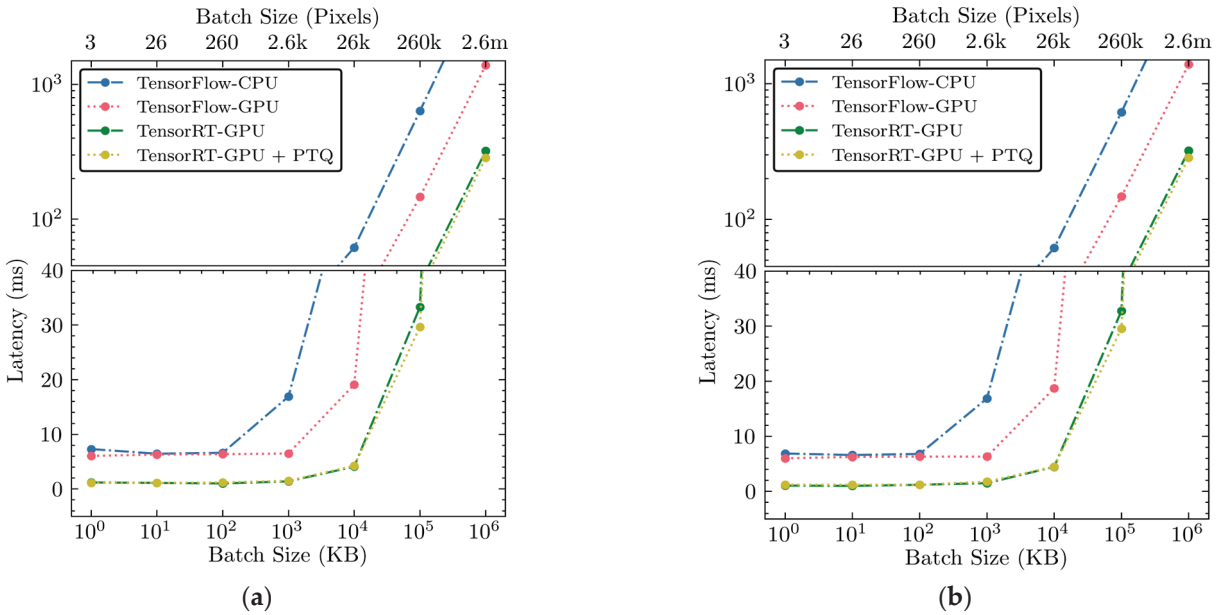


Figure 5. (**a**) Latency versus batch size for sub-engine E1 across all four inference engines; (**b**) latency versus batch size for sub-engine E2 across all four inference engines.

Initially, the unoptimized TensorFlow-CPU and TensorFlow-GPU engines showed negligible latency increments at smaller batch sizes, with latencies ranging from 6.87 ms to 16.86 ms for both inference engines up to 100 KB. However, when the batch size was

increased to the maximum batch size of 1 GB, the TensorFlow-CPU exhibited a steep increase in latency, reaching up to 6.851 s. While the TensorFlow-GPU showed improvements compared to the TensorFlow-CPU, it still showed significant latency increases, peaking at 1.391 s.

In contrast, the hardware-accelerated TensorRT-GPU showed a significant reduction in latency across all batch sizes when compared to the unoptimized inference engines, achieving up to 5 times faster performance than TensorFlow-GPU and 21 times faster than TensorFlow-CPU. At the minimal batch size of 1 KB, latency significantly decreased to 1.2 ms and remained impressively low even at the maximum batch size of 1 GB, where it recorded approximately 322 ms. The hardware acceleration and the introduction of model compression via PTQ in TensorRT-GPU-PTQ further reduced latency, especially for larger batch sizes. This optimization strategy using hardware-accelerated PTQ achieved the lowest latencies, maintaining exceptionally low latency for the largest batch size, capped at around 285 ms for both inference channels.

These findings confirm that using hardware acceleration and model compression not only preserves the quality of predictions but also enhances speed significantly, thus supporting high-efficiency operational needs in poultry processing. The results of this latency analysis suggest that the applied optimizations—particularly the use of PTQ in conjunction with GPU acceleration—provide a robust solution for maintaining throughput without sacrificing performance, marking a substantial advancement in the deployment of deep learning models in real-time detection systems.

3.3.2. Throughputs

Throughput, measured in MB/s, represents the amount of data processed per second and serves as a key measure of model efficiency, particularly in high-throughput environments like poultry processing lines. We assessed the throughputs of four different inference engines for the two sub-engines utilizing muscle and non-muscle GAN discriminators at the pixel level, as shown in Figure 6.

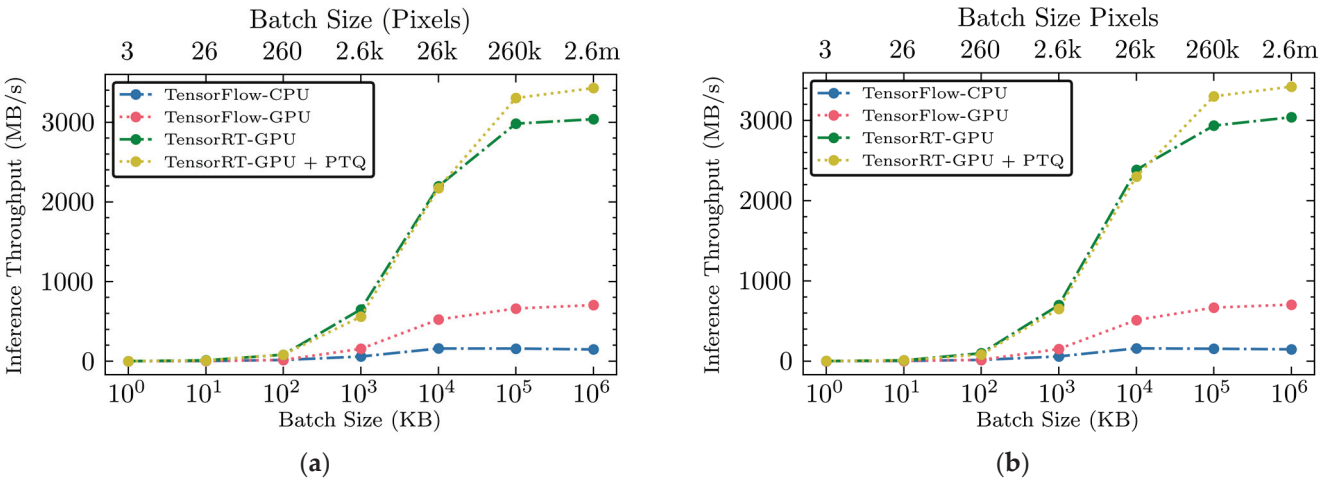


Figure 6. (**a**) Inference throughput versus batch size for sub-engine E1 across all four inference engines; (**b**) Iinference throughput versus batch size for sub-engine E2 across all four inference engines.

The unoptimized TensorFlow-CPU engine showed moderate throughput at smaller batch sizes, starting at 0.13 MB/s for 1 KB. As the batch sizes increased, there was a noticeable improvement in throughput, peaking at 158.5 MB/s for a batch size of 10 MB. At the larger batch sizes of 100 MB and 1 GB, throughput decreased slightly. The unoptimized TensorFlow-GPU engine comprising two sub-engines exhibited better throughput from the

beginning, showing 0.16 MB/s at a 1 KB batch and significantly rising to 703.13 MB/s at the maximum batch size of 1 GB.

In contrast, the hardware-accelerated TensorRT-GPU inference engine demonstrated a substantial enhancement in throughput across all tested batch sizes compared to the unoptimized engines. The throughput with this inference engine started with a higher throughput of 0.81 MB/s at 1 KB and achieved exceptional performance at larger batch sizes, reaching up to 2.933 GB/s at a batch size of 100 MB, and maintaining a high throughput of 3.037 GB/s at the largest batch size of 1 GB. The model compression in conjunction with GPU acceleration through TensorRT marked a significant advancement. Compared to the TensorRT-GPU, the TensorRT-GPU-PTQ maintained similar throughput rates at smaller batch sizes but exhibited superior performance at larger batch sizes. At a batch size of 100 MB, this engine achieved a throughput of 3.298 GB/s, the highest among all configurations. This peak throughput underscores the model's ability to efficiently handle large volumes of data, which is crucial for real-time applications that require rapid processing.

Interestingly, across all tested configurations, throughput tended to converge around the 100 MB batch size (i.e., 260k pixels). This convergence suggests a level-off point where each engine reaches its processing capacity limit under the hardware constraints used in the study. The throughput results indicate that while all engines show increased efficiency with larger batch sizes, the optimized GPU-accelerated inference engines, especially those optimized with PTQ, provide significant improvements in processing speed. This demonstrates their suitability for deployment in industrial settings, where both high data volume and swift processing are crucial.

3.3.3. Hypercube Rates

In this analysis, we evaluated the performance of four different inference engines based on their ability to process hyperspectral images per second, a crucial measure for real-time applications in poultry processing. This metric indicates how many hyperspectral data cubes each inference engine can process in one second, using two types of hypothetical hyperspectral cameras mentioned in Section 2.4.1. Figures 7 and 8 show the processing rates for each camera across different batch sizes and inference engines with two sub-engines.

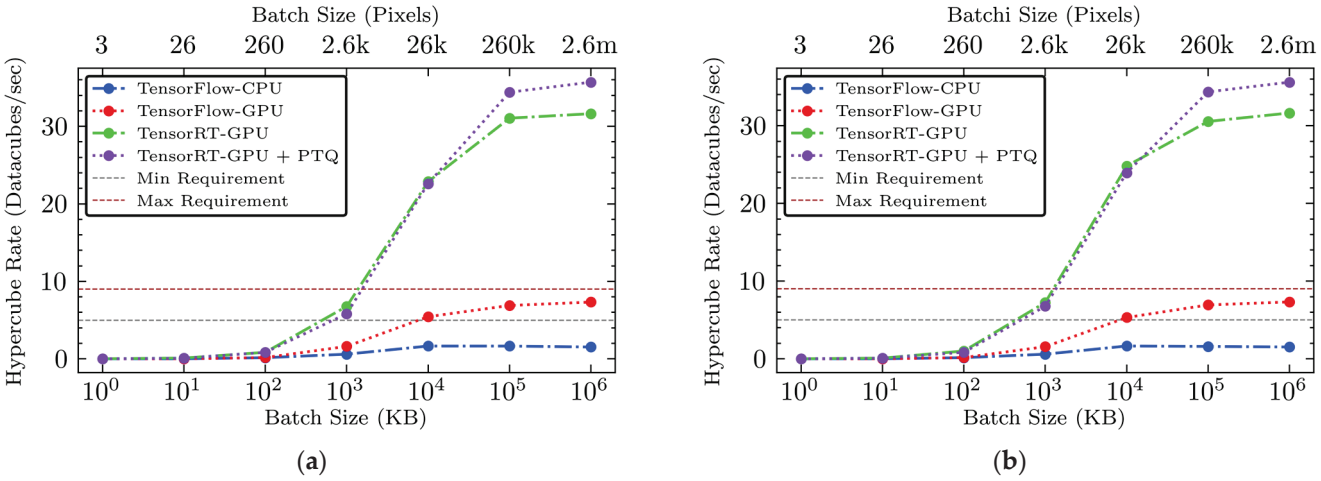


Figure 7. (**a**) The 256k-pixel SWIR camera hypercube rate versus batch size for sub-engine E1 (muscle discriminator) across all four inference engines; (**b**) 256k-pixel SWIR camera hypercube rate versus batch size for sub-engine E2 (non-muscle discriminator) across all four inference engines.

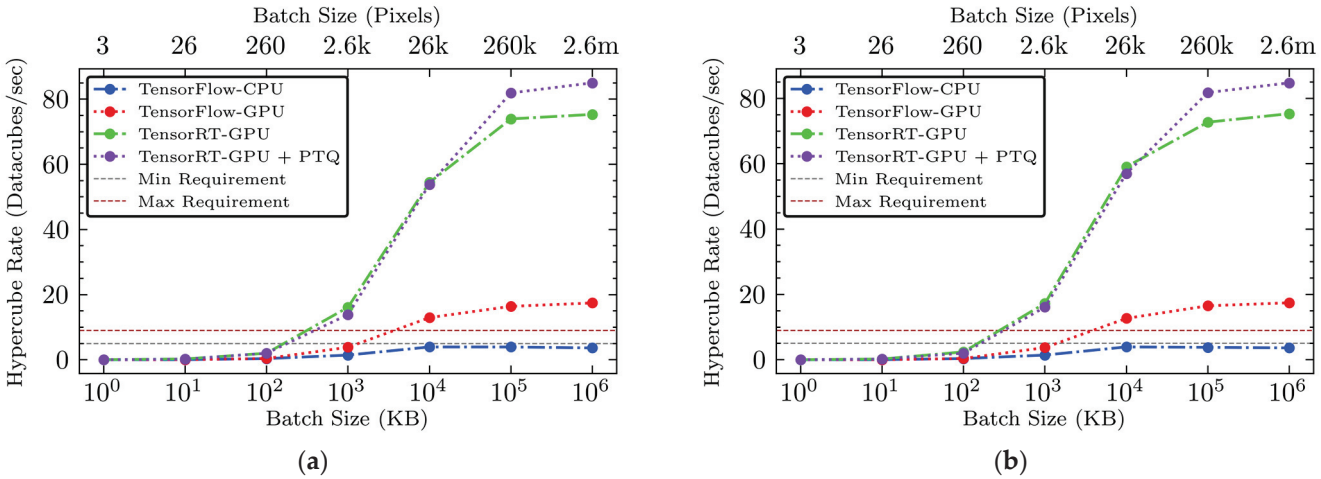


Figure 8. (**a**) The 108k-pixel NIR camera hypercube rate versus batch size for sub-engine E1 (utilizing muscle discriminator) across all four inference engines; (**b**) 108k-pixel NIR camera hypercube rate versus batch size for sub-engine E2 (utilizing non-muscle discriminator) across all four inference engines.

The unoptimized TensorFlow-CPU engine demonstrated limited processing capabilities, achieving only two images per second for the 256k-pixel SWIR camera and three images per second for the 108k-pixel NIR camera. The unoptimized TensorFlow-GPU engine showed improved performance, allowing the processing of seven hyperspectral images per second for the 256k-pixel SWIR camera and 16 for the 108k-pixel NIR camera at its maximum. Despite this enhancement, the performance of the 256k-pixel camera, which produces 2.4 times more data than the 108k-pixel camera, still falls short of meeting the high-speed requirements for processing approximately 5–9 hypercubes per second. This capability is essential for inspecting 4.7–8.3 halved fillets per second at 140 and 250 BPM, respectively, in poultry processing, highlighting the need for more efficient processing solutions. A significant advancement was achieved with the optimized inference engines utilizing TensorRT, which significantly increased hypercube rates when the batch size reached approximately 26k pixels (10 MB) or more. The hardware-accelerated TensorRT-GPU engine achieved processing speeds up to 31 hypercubes per second for the 256k-pixel SWIR camera and 75 hypercubes per second for the 108k-pixel NIR cameras at the largest batch size of 2.6 million pixels (1 GB). This improvement emphasizes the substantial impact of hardware-accelerated optimization on both throughput and real-time processing capabilities. Further enhancements were achieved with the integration of model compression in TensorRT-GPU-PTQ, establishing new benchmarks in processing efficiency. The TensorRT-GPU-PTQ processed over 35 hypercubes per second for the 256k-pixel SWIR camera and 84 hypercubes per second for the 108k-pixel NIR camera at the largest batch size of 1 GB, clearly surpassing the throughput requirements for real-time processing and confirming its suitability for industrial deployment. Models, especially those optimized with PTQ, provide significant improvements in processing speed. This demonstrates their suitability for deployment in industrial settings.

3.4. Discussion

The findings of this study suggest the significant advantages of advanced optimization techniques, specifically the combination of model compression and GPU hardware acceleration, for the FM detection application utilizing a deep learning model in poultry. Although the simulation conducted in this study remains partially theoretical, the validated computational workflows and optimizations provide a conceptual framework for potential implementation in live operational settings. The integration of these optimization technolo-

gies shows promise for substantial improvements in poultry meat safety and processing efficiency, aligning with the industry's shift toward more automated and precise contamination detection systems. The optimization strategy with GPU-based acceleration and model compression significantly enhanced real-time hyperspectral imaging capabilities under high-speed imaging conditions. The optimized inference engine, equipped with model compression and hardware acceleration, demonstrated the ability to process up to 35 and 82 hyperspectral images per second for two hypothetical 256k-pixel SWIR and 108k-pixel NIR hyperspectral cameras, respectively. When each processed hyperspectral image covers the entire surface of a halved chicken breast fillet, these results suggest that the proposed optimized methods could monitor approximately 2000 to 5000 BPM, far exceeding the typical poultry processing line speed of around 140, peaking at 250 per minute (i.e., 280–500 halved fillets per minute). Such high performance will allow for the immediate detection and removal of FMs using a sorting machine, such as a mechanical diverter, which is critical for maintaining quality and safety standards. The experiment with dynamic batch sizes demonstrated that this approach increased the model's adaptability and ensured flexibility, enabling it to effectively respond to varying input sizes while maintaining consistent output under diverse imaging conditions. These optimizations enhance processing speeds and can be applied to other rapid processing applications. Overall, the results indicate that the hardware-accelerated model compression can significantly advance industrial hyperspectral imaging applications in other food processing contexts that require both rapid and reliable real-time detection. However, this study has several limitations. One primary limitation is its current hardware dependence. The optimizations were heavily reliant on NVIDIA GPUs, which may limit the applicability of the performance enhancements to other hardware platforms, such as FPGAs, TPUs, or custom AI accelerators, where different frameworks or precision formats may be required. Although NVIDIA GPUs are considered an industry standard in GPU computing, we recognize the growing diversity of hardware solutions available. In principle, the same core optimizations—post-training quantization and hardware-aware model acceleration—can be adapted to alternative architectures by using equivalent vendor toolchains (e.g., Vitis AI for FPGAs or TensorFlow Lite for edge TPUs). Future research will investigate the feasibility and performance trade-offs of such implementations to broaden the accessibility and cost-effectiveness of these techniques.

Moreover, our work focused exclusively on FP16 post-training quantization (PTQ). While FP16 effectively balances memory efficiency and model accuracy, more aggressive approaches like INT8 quantization could further reduce computational overhead and storage requirements. However, this often requires careful calibration using a representative calibration dataset and, in some cases, fine-tuning to maintain the high sensitivity required by hyperspectral data, which goes beyond our study scope of optimizing the model without any re-training. In future work, we plan to investigate INT8 PTQ to evaluate its potential benefits in terms of both model size and real-time inference speed, while also exploring strategies to minimize any performance degradation due to lower precision.

It is important to note that while this study focused specifically on detecting FMs on boneless, skinless chicken breast fillets, the compression and acceleration techniques proposed in this study are largely independent of the detection model. Adapting our pipeline for other types of meat, such as pork or beef, or even plant-based products, would just need to retrain or fine-tune the underlying deep learning model using the relevant hyperspectral data. However, the proposed optimization methods, such as quantization and GPU acceleration, would still be applicable and effective.

Furthermore, while the experimental validation was robust, again, it was conducted under controlled conditions and might not fully capture the complexities or unpredictable variables present in real-world production scenarios. The current dataset was limited to

the NIR spectral range of 1000–1700 nm and a restricted set of FMs for testing. This NIR wavelength range may not fully represent the real-world hyperspectral imaging conditions of chicken fillets, including potential FM contaminants and environmental factors, such as variations in lighting–material interactions, surface textures, and seasonal changes in chicken breast fillets. Future studies should include a broader variety of fillet types for training the model and include additional contaminant types for testing, ensuring greater applicability. An expanded dataset, combined with more diverse sampling conditions, could provide a better understanding of the optimized inference model's robustness and its capacity to adapt to the challenges faced in poultry processing. Additionally, advanced data augmentation strategies that simulate variable sample surfaces and types may help bridge the gap between laboratory simulations and real-world applications.

Lastly, the inference model, which consists of two discriminators, was not executed on multiple GPUs, nor integrated as a single, fully parallelized inference engine in this study. A straightforward parallelization strategy could involve running each sub-engine on a separate GPU, which could theoretically reduce the total inference latency by half if hardware resources support concurrent execution. Alternatively, multi-stream processing or custom CUDA kernels on a single GPU might enable parallel execution of both discriminators. Nevertheless, our current serial approach still meets industrial speed requirements. Even if performance is reduced by half for a batch size of 26k pixels or more, the model exceeds typical throughput demands (as shown in Figures 7 and 8). While the implementation and benchmarking of these advanced strategies are beyond the scope of this study, our latency and throughput data (Section 3.3) suggest that a parallel approach could yield additional gains in high-throughput environments. Therefore, future research may involve deploying two GPUs simultaneously or designing a single GPU-based inference engine with synchronized custom CUDA kernels for dual inference channels, thereby enhancing performance in real-world poultry processing lines.

To summarize, there are several potential future research avenues to expand upon the current findings and further advance hyperspectral imaging technology. Investigating advanced quantization techniques, particularly INT8 quantization, could lead to more compact models with reduced computational demands. However, this approach would require model tuning and calibration to maintain accuracy. Exploring computationally efficient architectures during model development could significantly reduce resource demands, making these systems more viable on constrained hardware platforms. Extending optimization efforts to include a broader range of platforms, such as FPGAs or specialized AI accelerators, could diversify the utility and application of these models across various industrial settings. Expanding the dataset to include various types of chicken fillets and FMs, as well as different imaging environments, could further validate the system's robustness and ensure its real-world applicability. Finally, implementing these optimized models in actual processing settings and gathering feedback on their operational performance would provide invaluable insights into their practical effectiveness and highlight areas needing refinement. This real-world implementation is crucial for fine-tuning the models and ensuring they meet the stringent requirements of modern industrial applications, ultimately leading to more effective and adaptable hyperspectral imaging solutions.

4. Conclusions

This study has effectively demonstrated the capabilities of GPU hardware acceleration and model compression techniques in enhancing the real-time detection of FMs in poultry processing, predicting processing rates of up to 82 chicken fillets per second. Such performance well exceeds the production rate of poultry processing lines, significantly improving safety and quality control measures in the poultry industry. The integration of advanced DL

technologies with hyperspectral imaging has established new benchmarks for processing speeds and accuracy, proving essential for high-throughput environments where rapid and reliable detection is paramount. The application of PTQ and GPU acceleration has successfully addressed critical challenges related to the high-dimensional data typical of hyperspectral imaging, optimizing the detection process without compromising accuracy. Furthermore, the system's ability to handle dynamic batch sizes ensured flexibility and consistent performance across varying imaging conditions. These advancements underscore the potential of such technologies to apply to many other industrial imaging applications for monitoring safety and efficiency. Nonetheless, this work is primarily designed for the NVIDIA GPU ecosystem. Adapting the proposed compression and acceleration strategies for alternative hardware platforms, such as TPUs, FPGAs, and edge accelerators, will present a valuable opportunity for exploration. This adaptation can enhance the accessibility and cost-effectiveness of real-time hyperspectral imaging solutions. Additionally, while our experiments focused on FP16 PTQ, implementing more aggressive quantization methods such as INT8 could further improve throughputs and memory savings, provided that they are carefully calibrated. Future investigations should explore these methods while balancing trade-offs between precision, detection accuracy, and practical deployment constraints. It is also essential to conduct future studies that incorporate additional contaminants and varying imaging conditions to validate model robustness under real production scenarios. This broader validation will confirm the system's feasibility for industrial uses and enhance its overall reliability. It is also important to note that our tests were conducted in a controlled laboratory environment. In contrast, industrial processing lines present a range of real-world variables—such as conveyor speeds of up to 250 birds per minute, mechanical vibrations, and fluctuations in temperature or humidity changes—that can affect hyperspectral imaging. Consequently, pilot-scale or commercial-level testing in the future will be crucial for assessing our system's resilience and identifying any additional optimizations necessary to maintain high detection accuracy under operational conditions. In conclusion, this study contributes significantly to the field of real-time hyperspectral imaging in food safety, pushing the boundaries of what is possible in FM detection. It paves the way for future innovations that could further enhance the robustness and efficiency of food processing technologies, ensuring they keep pace with the demands of modern production environments and regulatory standards.

Author Contributions: Conceptualization, S.-C.Y.; methodology, Z.K., S.-C.Y. and S.M.B.; software, Z.K.; writing—original draft preparation, Z.K.; writing—review and editing, Z.K., S.-C.Y. and S.M.B.; project administration, S.-C.Y. and S.M.B.; funding acquisition, S.-C.Y. and S.M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the USDA—Agriculture Research Service through a research support agreement between the USDA—Agriculture Research Service and the University of Georgia, with agreement number 58-6040-4-033.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author but are not publicly accessible without authorization from the USDA-ARS.

Acknowledgments: We would like to express our gratitude to Rodrigo Campos, a graduate student at the School of Computing, University of Georgia, for his technical support. Additionally, we extend our thanks to Soo Chung, an assistant professor in the Department of Biosystems Engineering at Seoul National University in Seoul, Republic of Korea, for his assistance with data collection during his time as a postdoctoral researcher at USDA-ARS.

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

References

1. Payne, K.; O'Bryan, C.A.; Marcy, J.A.; Crandall, P.G. Detection and Prevention of Foreign Material in Food: A Review. *Heliyon* **2023**, *9*, e19574. [CrossRef] [PubMed]
2. Barbut, S.; Leishman, E.M. Quality and Processability of Modern Poultry Meat. *Animals* **2022**, *12*, 2766. [CrossRef] [PubMed]
3. USDA-FSIS Presence of Foreign Material in Meat or Poultry Products—Revision 4. Available online: https://www.fsis.usda.gov/policy/fsis-directives/7310.5 (accessed on 20 November 2024).
4. Zhu, L.; Spachos, P.; Pensini, E.; Plataniotis, K.N. Deep Learning and Machine Vision for Food Processing: A Survey. *Curr. Res. Food Sci.* **2021**, *4*, 233–249. [CrossRef] [PubMed]
5. Kwon, J.-S.; Lee, J.-M.; Kim, W.-Y. Real-Time Detection of Foreign Objects Using X-Ray Imaging for Dry Food Manufacturing Line. In Proceedings of the 2008 IEEE International Symposium on Consumer Electronics, Vilamoura, Portugal, 14 April 2008.
6. Jha, S.N.; Narsaiah, K.; Basediya, A.L.; Sharma, R.; Jaiswal, P.; Kumar, R.; Bhardwaj, R. Measurement Techniques and Application of Electrical Properties for Nondestructive Quality Evaluation of Foods—A Review. *J. Food Sci. Technol.* **2011**, *48*, 387–411. [CrossRef]
7. Pallav, P.; Diamond, G.G.; Hutchins, D.A.; Green, R.J.; Gan, T.H. A Near-Infrared (NIR) Technique for Imaging Food Materials. *J. Food Sci.* **2009**, *74*, E23–E33. [CrossRef]
8. Ristau, K. Using Foreign Material Detection Methods to Assist the Food Industry. Master's Thesis, Kansas State University, Manhattan, KS, USA, 2022.
9. Soni, A.; Dixit, Y.; Reis, M.M.; Brightwell, G. Hyperspectral Imaging and Machine Learning in Food Microbiology: Developments and Challenges in Detection of Bacterial, Fungal, and Viral Contaminants. *Compr. Rev. Food Sci. Food Saf.* **2022**, *21*, 3717–3745. [CrossRef]
10. Campos, R.L.; Yoon, S.-C.; Chung, S.; Bhandarkar, S.M. Semisupervised Deep Learning for the Detection of Foreign Materials on Poultry Meat with Near-Infrared Hyperspectral Imaging. *Sensors* **2023**, *23*, 7014. [CrossRef]
11. Bera, S.; Shrivastava, V.K.; Satapathy, S.C. Advances in Hyperspectral Image Classification Based on Convolutional Neural Networks: A Review. *CMES-Comput. Model. Eng. Sci.* **2022**, *133*, 219–250. [CrossRef]
12. Chung, S.; Yoon, S.-C. Detection of Foreign Materials on Broiler Breast Meat Using a Fusion of Visible Near-Infrared and Short-Wave Infrared Hyperspectral Imaging. *Appl. Sci.* **2021**, *11*, 11987. [CrossRef]
13. Seo, Y.; Kim, G.; Lim, J.; Lee, A.; Kim, B.; Jang, J.; Mo, C.; Kim, M.S. Non-Destructive Detection Pilot Study of Vegetable Organic Residues Using VNIR Hyperspectral Imaging and Deep Learning Techniques. *Sensors* **2021**, *21*, 2899. [CrossRef]
14. Yang, D.; Jiang, J.; Jie, Y.; Li, Q.; Shi, T. Detection of the Moldy Status of the Stored Maize Kernels Using Hyperspectral Imaging and Deep Learning Algorithms. *Int. J. Food Prop.* **2022**, *25*, 170–186. [CrossRef]
15. Yoon, S.C.; Park, B.; Lawrence, K.C.; Windham, W.R.; Heitschmidt, G.W. Line-Scan Hyperspectral Imaging System for Real-Time Inspection of Poultry Carcasses with Fecal Material and Ingesta. *Comput. Electron. Agric.* **2011**, *79*, 159–168. [CrossRef]
16. Rokh, B.; Azarpeyvand, A.; Khanteymoori, A. A Comprehensive Survey on Model Quantization for Deep Neural Networks in Image Classification. *ACM Trans. Intell. Syst. Technol.* **2023**, *14*, 1–50. [CrossRef]
17. Boggust, A.; Sivaraman, V.; Assogba, Y.; Ren, D.; Moritz, D.; Hohman, F. Compress and Compare: Interactively Evaluating Efficiency and Behavior Across ML Model Compression Experiments. *IEEE Trans. Vis. Comput. Graph.* **2024**, *31*, 809–819. [CrossRef]
18. Krichen, M.; Abdalzaher, M.S. Performance Enhancement of Artificial Intelligence: A Survey. *J. Netw. Comput. Appl.* **2024**, *232*, 104034. [CrossRef]
19. Li, Z.; Li, H.; Meng, L. Model Compression for Deep Neural Networks: A Survey. *Computers* **2023**, *12*, 60. [CrossRef]
20. Menghani, G. Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better. *ACM Comput. Surv.* **2023**, *55*, 1–37. [CrossRef]
21. Choudhary, T.; Mishra, V.; Goswami, A.; Sarangapani, J. A Comprehensive Survey on Model Compression and Acceleration. *Artif. Intell. Rev.* **2020**, *53*, 5113–5155. [CrossRef]
22. Marinó, G.C.; Petrini, A.; Malchiodi, D.; Frasca, M. Deep Neural Networks Compression: A Comparative Survey and Choice Recommendations. *Neurocomputing* **2023**, *520*, 152–170. [CrossRef]
23. Deng, L.; Li, G.; Han, S.; Shi, L.; Xie, Y. Model Compression and Hardware Acceleration for Neural Networks: A Comprehensive Survey. *Proc. IEEE* **2020**, *108*, 485–532. [CrossRef]
24. Han, S.; Pool, J.; Tran, J.; Dally, W.J. Learning Both Weights and Connections for Efficient Neural Networks. In Proceedings of the 28th International Conference on Neural Information Processing Systems, Cambridge, MA, USA, 7 December 2015.
25. Buluc, A.; Gilbert, J.R. Challenges and Advances in Parallel Sparse Matrix-Matrix Multiplication. In Proceedings of the 2008 37th International Conference on Parallel Processing, Portland, OR, USA, 12 September 2008.

26. Mao, H.; Han, S.; Pool, J.; Li, W.; Liu, X.; Wang, Y.; Dally, W.J. Exploring the Granularity of Sparsity in Convolutional Neural Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21 July 2017.
27. Gholami, A.; Kim, S.; Dong, Z.; Yao, Z.; Mahoney, M.W.; Keutzer, K. A Survey of Quantization Methods for Efficient Neural Network Inference, 1st Edition. In *Low-Power Computer Vision*; Chapman and Hall/CRC: New York, NY, USA, 2022; p. 438.
28. Jacob, B.; Kligys, S.; Chen, B.; Zhu, M.; Tang, M.; Howard, A.; Adam, H.; Kalenichenko, D. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 19 June 2018.
29. Horowitz, M. Computing's Energy Problem (and What We Can Do About It). In Proceedings of the 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), San Francisco, CA, USA, 13 February 2014.
30. Bengio, Y.; Léonard, N.; Courville, A. Estimating or Propagating Gradients Through Stochastic Neurons for Conditional Computation. *arXiv* **2013**, arXiv:1308.3432.
31. Krishnamoorthi, R. Quantizing Deep Convolutional Networks for Efficient Inference: A Whitepaper. *arXiv* **2018**, arXiv:1806.08342.
32. Nagel, M.; Baalen, M.V.; Blankevoort, T.; Welling, M. Data-Free Quantization Through Weight Equalization and Bias Correction. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1325–1334.
33. Dai, S.; Venkatesan, R.; Ren, H.; Zimmer, B.; Dally, W.J.; Khailany, B. VS-Quant: Per-Vector Scaled Quantization for Accurate Low-Precision Neural Network Inference. *arXiv* **2021**, arXiv:2102.04503.
34. Banner, R.; Nahshan, Y.; Soudry, D. Post Training 4-Bit Quantization of Convolutional Networks for Rapid-Deployment. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 8 December 2019.
35. 8-Bit Inference with TensorRT. Available online: https://www.cse.iitd.ac.in/~rijurekha/course/tensorrt.pdf (accessed on 20 November 2024).
36. Sakr, C.; Dai, S.; Venkatesan, R.; Zimmer, B.; Dally, W.; Khailany, B. Optimal Clipping and Magnitude-Aware Differentiation for Improved Quantization-Aware Training. In Proceedings of the 39th International Conference on Machine Learning, Baltimore, MD, USA, 28 June 2022.
37. Nvidia Ampere GPU Architecture Whitepaper V2. Available online: https://www.nvidia.com/content/PDF/nvidia-ampere-ga-102-gpu-architecture-whitepaper-v2.pdf (accessed on 22 January 2025).
38. Micikevicius, P.; Narang, S.; Alben, J.; Diamos, G.; Elsen, E.; Garcia, D.; Ginsburg, B.; Houston, M.; Kuchaiev, O.; Venkatesh, G.; et al. Mixed Precision Training. *arXiv* **2018**, arXiv:1710.03740.
39. Ma, T.; Schimleck, L.; Dahlen, J.; Yoon, S.-C.; Inagaki, T.; Tsuchikawa, S.; Sandak, A.; Sandak, J. Comparative Performance of NIR-Hyperspectral Imaging Systems. *Foundations* **2022**, *2*, 523–540. [CrossRef]
40. Mehaffey, J.M.; Pradhan, S.P.; Meullenet, J.F.; Emmert, J.L.; McKee; Owens, C.M. Meat Quality Evaluation of Minimally Aged Broiler Breast Fillets from Five Commercial Genetic Strains. *Poult. Sci.* **2006**, *85*, 902–908. [CrossRef]
41. Mallmann, B.d.A.; Tellez-Isaias, G.; Mauromoustakos, A.; Coon, C.N.; Owens, C.M. Fillet Dimensions and Meat Quality Attributes Associated with Woody Breast in Broilers. *Meat Muscle Biol.* **2020**, *4*, 1–9.

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.

sensors

MDPI

Article

Automatic Measurement of Frontomaxillary Facial Angle in Fetal Ultrasound Images Using Deep Learning

Zhonghua Liu [1,*], Jin Wang [2], Guorong Lyu [3], Haisheng Song [2], Weifeng Yu [1], Peizhong Liu [4], Yuling Fan [4] and Yaocheng Wan [5]

[1] Department of Ultrasound, Quanzhou First Hospital Affiliated to Fujian Medical University, Quanzhou 350122, China; yuweifengab@163.com
[2] School of Physics and Electronic Engineering, Northwest Normal University, Lanzhou 730070, China; jinwang202406@163.com (J.W.); songhs@nwnu.edu.cn (H.S.)
[3] Department of Ultrasound, Second Affiliated Hospital of Fujian Medical University, Quanzhou 362000, China
[4] School of Engineering, Huaqiao University, Quanzhou 362021, China; pzliu@hqu.edu.cn (P.L.); yling_fan@126.com (Y.F.)
[5] School of Electronic and Information Engineering, Changchun University of Science and Technology, Changchun 130022, China; wyc202412@163.com
* Correspondence: lgr_feus@sina.com (G.L.); liuzhonghua2005@126.com (Z.L.)

Abstract: Accurate measurement of frontomaxillary facial (FMF) angles in prenatal ultrasound (US) scans plays a pivotal role in the screening of trisomy 21. Nevertheless, this intricate procedure heavily relies on the proficiency of the ultrasonographer and tends to be a time-intensive task. Furthermore, FMF angles are subjective when measured manually. To address this challenge, we propose a deep learning-based assisted examination framework for automatically measuring FMF angles on 2D ultrasound images. Firstly, we trained a deep learning network using 1549 fetal ultrasound images to achieve automatic and accurate segmentation of critical areas. Subsequently, a key point detection network was employed to predict the coordinates of the requisite points for calculating FMF angles. Finally, FMF angles were obtained through computational means. We employed Pearson correlation coefficients and Bland–Altman plots to assess the correlation and consistency between the model's predictions and manual measurements. Notably, our method exhibited a mean absolute error of 2.354°, outperforming the typical standards of the junior expert. This indicates the high degree of accuracy and reliability achieved by our approach.

Keywords: automatic measurement; deep learning; semantic segmentation; ultrasound image

1. Introduction

Trisomy 21, commonly referred to as Down syndrome (DS), is a prevalent chromosomal abnormality that often manifests in cognitive impairment, facial dysmorphism, delayed growth, developmental delays, and a diverse array of congenital malformations. While people without DS have 46 chromosomes, people with DS have one more chromosome 21, and the extra chromosome monosomy affects their physical and intellectual development [1]. In a study, it was established that there is a correlation between the occurrence of DS and the age of pregnancy. Specifically, the probability of this condition increases concomitantly with advancing maternal age [1]. DS is an incurable lifelong condition that significantly affects individuals' health, so early detection helps families and wider society to prepare for this. Blood tests and ultrasound can be performed in early- or mid-pregnancy for DS. Blood testing is not a conclusive diagnostic tool for DS, but it can flag a heightened risk of the condition.

Meanwhile, more invasive procedures like amniocentesis and chorionic villus sampling (CVS) offer a more definitive diagnosis, albeit with a slight risk of complications that may potentially result in miscarriage. Ultrasonography is a non-invasive method, as patients with DS often have specific facial features that ultrasound can observe [2,3]. With the continuous development of ultrasound medical technology, ultrasound-based fetal facial examination is the most intuitive and effective means of clinical examination. Ultrasound is a non-invasive, non-radioactive, painless, and affordable technology that can visualize facial disorders by receiving echo information from the fetus' face and its collateral structures. Ultrasound is a safe, radiation-free test that causes no harm even when performed multiple times during pregnancy [3,4]. It has been the first choice for prenatal screening of fetuses for decades.

Recent research has emphasized the potential utility of measuring the FMF angle as a supplementary screening tool for trisomy 21 [5]. Moreover, ultrasound evaluations during pregnancy have revealed that fetuses with DS exhibit shorter maxillary lengths and shallower depth in comparison to haploid fetuses [6]. In fetuses with trisomy 21, the FMF angle is notably elevated compared to that observed in haploid fetuses [7]. The study's primary metric, termed FMFbone, represents the angle formed between the superior margin of the upper palate and the bony forehead. As depicted in Figure 1, the first ray traverses the anterior section of the upper palate, commencing from the superior margin of the maxilla, while the second ray originates from the apex along the frontal bone, defining the FMFbone angle. The measurement of FMF angle remains a challenging task with three main difficulties, as shown in Figure 2:

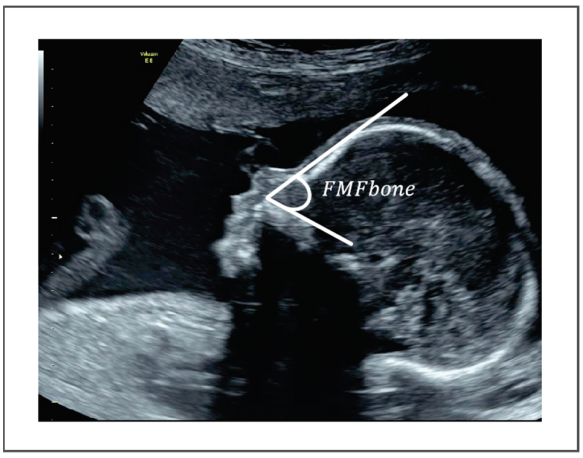


Figure 1. Measurement method of the FMFbone.

- Poor image quality may affect the doctor's judgment. Due to the complex interaction between US waves and the biological tissues of the mother and fetus, ultrasound images can be affected by speckle noise, acoustic shadow, motion blurring, missing boundaries, and poor picture clarity. This situation can affect the physician's measure result. For instance, the definition of the palatal margin might become indistinct, evident from the white arrow depicted in Figure 2a,b;
- Judgment of critical structures on ultrasound images is more difficult. The anterior segment of the palate stands out as the most prominent, while the posterior region might exhibit thickening and irregularity, as illustrated by the blue arrow in Figure 2c,d, which can affect the physician's judgment;
- Manual measurement inherently incorporates a degree of subjectivity. The subjective nature of manual measurement can be problematic. The poor performance of manual measurement can be attributed to the lack of standardized training.

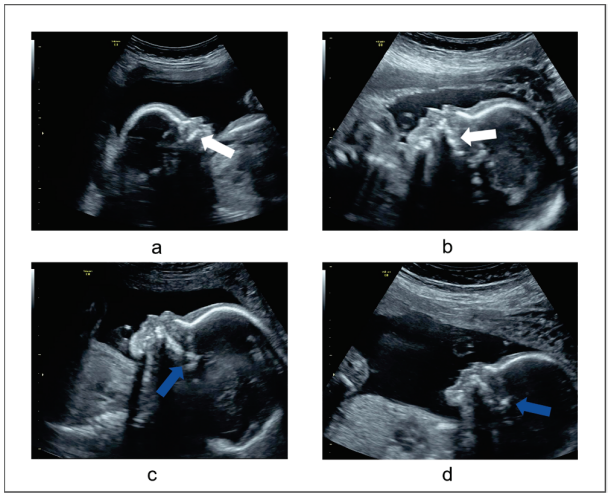


Figure 2. (**a**) Ambiguous maxillary structure. the white arrows indicate the definition of the palatal margin might become indistinct. (**b**) Irregular maxillary structure, Indicated by the white arrow (**c**) Thickening of the posterior part of the maxilla. The blue arrow points it out. (**d**) Irregularity of the posterior part of the maxilla, As shown by the blue arrow.

Prior investigations have demonstrated the remarkable capabilities of deep learning in attaining cutting-edge results across the domains of computer vision and biomedical measurement [8]. Deep learning has proven to be an effective tool for processing medical images. Ref. [9] completed the measurement of the FMF angle during the first trimester. Chen et al. [10] successfully implemented an automated approach for measuring the width of the fetal lateral ventricle (LV) in 2D ultrasound images. The methodology achieved an average absolute measurement error of 1.8 mm, outperforming manual measurement techniques. Van et al. [11] proposed a multitask learning network, MRNet, which implements segmentation of the lumbar spine and automatic parameter prediction tasks for lumbosacral spine examination. Hyun et al. [12] used deep learning for the first time to provide an automated assessment of amniotic fluid index on ultrasound images to provide an automated measurement of amniotic fluid index (AFI). Jang et al. [13] used a convolutional neural network to classify the image and then used the Hough transform method to automatically measure the abdominal circumference (AC); following acceptance check experiments, the accuracy level obtained with the expert was 0.809. Pluym et al. Ref. [14] aims to compare the accuracy of measurements of fetal intracranial structures using the artificial intelligence tool SonoCNS Fetal Brain with manual measurements. This device comes from the General Electric Company (Chicago, IL, USA). The research results indicate that automated measurements exhibit good reliability (with Intraclass Correlation Coefficients, ICCs, ranging from 0.80 to 0.88) compared to manual measurements for biparietal diameter (BPD) and head circumference (HC). Li et al., in [15], introduced a variety of novel biometric markers, including biparietal diameter (BPD) and occipitofrontal diameter (OFD). Deep learning is widely used in ultrasound images [16–20]. Computer-aided diagnostic techniques remove the need for manual tracing and could both increase efficiency and improve reproducibility. FMF angles are usually measured manually by ultrasonographers. This requires a high level of medical expertise and clinical experience. Subjectivity makes it difficult to perform the examination in primary hospitals. To efficiently harness the time and expertise of specialized physicians, we assert that cutting-edge artificial intelligence technology is imperative for automating biometric measurements. This automation not only enhances speed and reproducibility, but also elevates measurement accuracy. Implementing automated measurement techniques holds paramount importance in refining the prenatal diagnosis workflow and enhancing its precision.

In this paper, we propose a novel deep learning-driven measurement method for the first time, with the aim of designing an automated approach for measuring the FMF angle using sagittal ultrasound images of 1549 fetuses during the second trimester for examination and diagnosis. The proposed methodology is founded upon an advanced segmentation network, which is capable of efficiently segmenting the ultrasound images. This segmentation process is then utilized to predict the coordinates of the points necessary for angle measurement, and subsequently, to calculate the angle. During the experimental phase, physicians manually annotated the training data at the pixel level, which were then used to train our deep learning network. The experimental results demonstrate that the proposed method achieves a high level of accuracy. The primary contributions of this study are as follows:

- This paper presents an innovative approach to FMF angle measurement in fetuses during the second trimester, employing a novel automated method based on deep learning.
- This method accelerates the process of prenatal diagnosis and enhances the accuracy of the diagnosis.
- The accuracy of our method is validated through statistical analysis of consistency with senior experts.

2. Materials and Methods

In this study, we proposed a method for automatic FMF angle measurement, divided into three steps: segmentation, prediction, and measurement. The method proposed in this study is illustrated in Figure 3. Initially, the segmentation process commences with the utilization of the key component segmentation module, specifically targeting the fetus's palate and frontal bone regions. Then, on the obtained segmentation map, the key point detection module is used to predict the three key points needed to measure the fetal frontal maxillary angle. Finally, the angle is automatically measured using the biometry measurement module.

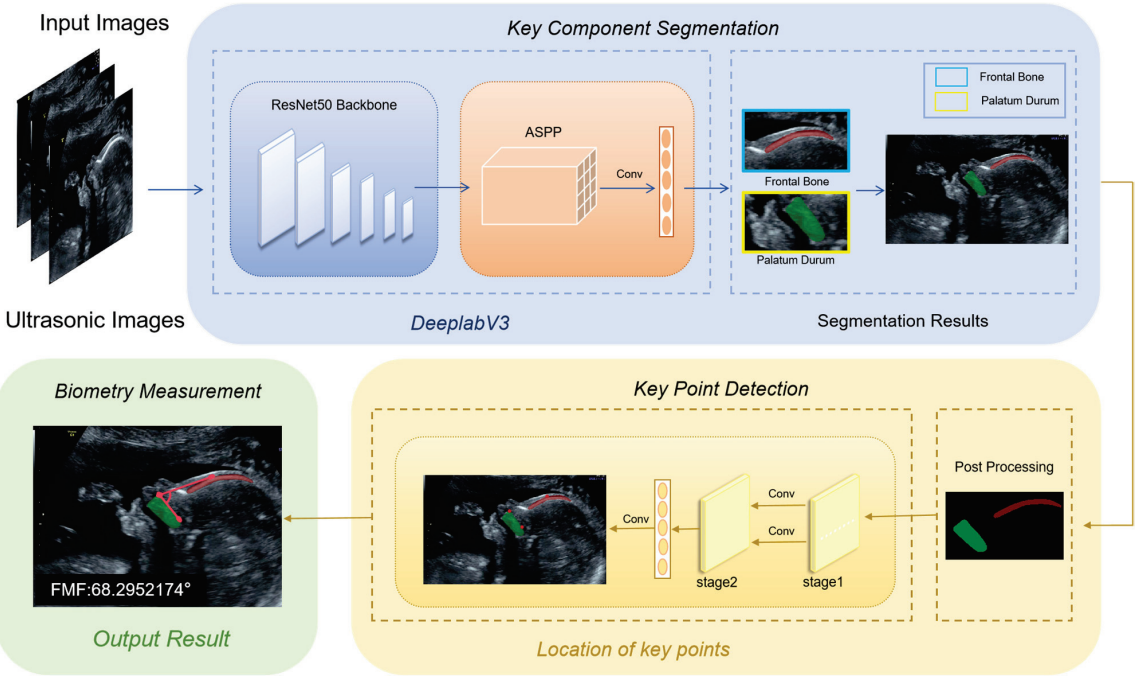


Figure 3. This study proposes a block diagram of the process of automatic measurement of the FMF angle, which demonstrates the measurement process through AI models Firstly, the segmentation is performed by Key Component Segmentation, then the Key Point Dection module is used for key point detection, and finally the angle is calculated using Biometry Measurement.

2.1. Study Population

This retrospective study utilized archived fetal images of singleton pregnancies obtained by mid-trimester ultrasound scanning between February 2020 and October 2022 at the Quanzhou First Hospital, affiliated to Fujian Medical University. For analysis, two-dimensional images depicting the median sagittal plane of the fetal face were chosen, spanning the gestational period between 14 and 27 weeks.

The inclusion criteria were as follows:

- We selected a median sagittal image of the fetal face, encompassing both the entire head and upper thorax, where these two regions account for over 65% of the total image area;
- The fetal face display is complete and clear. The ultrasonic images include the forehead, nasal bone, nasion, palate, mandible, chin, and upper/lower lip;
- The fetal image is not obscured by the umbilical cord or limb.

Individuals with blurred images, rendering fetal facial structures indistinct, were excluded from the study. The data employed in this research were rendered anonymous, ensuring adherence to ethical standards and alleviating any potential concerns.

2.2. Data Annotation

The COCO-Annotator tool can be employed to annotate key points for the key point identification task of the key point detection module. COCO-Annotator, a web-oriented tool tailored specifically for computer vision researchers and practitioners, serves as an efficient means for annotation. Utilizing the Labelme (Version 4.5.9) tool, the senior expert manually annotated key anatomical structures on 1549 fetal ultrasound images at pixel-level precision. This annotation process served as the foundation for executing the semantic segmentation task within the critical component segmentation module. Labelme, an image annotation tool originating from the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology (MIT), is crafted in Python and PyQT. It enables the precise annotation of images in various forms, including polygons, rectangles, circles, polylines, line segments, and dots. Labelme's versatility is evident in its numerous image segmentation tasks.

2.3. Deep Learning Model Architecture

This research aims to devise an automated biometric system, employing a deep learning technique for accurately gauging FMF angles through the utilization of fetal ultrasound imagery. The method proposed in this study is designed to measure the FMF angle. The automated measurement of FMF angle frames is structured into three distinct subtasks, which are visualized in Figure 3. As an initial step, the ultrasound image undergoes semantic segmentation. Subsequently, the segmentation network model undergoes a learning process to extract pertinent features from the image, ultimately identifying two crucial anatomical structures: the frontal bone and the palate. Then, the acquired anatomical structure images are trained a second time using a keypoint detection network to obtain the points needed to measure the angle of FMF, which we proposed according to HRnet [21] to obtain the points required to measure the FMF angle. Finally, the FMF angle is calculated based on the pixel points.

The key component segmentation module employs the DeepLabV3 [22] segmentation network to delineate the key structures of the fetal face, resulting in the classification of three categories: palate, frontal bone, and background. The utilization of deep convolutional neural networks (DCNNs) in semantic segmentation tasks often reduces spatial resolution. Ref. [23] shows that in semantic segmentation, tasks result in a decrease in spatial resolution. This is because deep convolutional neural networks undergo consec-

utive pooling or convolution operations. While these operations increase the receptive field of the DCNNs, they also generate the phenomenon of information loss during the downsampling processes.

To mitigate the challenge of information loss stemming from downsampling, DeepLabV3 introduces atrous convolution, a technique that enlarges the receptive field and facilitates the acquisition of multi-scale contextual data. This approach has been demonstrated to be effective for semantic image segmentation [24]. Atrous convolution, also known as dilation convolution or inflationary convolution, expands the receptive field by adding voids and introduces a new hyperparameter, the dilation rate, to the convolutional layer. The dilation rate dictates the spacing of values during the convolution kernel's data processing. By adjusting this hyperparameter, atrous convolution effectively expands the receptive field to varying degrees, ultimately enhancing the precision of segmentation outcomes. This approach enables capturing multi-scale contextual information, thereby optimizing segmentation accuracy. Figure 4 illustrates the structure of our segmentation network, which comprises an input layer with dimensions of $480 \times 480 \times 3$. The features are extracted using ResNet [25] as the backbone, after which the image is downsampled by a factor of 8. This is followed by an atrous spatial pyramid pooling (ASPP) structure, which applies atrous convolution with different rates on the feature map. The ASPP algorithm employs a different rate of atrous convolution and is an efficient method for capturing multi-scale information. Its structure is shown in Figure 5. This network architecture consists of five parallel branches, encompassing a 1×1 convolutional layer, three 3×3 atrous convolutional layers, and a global average pooling layer. Notably, the dilation rates for the atrous convolutions are set to 12, 24, and 36, respectively, to achieve optimal performance. The global average pooling layer increases the global context information, followed by a 1×1 convolutional layer. Finally, a bilinear interpolation method is employed to incorporate global context information, after which the outputs of the five branches are obtained by converting the five. Subsequently, the outputs from the five branches are concatenated, and a convolutional layer is utilized to integrate the information from each individual branch. This is followed by a reduction to the original size by bilinear interpolation. The cross-entropy loss function is employed to address the loss incurred by the segmentation network. By referencing the weight parameter, we can reduce the weight assigned to the background, thereby improving the issue of class imbalance. The specific formula is shown in (1).

$$Loss = -\sum_{i=1}^{n} \omega \times P(i) \times log_2 Q(i) \quad (1)$$

where i denotes the index of the class, $P(i)$ represents the value of the i-th class in the true label, $Q(i)$ represents the probability predicted by the model for the i-th class, ω is the weight, and n is the number of categories.

In the key point detection module, we predict two crucial points: the upper edge of the palate and the outermost point of maximum anterior deviation on the frontal bone's exterior surface. Algorithms for keypoint detection can be classified according to the output format of the model into regression and heatmap methods. Utilizing the regression approach, the input image is introduced into a convolutional network. Upon processing through a fully connected layer, this network subsequently generates the coordinates of the key points. This method has proven effective in [26,27]. The heatmap-based supervised model [28,29] learns a Gaussian probability distribution map, which renders each point in the ground truth as a Gaussian heatmap. Finally, the network outputs a heatmap of size $W' \times H'$, $\{H_1, H_2, \ldots, H_k\}$, where H_k is the location confidence of the kth keypoint. k feature maps correspond to N keypoints. The HRnet method is a technique for predicting keypoints based on heat maps, it is one of the most popular methods in existing research,

parallel connection resolution from high to low subnetworks can maintain high resolution throughout the training process. It entails the generation of a heat map for each keypoint, with the subsequent selection of the location with the highest heat values as the keypoint. HRNet comprises four modules, with each module comprising a connection part and a core idea of continuously fusing information on different scales.

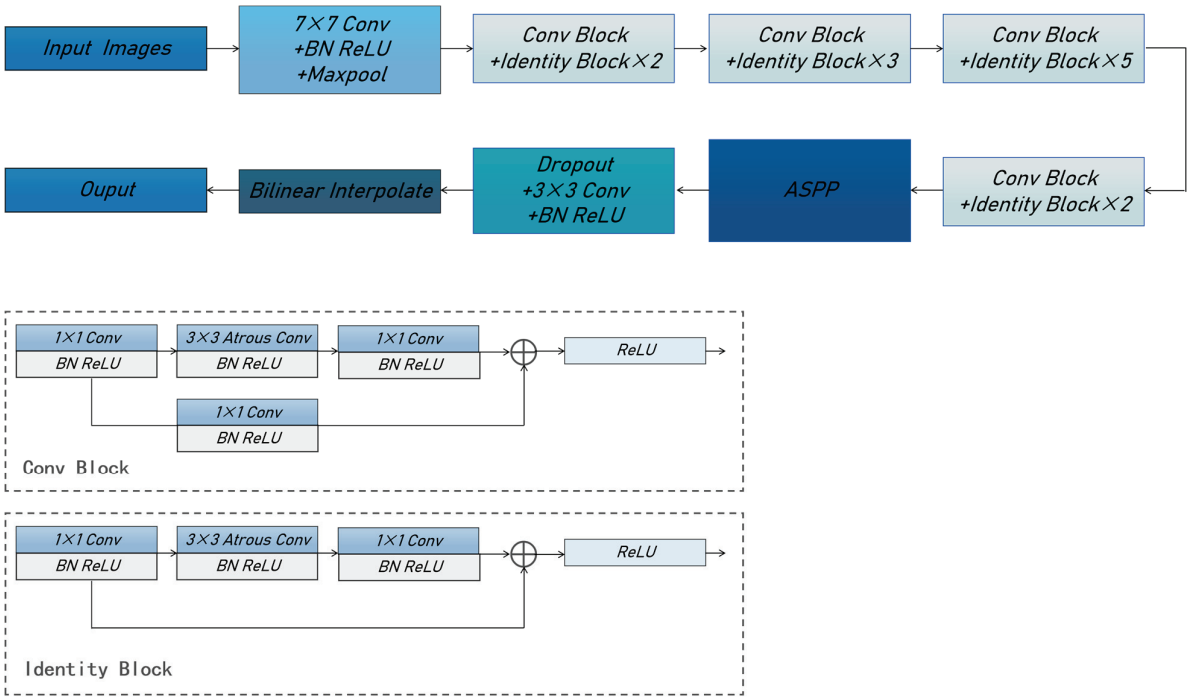


Figure 4. An end-to-end segmentation model is adopted in this study. The architecture comprises our network model with three modules: the identity block, the transformation block, and the atrous spatial pyramid pooling (ASPP) block.

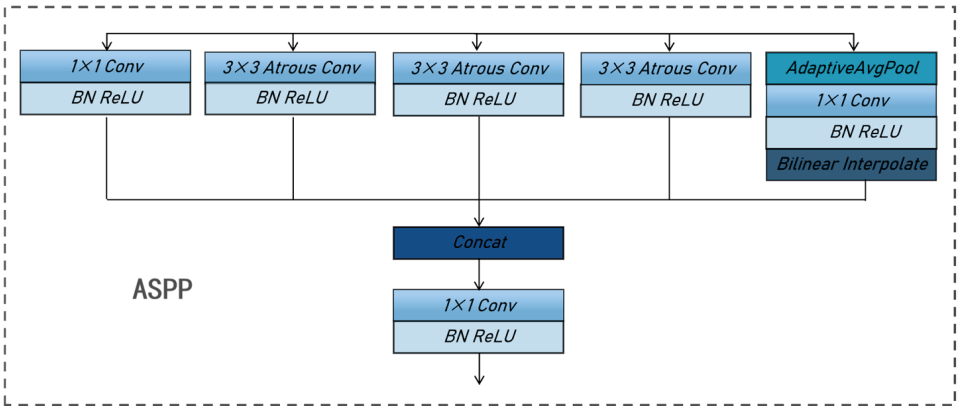


Figure 5. ASPP structure diagram.

As illustrated in Figure 6, we built two parallel sub-network branches based on the idea of fusing different scales. The network was initially downsampled four times through two convolutional layers of size 3 × 3 with a step size of two, then with four bottlenecks. For each scale branch, the stage structure fuses the information on different scales by downsampling and upsampling. To achieve downsampling, a convolutional layer with a 3 × 3 kernel and stride of 2 is employed. Upsampling is accomplished through nearest neighbor interpolation. Subsequently, a 1 × 1 convolutional layer with an applied convolutional kernel produces the feature layer, or heatmap, corresponding

to each key point. The network undergoes training utilizing mean squared error loss, as expressed in (2).

$$Loss = \frac{1}{n}\sum_{i=1}^{n}(P(i) - Q(i))^2 \tag{2}$$

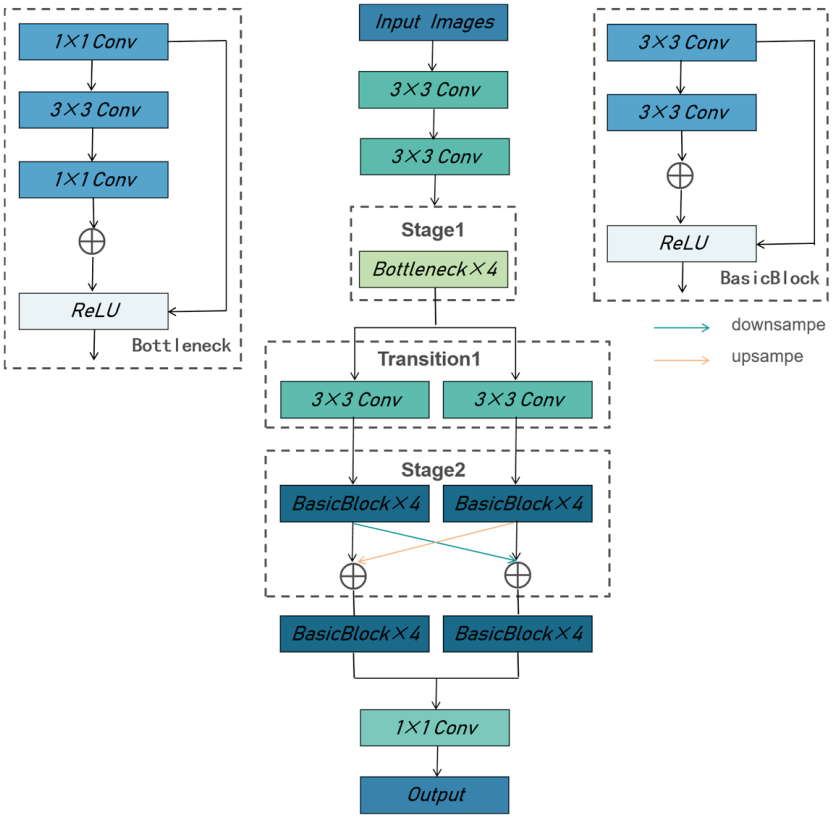


Figure 6. Structure of the keypoint detection network. The green arrow represents downsampling, and the orange arrow represents upsampling.

$Q(i)$ and $P(i)$ represent the model's predicted value and true sample value, respectively, and n is the number of categories.

The third step is the biometry measurement module. This module first obtains the coordinates of the keypoints predicted by the keypoint detection network and forms a triangle through the three keypoints. It then calculates the lengths of the three sides of the triangle using (3), where x_1, x_2, y_1, and y_2 are the horizontal and vertical coordinates of the endpoints of the corresponding line segments, respectively. H is the length of the side of the triangle formed. As depicted in Figure 7, the radian value corresponding to the angle associated with each side of the triangle is subsequently determined using the inverse cosine function. Specifically, the lengths of the three sides of the triangle are represented by the values a, b, and c, while θ signifies the radian value of the desired angle. The specific formula is shown in (4). Finally, the radian values are converted to angle values using the degrees function in the math library. The specific flow is shown in Figure 7.

$$H = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{3}$$

$$\theta = a\cos\frac{b^2 + c^2 - a^2}{2bc} \tag{4}$$

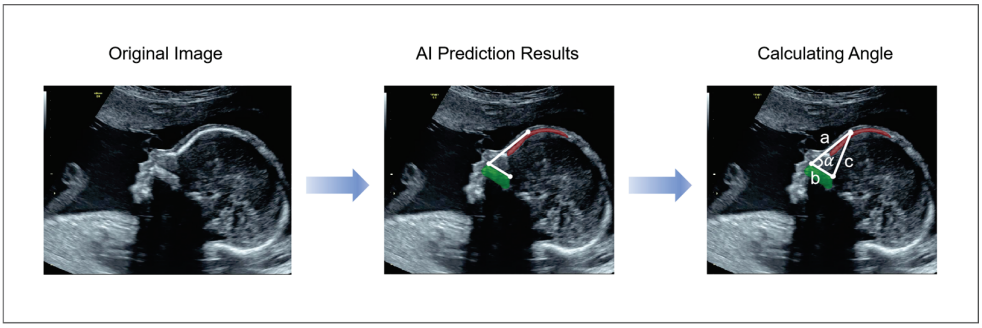


Figure 7. General flowchart of the biometry measurement module. The angle is calculated from the coordinates of the points.

3. Experiments

In accordance with the established criteria for inclusion and exclusion, a total of 1549 mid-trimester fetal ultrasound images were obtained. From this dataset, 116 fetal cases were randomly selected to constitute the validation set, while the remaining dataset was divided into a training set and a test set in a 9:1 ratio. The training set was utilized for the construction and training of the model, and the test set provided an independent evaluation of its performance. The validation set was used to adjust the model hyperparameters and monitor for overfitting. The model was trained on a computer with a 3.8 GHz Intel i7-10700K CPU and an NVIDIA GeForce RTX 2060 GPU, which produced by NVIDIA Corporation (Santa Clara, CA, USA). The model was constructed using the PyTorch (Version 1.10) deep learning framework on the Windows 10 operating system. We empirically determined the hyperparameter settings based on our experiences. The initial learning rate for the semantic segmentation model was set to 2×10^{-3} with a batch size of four and a polynomial learning rate strategy. The input images were resized to 480×480, and the semantic segmentation network was pre-trained using the COCO dataset. During the training phase, data augmentation was employed to randomly scale the images between 0.5 and 2, and flipping and random cropping were performed to ensure the generalization performance of the network. The mean and standard deviation values used for pre-training normalization were [0.163, 0.178, 0.193] and [0.185, 0.199, 0.208], respectively. The SGD optimizer was used with a momentum parameter set to 0.9 and a weight decay of 1×10^{-4}. The network was trained using cross-entropy loss over 200 epochs. For the keypoint detection network, the initial learning rate was set to 3×10^{-3}, with a batch size of 16. Data augmentation techniques, including scaling, rotation, and resizing to a fixed size, were applied to the dataset. The scaling range was between 0.65 and 1.35, and the rotation angle was between -45 degrees and 45 degrees. Image tensors were normalized using specified mean values [1.295, 1.467, 1.659] and standard deviations [0.305, 0.401, 0.466]. The AdamW optimizer was utilized with $\beta 1$ set to 0.9, $\beta 2$ set to 0.999, $\in$ set to 1×10^{-8} and a weight decay of 1×10^{-4}. The network was trained using mean squared error loss over 100 epochs.

3.1. Evaluation of Segmentation

To assess the efficacy of the semantic segmentation model, three pivotal metrics were chosen: pixel accuracy (PA) [30], intersection over union (IOU) [31], and dice coefficient (Dice) [32]. PA captures the proportion of accurately classified pixels among all pixels. IOU measures the overlap between predicted and ground truth pixel values. Meanwhile, Dice serves as a measure of how closely the model's predicted segmentation results align with the actual labels.

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$IOU = \frac{TP}{TP + TN + FN} \tag{6}$$

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{7}$$

In (5)–(7) *TP* signifies the count of instances where the model correctly forecasts a positive outcome as truly positive; *TN* represents the count of cases where the model accurately predicts a negative instance as negative; *FP* denotes the number of occurrences where the model wrongly classifies a negative sample as positive; and *FN* refers to the number of instances where the model incorrectly predicts a positive sample as negative. Table 1 presents the results of our experiments in segmenting specific anatomical structures using DeepLabV3. The model performance was evaluated using IOU, PA, and Dice. Utilizing the DeepLabV3-based segmentation model, our dataset achieved remarkable accuracy. Specifically, for the palate and frontal bone, the precision–recall (PA) scores were 91.6 and 95.1, respectively. Furthermore, the intersection over union (IOU) values were 62.9 and 70.3, while the Dice similarity coefficients were 77.3 and 82.5 for the palate and frontal bone. Table 2 shows the average Dice values for the fivefold cross-validation experiments. The automatic segmentation based on DeepLabV3 achieves high accuracy on cross-validation of our dataset, with average Dice of 77.3 $\pm$ 1.830 and 82.4 $\pm$ 0.923 for the palate and frontal bone, respectively. Figure 8 presents the segmentation outcomes of the DeepLabV3 model, juxtaposed with the ground truth data, for a comparative analysis.

Table 1. Using DeepLabV3 to segment anatomical structure: IOU, PA, Dice.

Class	PA (%)	IOU (%)	Dice (%)
Palate	91.6	62.9	77.3
Frontal bone	95.1	70.3	82.5

Table 2. Five-fold cross validation with Dice.

Class	Fold 1 (%)	Fold 2 (%)	Fold 3 (%)	Fold 4 (%)	Fold 5 (%)	Average (SD) (%)
Palate	78.0	77.4	76.5	78.3	76.2	77.3 $\pm$ 1.830
Frontal bone	82.6	81.8	82.0	84.1	81.5	82.4 $\pm$ 0.923

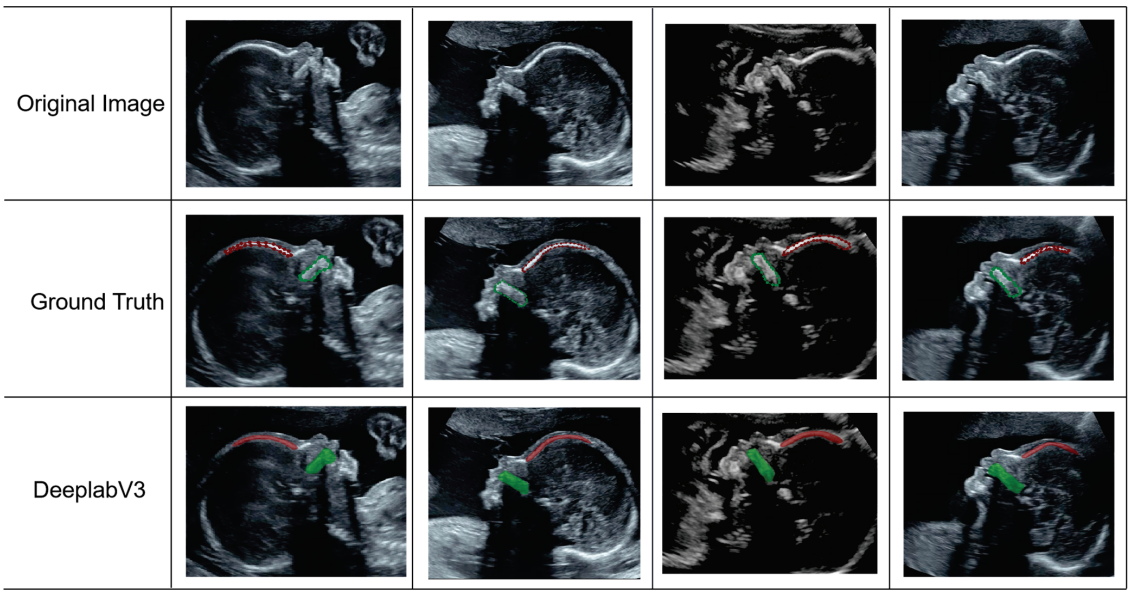


Figure 8. Fetal facial ultrasound images, from top to bottom, the first row is the original image, the second row is the annotated ground truth, and the third row is the segmentation result of DeepLabV3.

3.2. Automatic Biometric Measurements Evaluation

The model is validated using external data. Firstly, the data were tested for normal distribution. If the measurements adhered to a normal distribution, we employed the mean absolute error (MAE), mean relative error (MRE), and standard deviation (SD) as metrics to assess their accuracy. Conversely, in cases where the measurements did not conform to a normal distribution, we represented them using the median and range. The MAE is expressed as the mean of the absolute value of the error. The MRE, standing for mean relative error, represents the ratio of the error to the true value. (8) and (9) illustrate the specific methodology for its calculation. Manual measurements conducted by the senior expert were employed as a reference standard. To evaluate the correlation between manual and automated measurements, we employed Pearson's correlation coefficient as well as the interclass correlation coefficient (ICC) with the "two-way stochastic" and "absolute agreement" variants. Furthermore, we utilized Bland–Altman plots to analyze the agreement between the two methodologies.

A comparison was conducted between the results of the senior expert and the proposed automated measurement method, with the results presented in Table 3. The AI group demonstrated satisfactory performance on the FMF angles measures. The mean absolute error (MAE) and mean relative error (MRE) both exhibited low values, specifically 2.354 and 0.35, respectively, accompanied by a minimal variance. Furthermore, the mean values were very close to the senior expert's. The AI measurements of FMF angles in the clinical validation set were $67.352^\circ \pm 3.877$, while the senior expert's measurements were $67.790^\circ \pm 4.609$. Finally, the mean measurement deviations were $0.686^\circ \pm 2.85$, and the significance level was $p < 0.001$. The comparative analysis of measurements between the junior expert and the senior expert is summarized in Table 4. In the validation set, the measurements of the junior expert were 65.922 ± 4.363, with mean measurement deviations of -1.869 ± 3.967 and MAE and MRE values of 3.428 and 0.050, respectively. The MAE and MRE metrics are higher than those observed between AI and the senior expert, indicating that the proposed method is more accurate than that employed by the junior expert, as illustrated in Figure 9. The distribution of errors is displayed in Figure 10.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i{}' - y_i\right| \tag{8}$$

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i{}' - y_i}{y_i}\right| \tag{9}$$

Table 3. Comparison of our proposed method with the senior expert in the clinical validation set (mean ± std).

Group (n = 116)	FMF Angle (°)
AI group	67.352 ± 3.877
Senior expert	67.790 ± 4.609
Mean of difference	0.686 ± 2.85
MAE	2.354
MRE	0.035
p	<0.001

Table 4. Comparison of the junior expert and the senior expert in the clinical validation set (mean ± std).

Group (n = 116)	FMF Angle (°)
Junior expert	65.922 ± 4.363
Senior expert	67.790 ± 4.609
Mean of difference	−1.869 ± 3.967
MAE	3.428
MRE	0.050
p	<0.001

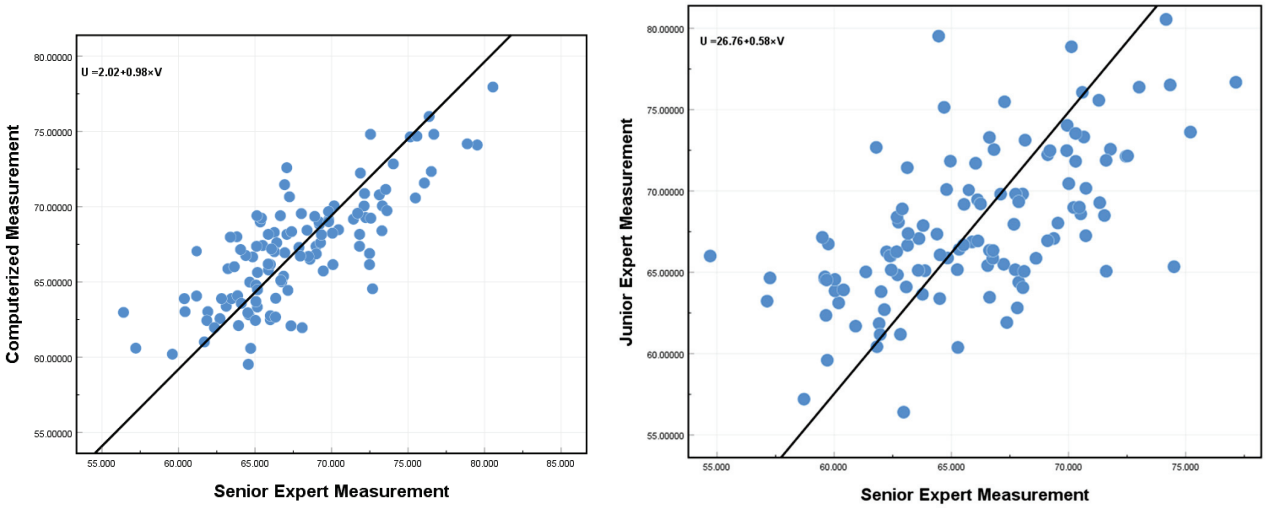


Figure 9. Pearson correlation coefficient plots showing the agreement of the measurements between the senior expert and AI are shown on the left, and the junior expert and the senior expert on the right. In the figure, a circle represents a single fetus, and the straight line is the fitting line of the points.

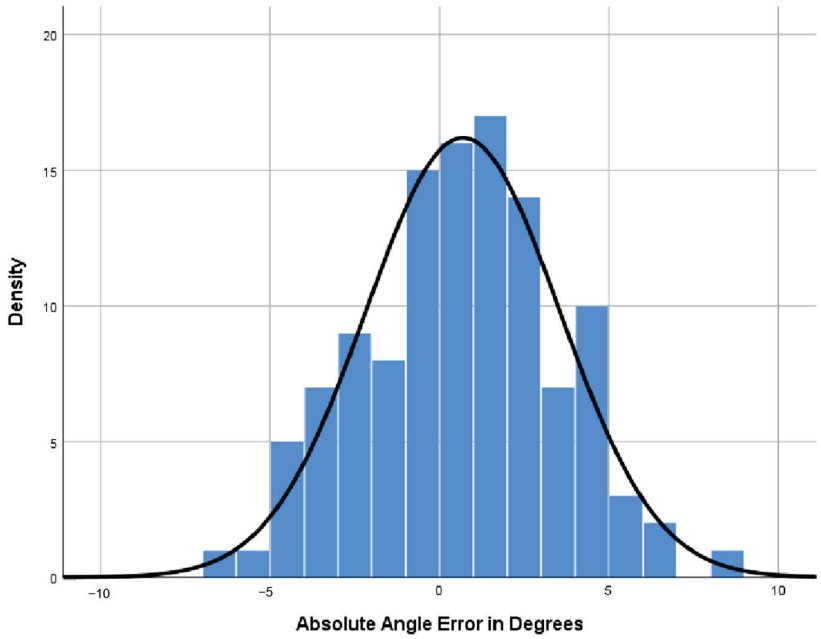


Figure 10. Histogram of the error distribution, The horizontal axis is the difference between the senior expert and AI, and the vertical axis is the frequency of the difference. The plot reveals that the error distribution is approximately centered around zero, with no apparent skewness in the errors.

To further evaluate the extent of correlation and concordance between the senior expert and AI, we undertook a comparative analysis of the FMF angle measurements' ICC reliability indices. Table 5 presents the detailed ICC data. Specifically, the ICC value comparing AI's automated measurements to the senior expert's was 0.760, whereas the ICC comparing junior expert's measurements to the senior expert's stood at 0.562. It can be observed that the junior expert had significantly lower ICC values for the FMF angle measurements than the AI group. It is postulated that the reason for this phenomenon is that speckle noise, boundary-blurring, and complex anatomical structures in ultrasound images influence the judgment of junior physicians. In conclusion, the aforementioned analyses indicate that AI can not only relieve the experienced senior expert of the arduous task of manual measurements, but can also facilitate the advancement of junior experts' diagnostic accuracy.

Table 5. Comparison of the reliability of FMF angle measurements between the junior expert and the senior expert.

Group (n = 116)	FMF Angle (°)	
	ICC	95% CI
AI and senior expert	0.760	0.666–0.829
Junior expert and senior expert	0.562	0.361–0.701

3.3. Bland–Altman Agreement Evaluation

To enhance the visualization of the concordance between the AI, senior expert, and junior expert measurements, we created Bland–Altman plots. These plots feature a solid central line that represents the average deviation between the two measurement sets, and upper and lower dashed lines that mark the 95% agreement intervals, respectively. This visualization approach aids in understanding the level of agreement between the various measurement methods. Figure 11 presents the results of the Bland–Altman plot analysis of the validation set. The AI measurements exhibited a substantial level of concurrence with the senior expert's reading, as evidenced by a difference of 0.686, which falls within a 95% confidence interval, ranging from −4.9133 to 6.286. Additionally, the difference between the junior expert and senior expert measurements was 1.869 (95% confidence interval: −5.906 to 9.644).

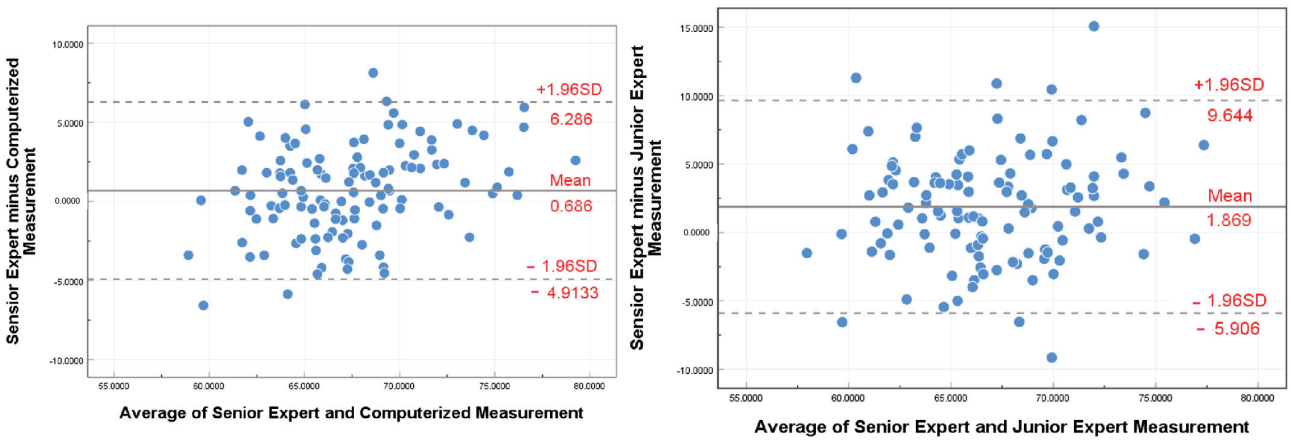


Figure 11. Bland–Altman analysis plot with measurements of the AI versus the senior expert on the left and the junior expert versus senior expert analyzed on the right. The solid line in the figure is the mean value of the difference, and the dotted line is 95% of the confidence interval.

4. Discussion

This study proposes an AI model that employs a deep learning approach to complete the FMF angle measurement task in mid-pregnancy. This is achieved using three modules: key component segmentation, key point detection, and biometry. The modules were designed to achieve the accurate segmentation of key structures, the accurate prediction of key points, and automatic measurement tasks, respectively. The advanced deep learning method was employed to achieve accurate segmentation of two structures, the palate and the frontal bone. The accuracy of the measurement is heavily influenced by the outcome of the segmentation process. Based on the accurate segmentation, three key points were accurately predicted using a keypoint detection algorithm. The findings from our experiments revealed a significant correlation between the proposed model and the assessments provided by senior expert, as illustrated in Figure 9. The Bland–Altman analysis (Figure 11) revealed a high degree of agreement between the computer-predicted and manually measured values for the FMF angle measurements. Figure 12 illustrates the comparison between AI and the senior expert. The results of this study may enhance the efficiency of sonographers and accelerate the process of clinical assessment of the fetus.

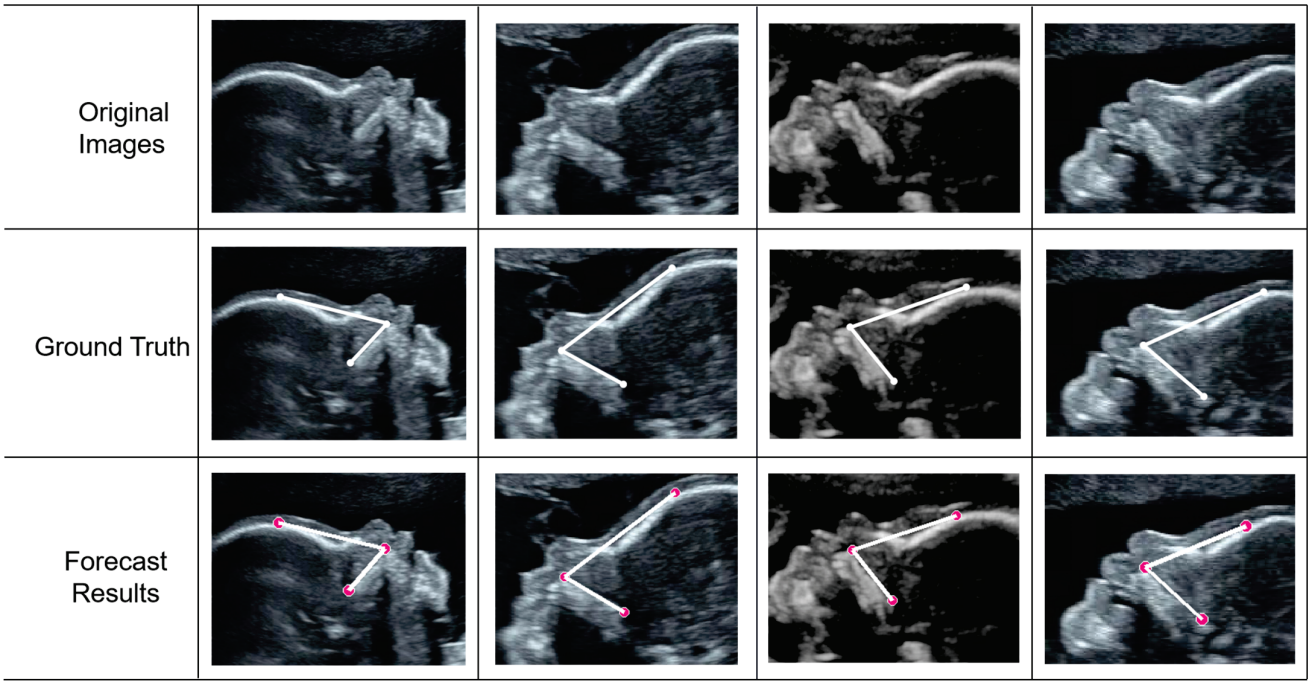


Figure 12. FMF measurement results. From top to bottom, the first row shows the original image, the second row shows the annotated ground truth, and the third row shows the results of the FMF prediction.

In late 2011, the United States introduced noninvasive prenatal testing utilizing massively parallel sequencing of cell-free DNA (cfDNA testing) in maternal plasma into its clinical prenatal care practices. The positive predictive values (PPV) for trisomy 21 were 45.5% (95% CI, 16.7 to 76.6) with cfDNA testing, as reported in [33]. Although it has been demonstrated that trisomy 21 is associated with certain sonographic markers, including increased nuchal fold thickness, a shortened femur, hyperechogenic bowel, or hydronephrosis, these distinctive characteristics have been noted in a limited number of fetuses, as reported in [34]. In contrast, using ultrasound to evaluate the FMF angle represents a highly sensitive approach [6]. The automation of biometry using AI has been demonstrated in numerous studies [35]. Automated biometric systems have the capability to reduce the impact of both intra- and inter-operator variability, thus improving the precision of diagnostic

assessments. In this study, an automated FMF angle measurement model was constructed for the first time, which reduces the workload of ultrasonographers and simultaneously facilitates the development of an intelligent healthcare system.

This study is the first to propose the use of deep learning for automatic FMF angle measurement, which includes 1549 normal fetal images for model development and validation. In clinical applications, traditional manual measurement methods necessitate the involvement of a trained sonographer. This process is often challenging and time-consuming. The method improves the interpretability of the diagnostic process by automatically segmenting key structures and locating landmarks. It should be noted that this study has certain limitations. First of all, the ultrasound images used for training and validation were obtained from normal fetuses, and the applicability of the method to abnormal fetuses has not been tested. Secondly, this study is retrospective in nature. To ensure the veracity of AI models, it is advisable to conduct further prospective investigations in the future. Another significant challenge pertains to device variability, as ultrasound machines from different manufacturers may produce images with varying resolutions and qualities. This has the potential to impact the efficacy of the deep learning models, which were trained on a particular dataset of ultrasound images. To address this challenge, future research should involve testing the framework on images acquired from a diverse range of ultrasound devices to ensure its robustness and generalizability. Furthermore, consideration must be given to the real-time capability of the measurement methodology. The incorporation of deep learning models into ultrasound machines is a potential solution to this issue, as it would enable the real-time processing and analysis of images. In the domain of deep learning, the transferability of a model across different datasets is a pivotal consideration, as it determines the model's performance on unseen data. The data presented herein were obtained from two distinct ultrasound instruments, namely the Philips EPlQ5 ultrasound machine and the GE Voluson E8 ultrasound machine. These two pieces of equipment originate from Philips (Amsterdam, The Netherlands) and General Electric Company (Boston, MA, USA), respectively. The dataset under consideration is diverse, and the model was trained using a training set comprising multiple data types. This approach facilitates the model's learning of richer feature representations, thereby enhancing its transferability. It has been observed that certain multi-adversarial learning methods have shown considerable potential in enhancing the generalization ability of the model. Ref. [36] proposed a Multi-Adversarial Open-Set Domain Adaptation Network (MAOSDAN), which effectively addressed the problem of partial overlap between the label space of the source domain and the target domain. Although these methods have not yet been included in our research, they provide new concepts for future research work. In subsequent studies, we aim to integrate these techniques into our model to enhance its performance and generalization capability.

5. Conclusions

This study proposes a deep learning-based artificial intelligence-assisted measurement method for the fast and accurate automatic measurement of FMF angles. This method comprises three steps: segmentation, keypoint detection, and measurement. It is effective for measuring FMF angles in clinical work. The method proposed in this study reduces the subjectivity of sonographers in the diagnostic process and improves the efficiency and accuracy of the diagnosis.

Author Contributions: Conceptualization, Z.L.; methodology, J.W.; software, J.W.; validation, G.L.; formal analysis, H.S.; investigation, W.Y.; resources, P.L.; data curation, Y.F.; writing—original draft preparation, Y.W.; writing—review and editing, J.W.; supervision, J.W.; project administration, G.L.; funding acquisition, Z.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 Fujian grant number (2021J011404, 2023J011773, 2023J011784), and by the Quanzhou scientific and technological planning projects grant number (2022NS057, 2023NS055).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the School of Medicine, Huaqiao University (protocol code: M2024006; date: May 2024).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Antonarakis, S.E.; Skotko, B.G.; Rafii, M.S.; Strydom, A.; Pape, S.E.; Bianchi, D.W.; Sherman, S.L.; Reeves, R.H. Down syndrome. *Nat. Rev. Dis. Primers* **2020**, *6*, 9. [CrossRef] [PubMed]
2. Down, J.L. Observations on an ethnic classification of idiots. 1866. *Ment. Retard.* **1995**, *33*, 54–56. [PubMed]
3. Bhide, A.; Acharya, G.; Bilardo, C.M.; Brezinka, C.; Cafici, D.; Hernandez-Andrade, E.; Kalache, K.; Kingdom, J.; Kiserud, T.; Lee, W.; et al. ISUOG practice guidelines: Use of Doppler ultrasonography in obstetrics. *Ultrasound Obstet. Gynecol.* **2013**, *41*, 233–239. [CrossRef]
4. Zaffino, P.; Moccia, S.; De Momi, E.; Spadea, M.F. A Review on Advances in Intra-operative Imaging for Surgery and Therapy: Imagining the Operating Room of the Future. *Ann. Biomed. Eng.* **2020**, *48*, 2171–2191. [CrossRef]
5. Sonek, J.; Borenstein, M.; Downing, C.; McKenna, D.; Neiger, R.; Croom, C.; Genrich, T.; Nicolaides, K.H. Frontomaxillary facial angles in screening for trisomy 21 at 14–23 weeks' gestation. *Am. J. Obstet. Gynecol.* **2007**, *197*, e161–e165. [CrossRef]
6. Bergann, A.; Bamberg, C.; Eder, K.; Proquitté, H.; Hartung, J.P.; Bollmann, R.; Kalache, K.D. Mid-facial anthropometry in second-trimester fetuses with trisomy 21: A three-dimensional ultrasound study. *Prenat. Diagn.* **2006**, *26*, 158–162. [CrossRef]
7. Molina, F.; Persico, N.; Borenstein, M.; Sonek, J.; Nicolaides, K.H. Frontomaxillary facial angle in trisomy 21 fetuses at 16–24 weeks of gestation. *Ultrasound Obstet. Gynecol.* **2008**, *31*, 384–387. [CrossRef]
8. Zheng, J.; Yuan, S.; Li, W.; Fu, H.; Yu, L.; Huang, J. A Review of Individual Tree Crown Detection and Delineation From Optical Remote Sensing Images: Current progress and future. *IEEE Geosci. Remote Sens. Mag.* 2024, *early access*. [CrossRef]
9. Lakshmi, P.; Geetha, M.; Menon, N.R.; Krishnan, V.; Nedungadi, P. Automated Screening for Trisomy 21 by measuring Nuchal Translucency and Frontomaxillary Facial Angle. In Proceedings of the 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 19–22 September 2018; pp. 1738–1743. [CrossRef]
10. Chen, X.; He, M.; Dan, T.; Wang, N.; Lin, M.; Zhang, L.; Xian, J.; Cai, H.; Xie, H. Automatic Measurements of Fetal Lateral Ventricles in 2D Ultrasound Images Using Deep Learning. *Front. Neurol.* **2020**, *11*, 526. [CrossRef]
11. Tran, V.L.; Lin, H.Y.; Liu, H.W. Multitask Deep Learning for Segmentation and Lumbosacral Spine Inspection. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 4005910. [CrossRef]
12. Cho, H.C.; Sun, S.; Min Hyun, C.; Kwon, J.Y.; Kim, B.; Park, Y.; Seo, J.K. Automated ultrasound assessment of amniotic fluid index using deep learning. *Med. Image Anal.* **2021**, *69*, 101951. [CrossRef] [PubMed]
13. Jang, J.; Park, Y.; Kim, B.; Lee, S.M.; Kwon, J.Y.; Seo, J.K. Automatic Estimation of Fetal Abdominal Circumference From Ultrasound Images. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1512–1520. [CrossRef] [PubMed]
14. Pluym, I.D.; Afshar, Y.; Holliman, K.; Kwan, L.; Bolagani, A.; Mok, T.; Silver, B.; Ramirez, E.; Han, C.S.; Platt, L.D. Accuracy of automated three-dimensional ultrasound imaging technique for fetal head biometry. *Ultrasound Obstet. Gynecol.* **2021**, *57*, 798–803. [CrossRef] [PubMed]
15. Li, P.; Zhao, H.; Liu, P.; Cao, F. Automated measurement network for accurate segmentation and parameter modification in fetal head ultrasound images. *Med. Biol. Eng. Comput.* **2020**, *58*, 2879–2892. [CrossRef] [PubMed]
16. Al Qurri, A.; Almekkawy, M. Improved UNet with Attention for Medical Image Segmentation. *Sensors* **2023**, *23*, 8589. [CrossRef]
17. Ibrahim, N.M.; Alanize, H.; Alqahtani, L.; Alqahtani, L.J.; Alabssi, R.; Alsindi, W.; Alabssi, H.; AlMuhanna, A.; Althani, H. Deep Learning Approaches for the Assessment of Germinal Matrix Hemorrhage Using Neonatal Head Ultrasound. *Sensors* **2024**, *24*, 7052. [CrossRef]
18. Anand, A.; Jung, S.; Lee, S. Breast Lesion Detection for Ultrasound Images Using MaskFormer. *Sensors* **2024**, *24*, 6890. [CrossRef]
19. Alruily, M.; Said, W.; Mostafa, A.M.; Ezz, M.; Elmezain, M. Breast Ultrasound Images Augmentation and Segmentation Using GAN with Identity Block and Modified U-Net 3+. *Sensors* **2023**, *23*, 8599. [CrossRef]

20. Tang, J.; Han, J.; Xie, B.; Xue, J.; Zhou, H.; Jiang, Y.; Hu, L.; Chen, C.; Zhang, K.; Zhu, F.; et al. The Two-Stage Ensemble Learning Model Based on Aggregated Facial Features in Screening for Fetal Genetic Diseases. *Int. J. Environ. Res. Public Health* **2023**, *20*, 2377. [CrossRef]
21. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 3349–3364. [CrossRef]
22. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.
23. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
24. Wang, H.; Sun, K.; Wang, Y. Zebra crossing segmentation based on dilated convolutions. In Proceedings of the 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), Changsha, China, 26–28 March 2021; pp. 522–527. [CrossRef]
25. 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. [CrossRef]
26. Carreira, J.; Agrawal, P.; Fragkiadaki, K.; Malik, J. Human Pose Estimation with Iterative Error Feedback. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 4733–4742.
27. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]
28. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 172–186. [CrossRef]
29. Yang, Z.; Zeng, A.; Yuan, C.; Li, Y. Effective Whole-body Pose Estimation with Two-stages Distillation. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Paris, France, 2–6 October 2023; pp. 4212–4222. [CrossRef]
30. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 3523–3542. [CrossRef]
31. Rahman, M.A.; Wang, Y. Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation. In Proceedings of the International Symposium on Visual Computing, Virtual, 12–14 December 2016.
32. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*, 29. [CrossRef]
33. Bianchi, D.W.; Parker, R.L.; Wentworth, J.; Madankumar, R.; Saffer, C.; Das, A.F.; Craig, J.A.; Chudova, D.I.; Devers, P.L.; Jones, K.W.; et al. DNA sequencing versus standard prenatal aneuploidy screening. *N. Engl. J. Med.* **2014**, *370*, 799–808. [CrossRef]
34. Smith-Bindman, R.; Hosmer, W.; Feldstein, V.A.; Deeks, J.J.; Goldberg, J.D. Second-trimester ultrasound to detect fetuses with Down syndrome: A meta-analysis. *JAMA* **2001**, *285*, 1044–1055. [CrossRef]
35. Chen, Z.; Liu, Z.; Du, M.; Wang, Z. Artificial Intelligence in Obstetric Ultrasound: An Update and Future Applications. *Front. Med.* **2021**, *8*, 733468. [CrossRef]
36. Zheng, J.; Wen, Y.; Chen, M.; Yuan, S.; Li, W.; Zhao, Y.; Wu, W.; Zhang, L.; Dong, R.; Fu, H. Open-set domain adaptation for scene classification using multi-adversarial learning. *ISPRS J. Photogramm. Remote Sens.* **2024**, *208*, 245–260. [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.

sensors

MDPI

Article

Seed Protein Content Estimation with Bench-Top Hyperspectral Imaging and Attentive Convolutional Neural Network Models

Imran Said [1], Vasit Sagan [1,2,3,*], Kyle T. Peterson [4], Haireti Alifu [2,3], Abuduwanli Maiwulanjiang [2], Abby Stylianou [1], Omar Al Akkad [2,3], Supria Sarkar [2] and Noor Al Shakarji [2]

[1] Department of Computer Science, Saint Louis University, Saint Louis, MO 63104, USA; imran.said@slu.edu (I.S.)
[2] Department of Earth, Environment and Geospatial Sciences, Saint Louis University, Saint Louis, MO 63108, USA; omar.alakkad@taylorgeospatial.org (O.A.A.); supria.sarkar@slu.edu (S.S.)
[3] Taylor Geospatial Institute, Saint Louis, MO 63108, USA
[4] Bayer Crop Science, 800 N Lindbergh Blvd, Creve Coeur, MO 63141, USA
* Correspondence: vasit.sagan@slu.edu

Abstract: Wheat is a globally cultivated cereal crop with substantial protein content present in its seeds. This research aimed to develop robust methods for predicting seed protein concentration in wheat seeds using bench-top hyperspectral imaging in the visible, near-infrared (VNIR), and shortwave infrared (SWIR) regions. To fully utilize the spectral and texture features of the full VNIR and SWIR spectral domains, a computer-vision-aided image co-registration methodology was implemented to seamlessly align the VNIR and SWIR bands. Sensitivity analyses were also conducted to identify the most sensitive bands for seed protein estimation. Convolutional neural networks (CNNs) with attention mechanisms were proposed along with traditional machine learning models based on feature engineering including Random Forest (RF) and Support Vector Machine (SVM) regression for comparative analysis. Additionally, the CNN classification approach was used to estimate low, medium, and high protein concentrations because this type of classification is more applicable for breeding efforts. Our results showed that the proposed CNN with attention mechanisms predicted wheat protein content with R^2 values of 0.70 and 0.65 for ventral and dorsal seed orientations, respectively. Although, the R^2 of the CNN approach was lower than of the best performing feature-based method, RF (R^2 of 0.77), end-to-end prediction capabilities with CNN hold great promise for the automation of wheat protein estimation for breeding. The CNN model achieved better classification of protein concentrations between low, medium, and high protein contents, with an R^2 of 0.82. This study's findings highlight the significant potential of hyperspectral imaging and machine learning techniques for advancing precision breeding practices, optimizing seed sorting processes, and enabling targeted agricultural input applications.

Keywords: hyperspectral imaging; seed composition estimation; machine learning; 3D CNN modeling; attentive models

1. Introduction

Wheat (*Triticum aestivum* L.) is a cereal which is cultivated extensively across the world, with millions of tons of grains being harvested annually, and serves as a staple for many communities worldwide [1]. A significant portion of the total mass of wheat seeds consists of proteins, which serve several different functions. Proteins in durum wheat are not sufficient as a primary provider of essential amino acids [2], mostly due to the

less anabolic nature of wheat proteins compared to animal proteins due to their lower comparative digestibility [3]. This low digestibility is compounded further by cooking processes such as baking [4]. However, when consumed alongside [3] or infused with other non-plant-based proteins [2], wheat products have been found to complement these primary protein sources well. Additionally, processes which have been introduced to solve their lower digestibility have made wheat proteins a great source of protein for senior consumers [5]. The proteins in wheat seeds also serve another crucial role: structural. In particular, the gluten proteins, which make up 70–80% of total grain proteins, form a mesh in which starch granules are deposited [6] and are foundational to the distinctive physical properties of dough and the bread-making quality of wheat. This invariably means that the gluten proteins have an influence on the technological performance of different wheat varieties in food production and on the final price of wheat [7]. This calls for the initiation and encouragement of processes that allow analysis and documentation of the qualitative and quantitative characteristics of protein in wheat seed kernels. Furthermore, these processes need to be non-destructive and efficient to accelerate selection processes in breeding programs, while also preserving valuable genetic material for subsequent breeding cycles.

Hyperspectral imaging (HSI) plays an important role in estimating seed composition to automate the process, as the distinct chemical composition of the sample contributes to the relative intensity in absorption, reflection, and transmission of incident radiation [8,9]. This can accelerate seed inspection and breeding efforts as chemical analysis methods often require large, ground samples and involve destructive analysis procedures that can be costly and time-consuming. HSI entails the use of specialized sensors to scan subjects across specific continuous wavelengths of the electromagnetic spectrum, creating three-dimensional datasets composed of two spatial dimensions, and a spectral dimension representing different reflectance values for each pixel of the image [10,11].

HSI has been used to predict seed composition traits of individual kernels, including fatty acids [12], oil [13], moisture [14], and proteins [15,16], enabling the rapid screening of large numbers of samples for the selection of seeds with desirable composition traits [11,13]. Protein content prediction using NIR and HSI for wheat seeds has been reported by several publications with spectral features extracted from imagery [11,17–19]. To the best of our knowledge, there have been no studies that (1) utilized both VNIR and SWIR images as direct input for end-to-end seed composition estimation and (2) investigated the contribution of ventral and dorsal seed orientations for wheat seed protein determination.

This can be attributed to previous studies showing that the NIR region contains the most informative wavelengths for protein content determination, with one study establishing a high vs. low protein content discriminatory technique based on NIR [20], whereas [21] applied PLSR on NIR data with different scatter correction techniques to predict protein content in wheat seeds. Among these studies, Caporaso et al. [11,15] found the wavelength of 1918 nm to be most informative, followed by 2008 nm, 2062 nm, and 2272 nm. Delwiche et al. [22] found wavelengths 1100 nm and 1398 nm to produce the lowest overall model error, while Liu et al. [23] identified five wavelengths between 1980 nm and 2350 nm that correlate strongest with gluten proteins. This region covers numerous overtones and combination bands of molecular vibrations, particularly those in the amide functional groups (N-H and C=O bonds), the chemical components that form the building blocks of proteins and which absorb more NIR light [8,24]. The first overtone of N-H stretching modes typically occur in the 1400–1600 nm range, while the combination bands involving N-H and C=O groups typically occur within the 2000–2300 nm range [25].

In this paper, we present a new approach to protein content estimation that combines attentive vision models and hyperspectral images of wheat seed scans. We also

present the calibration and data preprocessing methods necessary to make this kind of modeling possible.

The objectives of this paper are as follows:

1. Investigating the potential of hyperspectral imaging to predict protein content values for wheat seeds;
2. Classifying wheat seeds according to protein content;
3. Determining optimal wavelengths that are the most informative to the prediction of protein content in seeds;
4. Determining the effect of the orientation of wheat kernels on the predictive ability of machine learning models.

2. Materials and Methods

2.1. Wheat Seed Sample Description

The wheat seed samples were provided in two batches by the Crop Science division of Bayer (Creve Coeur, MO, USA). The first batch consisted of individual seeds sorted according to shared genotypical traits. In total, we had 9 genotype classes, whose frequencies are listed in Table 1. To allow for a parallel comparison of ventral and dorsal scans of each seed sample, they were retained separately in a tray with fixed labels before and after scanning.

Table 1. Number of instances for each genotype class.

Class	B1	B9	M	MT	N	NT	W	WB	WBT
Frequency	384	94	207	422	373	118	304	239	257

The second batch of seeds was grouped in samples based on their protein content obtained through NIR spectroscopy scans. These were later divided into smaller subsamples for a total of 188 subsamples, with at least 60 seeds per subsample, which were scanned separately, and then protein content values for each subsample were obtained separately.

2.2. HySpex Hyperspectral Imaging System

The scans for the seeds were acquired using two hyperspectral imaging sensors of the HySpex brand manufactured by Norsk Elektro Optikk (Oslo, Norway), mounted on an indoor scanning stage as illustrated in Figure 1. The first system, the HySpex VNIR-3000 N [26], has a pixel size of 3.45 μm, which ensures that even narrow band features are resolved with great precision. It also features 3000 spatial pixels and 300 bands covering wavelengths from 400 nm to 1000 nm. The second imaging system, the HySpex SWIR-640 [27], consists of a Mercury Cadmium Telluride (MCT) sensor, which offers a spatial resolution of 32 μm and features a cryogenic cooler that cools the focal plane array to 150 K to minimize background noise and offer exceptional signal-to-noise ratio levels. The SWIR sensor has 362 bands from wavelengths of 960 nm to 2500 nm. The spectral resolutions of the VNIR and SWIR regions are 1.2 nm and 1.7 nm, respectively.

For both systems, the source of illumination of the samples consisted of two custom-made lamps which were manually adjusted to focus the illumination to a line overlapping with each camera's field of view [28]. They cover a broad spectral range from 400 nm to 2500 nm, which sufficiently covers the spectral range of both cameras, and are powered by a 12 V DC supply. The scanners capture the image line by line (line-scanning) while the specimen on the stage is passed across their fields of view in a technique known as pushbroom scanning. A 50% spectralon reflectance panel was employed for calibration.

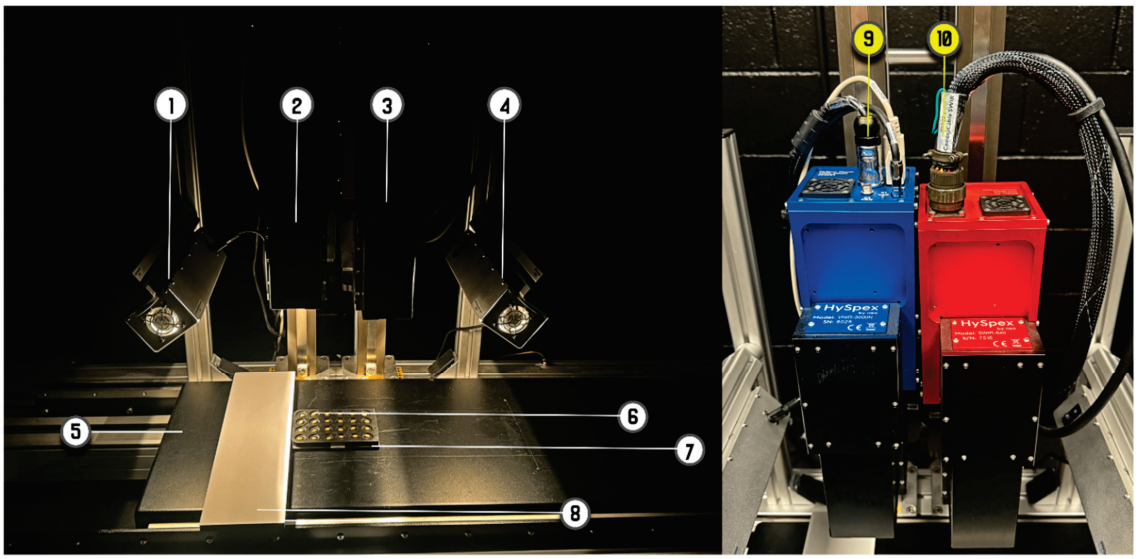


Figure 1. HySpex bench-top scanning machine used in this study. Components as labeled in order: (1) lamp serving as illumination source for VNIR scanner, (2) HySpex VNIR-3000N sensor, (3) HySpex SWIR-640 sensor, (4) SWIR lamp, (5) scanning stage, (6) wheat seed sample, (7) tray holding the seed samples, (8) reflectance panel, (9) VNIR power source, and (10) SWIR power source.

The combination of mechanisms on the imaging system to reduce signal-to-noise ratio levels ensures that less time is spent in calibrating the images during the post-processing stage. The high spatial resolutions of the scanners ensures that we capture as much textural detail as possible, while the precise spectral resolution and broad wavelength coverage allow for the accurate detection of spectral signatures associated with protein and other seed components, enabling more comprehensive chemical and physical analyses.

2.3. Hyperspectral Scanning

The seeds were scanned in two batches. The first batch was scanned over a period of two months, between May and June of 2023, and involved around 2398 samples. Single seeds were placed on circular troughs on the scanning plate, spanning 4 wide and 6 lengthwise to make a total of 24 seeds scanned at a time. The seeds were all scanned with the crease facing upwards (ventral view) and then a second time with the crease facing downwards (dorsal view). The seeds were then immediately replaced in their storage trays and the lamps were turned off once the scanning was completed to prevent excessive moisture loss. The second batch of seeds was scanned in November 2023 over a period of three weeks and involved around 11,000 samples. This time, the seeds were placed randomly on the scanning stage, with heterogeneous orientations of the seeds available with each scan.

Given that there is no commercially available software to co-register VNIR and SWIR spectral datasets, we implemented a novel approach by introducing fiducial markers within the cameras' field of view. These tie pointswere placed on a specialized paper substrate and served as invariant reference points across both spectral ranges. Our team developed a proprietary co-registration algorithm that leverages these fiducial markers to achieve high-precision spatial alignment between the VNIR and SWIR hyperspectral data cubes.

2.4. Reference Data Collection

The protein content of the seeds was obtained after scanning of the images via Dumas combustion chemical analysis. Dumas has been used in previous studies to obtain nitrogen [29] and protein content [30,31] in cereals and other foodstuffs and has been documented as outperforming other chemical analysis methods for most applications [32,33].

The process is based on the concept of the extraction of nitrogen oxides from the materials of study through rapid combustion of the material in a high-heat chamber, followed by a reduction of the nitrogen oxides to nitrogen gas (N2). The final step involves the quantification of N2 using thermal conductivity.

From the chemical analysis, we were able to obtain percentages of protein mass against total seed mass with a range of 11.96%, with a minimum of 7.69% and a maximum value of 19.65%. The rest of the statistics for each orientation are summarized in Table 2.

Table 2. Ground-truth statistics.

Orientation	Central Tendency			Spread				Position		Shape	
	Mean	Median	Mode	Range	Variance	Standard Deviation	IQR	Q1	Q3	Skewness	Kurtosis
Bottom	13.22978	12.81	10.44	11.96	9.586461	3.094586	5.36	10.63	15.99	0.200924	−0.95656
Top	13.48322	13.06	10.44	11.96	8.773358	2.961986	5.12	10.99	16.11	0.212506	−0.95651

We defined protein content classification labels for each of the seeds according to the amount of protein content, with anything below 10% labeled as low, 10–14% as medium, and above 14% as high.

2.5. Workflow

An overview of the methods applied within our study is presented in this section and summarized in Figure 2. The three main stages of the processes include the data collection and processing step, followed by the feature extraction and image preprocessing stage, which is further divided into feature-based approaches and image-based approaches, and finally the modeling and feature evaluation stage.

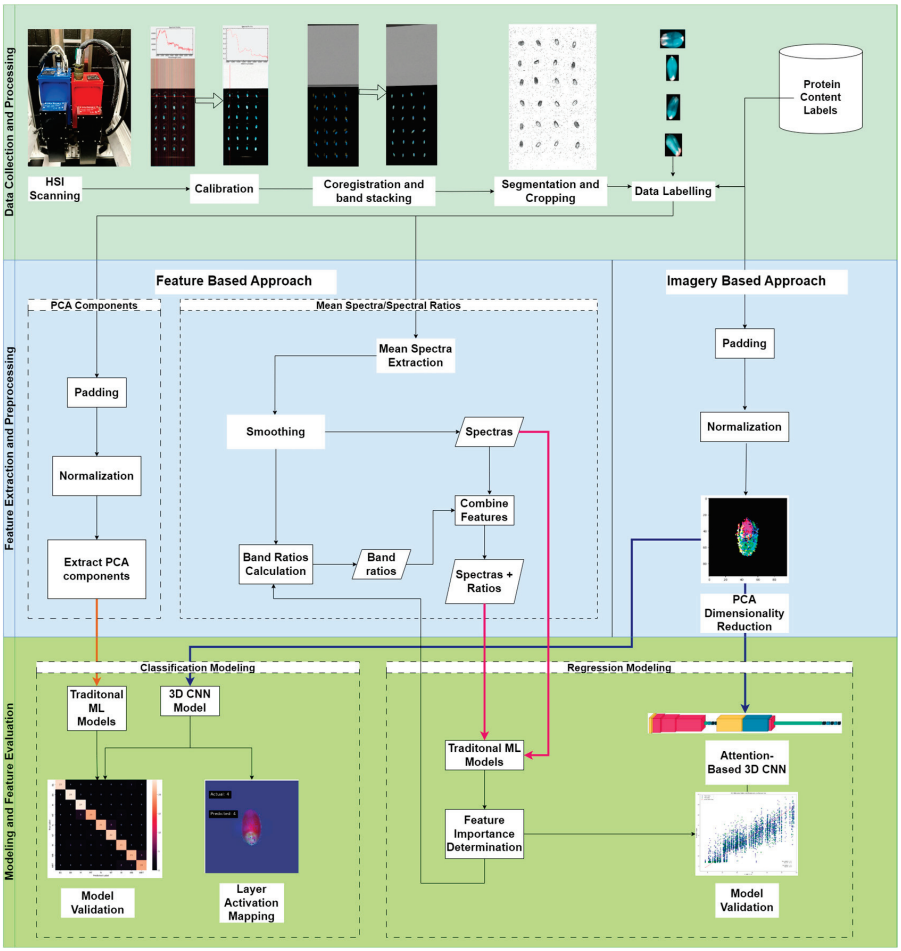


Figure 2. Overview of the workflow for our pipelines from data collection all the way to model validation.

2.6. Radiometric Calibration

Wheat seeds have intrinsically anisotropic reflection properties caused by their uneven surfaces and variable color depth characteristics, which can significantly influence the spectral information captured by hyperspectral imaging systems by causing drastic variability in the data, depending on incident light and sensor viewing angles. The size of a single wheat kernel (about 3 mm along/1.25 mm across) is often smaller than the irradiated diameter (spot size) of the sensor, causing a significant presence of scattering of the reflected signal from the edges of the kernel [34] which contributes to the overall deterioration of signal quality as demonstrated in the pre-calibration spectral profile of the seed in Figure 3. This presents a challenge for fidelity of the data and necessitates the application of reflectance correction methodologies to normalize the hyperspectral data and therefore ensure that it accurately represents the seeds' true spectral properties across different samples and scanning sessions. For hyperspectral analysis of wheat seeds, the fidelity of the data has a direct influence on the reliable estimation of biophysical properties and can prevent inconsistencies in comparative studies across different datasets.

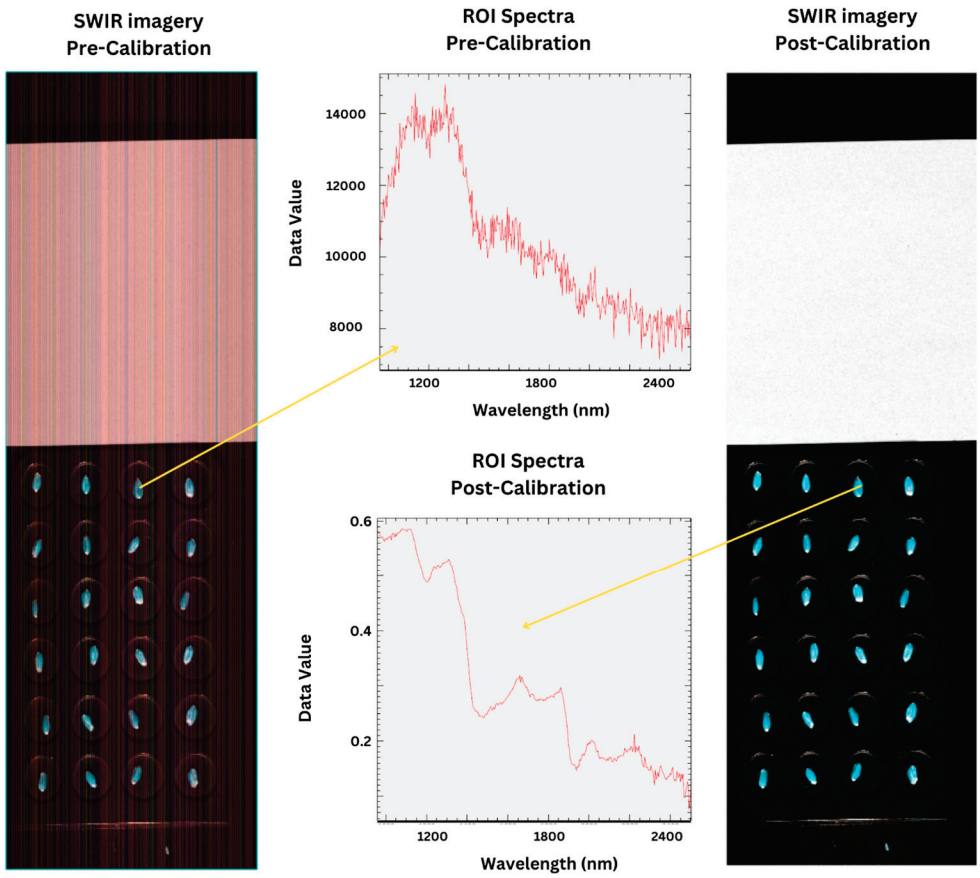


Figure 3. A comparison of the imagery and spectral profiles of the scan scenes (confined to SWIR scans) to highlight the effect of radiometric calibration.

Previous studies have successfully managed to address the reflectance correction challenge of hyperspectral imagery through various methods, including absolute radiometric correction [35], the empirical line method [36], flat field correction [37,38], and Bidirectional Reflectance Distribution Function (BRDF) correction [39,40], among others. Our methodology adopts a reflectance correction approach tailored to bench-top hyperspectral scanning systems and is facilitated by the standard calibration procedure provided by the instrument manufacturer, Norsk Elektro Optikk (NEO). This calibration workflow consisted of the following steps [41]:

1. Background signal subtraction: The sensor's dark current response (*Sbg*), measured when no light is supplied to the sensor, is subtracted from the raw signal (*Sraw*) to correct for background electronic noise. For pixel (i, j) this is summarized as

$$S_{i,j} = S_{\text{raw},i,j} - S_{\text{bg},i,j} \tag{1}$$

2. Radiometric response: A bright scan using the reflectance panel referenced above, with a known reflectance value L_{ref}, is used to obtain a signal $S_{bright,i,j}$ which is then used to calculate the radiometric response $R_{i,j}$ using the following formula:

$$R_{i,j} = \frac{S_{\text{bright},i,j} - S_{\text{bg},i,j}}{L_{ref}} \tag{2}$$

3. Radiometric calibration: The background-subtracted signal is converted to spectral radiance L in units $(\text{W/sr}\cdot\text{m}^2\cdot\text{nm})$ incident at the sensor's entrance pupil using the radiometric response matrix R, calculated above, and the integration time t_{int}, as per the following equation:

$$L_{i,j}(\lambda_{i,j}) = S_{i,j} \cdot \frac{1}{R_{i,j} \cdot t_{int}} \tag{3}$$

2.7. Co-Registration of Images

To take advantage of both spectral and textural features of our subjects, the VNIR and SWIR images need to be combined/stacked correctly and efficiently. To do so, we need to map coordinates of our fine-resolution VNIR (with spatial resolution of 32 μm) scans to our coarse-resolution SWIR (spatial resolution of 3.45 μm) scans. A common approach to achieving this involves resampling the VNIR image by scaling the dimensions of the VNIR (target) using an interpolation function *interp* like INTER_AREA [42]. If we assume that h_{VNIR} and w_{VNIR} represent the height and width of VNIR image, respectively, and h_{SWIR} and w_{SWIR} represent the height and width of the SWIR image, respectively, $scale_h$ and $scale_w$ represent the scaling factors height and width, respectively, which can be calculated as follows:

$$scale_h = \frac{h_{VNIR}}{h_{SWIR}}, \;\; scale_w = \frac{w_{VNIR}}{w_{SWIR}} \tag{4}$$

However, this method demonstrates invariance to both horizontal and vertical translations, as shown in Figure 4. Furthermore, the differences in sensor height resulted in a noticeable horizontal shift in the VNIR images relative to the SWIR images. In this case, differences in scale, rotation, and translation between the source and target images necessitate the use of co-registration [43,44]. In our study, we needed to address two of these three: scale and translation.

Solutions from the public domains that we tested did not produce results within acceptable margins of errors for various reasons—for instance, applying traditional co-registration methods using feature detection and matching, such as Scale-Invariant Feature Transformation or (SIFT) [45] and Speeded-Up Robust Features (or SURF) [46]—and optical flow was hampered by (a) differences in vertical positioning of the scanners, resulting in differences in amounts of detail captured by the different sensors; (b) the differences in spectral ranges captured by the two sensors (400–1000 nm for VNIR and 960–2500 nm for SWIR), which caused the reflectance characteristics of the imaged scene elements to differ significantly; (c) differences in intensity values for the same spectral bands overlapping across the two sensors; and (d) the presence of numerous visually similar artifacts within the imaged scene (in this case, the seeds), causing ambiguity due to the repetitive nature of seed edges and creases.

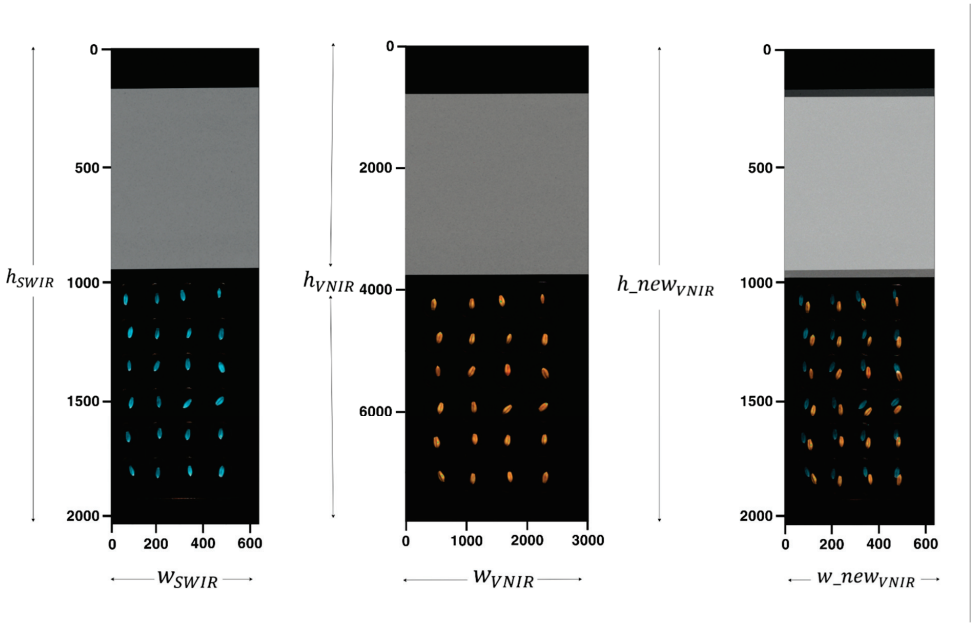


Figure 4. Validation of the need for co-registration. The image on the right shows the result of stacking SWIR and VNIR images after resampling of the VNIR image. Here, the new dimensions of VNIR, $w_new_{VNIR} = round(\frac{w_{VNIR}}{scale_w})$ and $h_new_{VNIR} = round(\frac{h_{VNIR}}{scale_h})$, exactly match the dimensions of the corresponding SWIR image, w_{SWIR} and h_{SWIR}, respectively. However, the invariance of the method to the horizontal and vertical translations of the image reveals a significant shift between the pixels of the stacked images and will result in more errors.

A customized approach to co-registration of our images was therefore necessary. For the first batch of images, fiducial tie point markers were identified manually, illustrated in Figure 5a, using ENVI software. Meanwhile, for the second batch, fiducial markers were placed during the scanning process as mentioned earlier in the text as shown in Figure 5b.

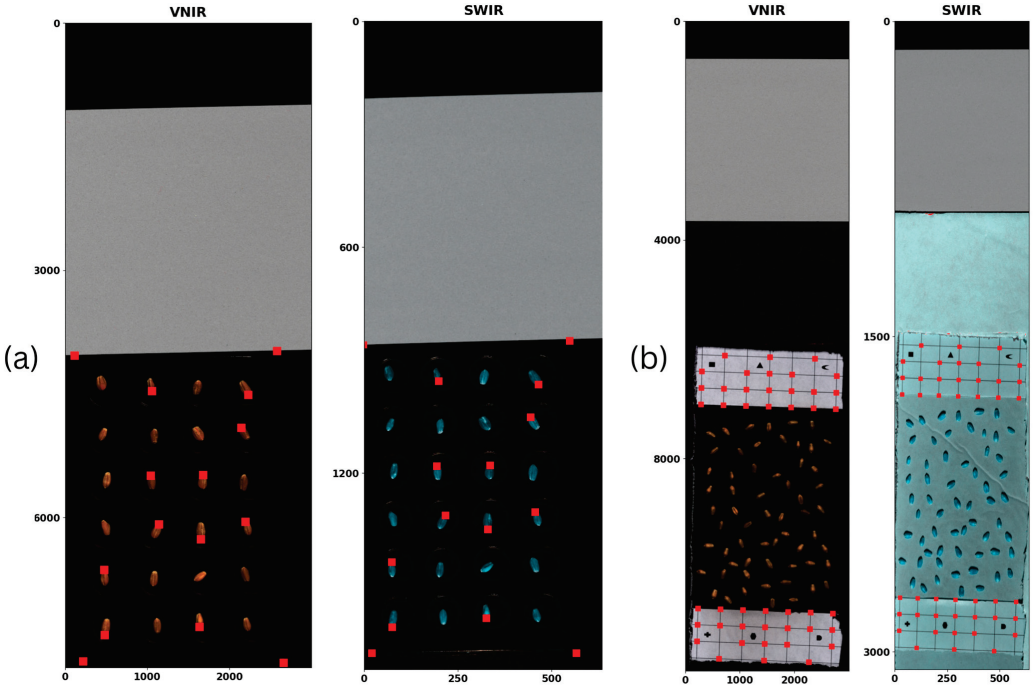


Figure 5. Tie points used to calculate homography for each batch of seed scans. (**a**) The red dots indicate tie points identified manually on the seeds and the background. These were meticulously picked and curated until the lowest error was achieved. (**b**) The second batch of seeds was scanned with tie points placed alongside the samples to ensure minimal manual identification of tie points.

Corresponding coordinates of the tie points from both images were then used for calculation of a homography matrix which maps the rest of the target image's pixels to their corresponding positions in the reference image. Homography describes the transformation of one point from an image to its corresponding point in another image with the assumption that the scene is flat, and the transformation can be estimated by a planar projective transformation. It can be represented in formula as

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \tag{5}$$

where H is our projective transformation 3×3 matrix with the elements of the matrix, i.e., $(h_{11}, h_{12}, \ldots, h_{33})$, being determined by the correspondence between several matched points in the two images, typically at least four of these. In our case, we obtained a minimum of 10 points and at most 17 matched points for our first scan batch and 32 for the second batch of scanned seeds, mapping original points x and y onto their corresponding coordinates x' and y' on the second image, as shown in the formula below:

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{6}$$

Here, (x', y', w') are the homogeneous coordinates. To convert these back to Cartesian coordinates, the homogenous coordinates are divided by w'.

$$x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}},\ y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \tag{7}$$

The transformation is not merely linear but projective as well, meaning it accounts for the perspective changes due to the angle of view. The fact that our sensors capture images from different heights and slightly different angles therefore makes the choice of using homography transformation the most viable one.

To calculate the homography matrices for the images, we used the robust RANSAC (Random Sampling Consensus) algorithm provided by the open-source library OpenCV [47] with a threshold of 7.0. The resulting transformation matrix was then applied to the target image using the warpPerspective [42] function from the same library, which warps the dimensions of the target image along the projective transformation matrix provided. To quantitatively assess the precision of our alignment method, we employed the mean squared error (MSE) and root mean squared error (RMSE) as metrics.

2.8. Background Segmentation, Cropping, and Labeling

Spurious signals from the pixels of the background tray and the reflectance panel will affect the overall mean reflectance spectra of our seed scans and therefore need to be removed. This will allow the predictive models to focus attention purely on the signals from the seed scans. Similar studies have removed spurious signals caused by background material in hyperspectral scans of vegetation using a rule-based approach [39,48], as well as removing soil background from maize [49] using vegetation indices, and also in a different study has been applied to distinguish Petri dish salmon samples using Otsu thresholding [50]. Specifically for bench-top scans, other researchers applied a combination of Otsu thresholding and morphological approaches to segment rice granules from a dark background [51,52].

For the first step of segmentation, a rule-based masking algorithm was applied, taking advantage of the intrinsic differences in reflectance properties between our wheat samples and the background materials, combined with morphological operations to retain seed pixels and remove noise. This approach is based on the distinct spectral signatures exhibited by the seeds, background tray, and reflectance panel, as evident from the spectral profiles shown in Figure 6. The algorithm utilizes reflectance values at specific wavelengths—410 nm, 456 nm, 553 nm, 654 nm, and 852 nm bands—which capture the most significant variations between the seeds and the background materials, abbreviated as B1, B2, G, R, and N, respectively, in the formula below:

$$p_i = \begin{cases} 1, & for\ condition1\ and\ condition2\ and\ condition3\ and\ condition4\ and\ condition5 \\ 0, & for\ condition6\ and\ condition7\ and\ condition8\ and\ condition9\ and\ condition10 \end{cases} \quad (8)$$

$$condition1 = p_{i,B1} < p_{i,B2}$$
$$condition2 = 0.7 * p_{i,B1} \leq p_{i,R} \geq 12.5 * p_{i,B1}$$
$$condition3 = 1.2 * p_{i,B1} \leq p_{i,N} \geq 15 * p_{i,B1}$$
$$condition4 = p_{i,B2} < p_{i,G}$$
$$condition5 = 1.2 * p_{i,B2} < p_{i,G}$$
$$condition6 = p_{i,B1} > p_{i,B2}$$
$$condition7 = 0.7 * p_{i,B1} > p_{i,R} < 12.5 * p_{i,B1}$$
$$condition8 = 1.2 * p_{i,B1} > p_{i,N} < 15 * p_{i,B1}$$
$$condition9 = p_{i,B2} > p_{i,G}$$
$$condition10 = 1.2 * p_{i,B2} > p_{i,G}$$

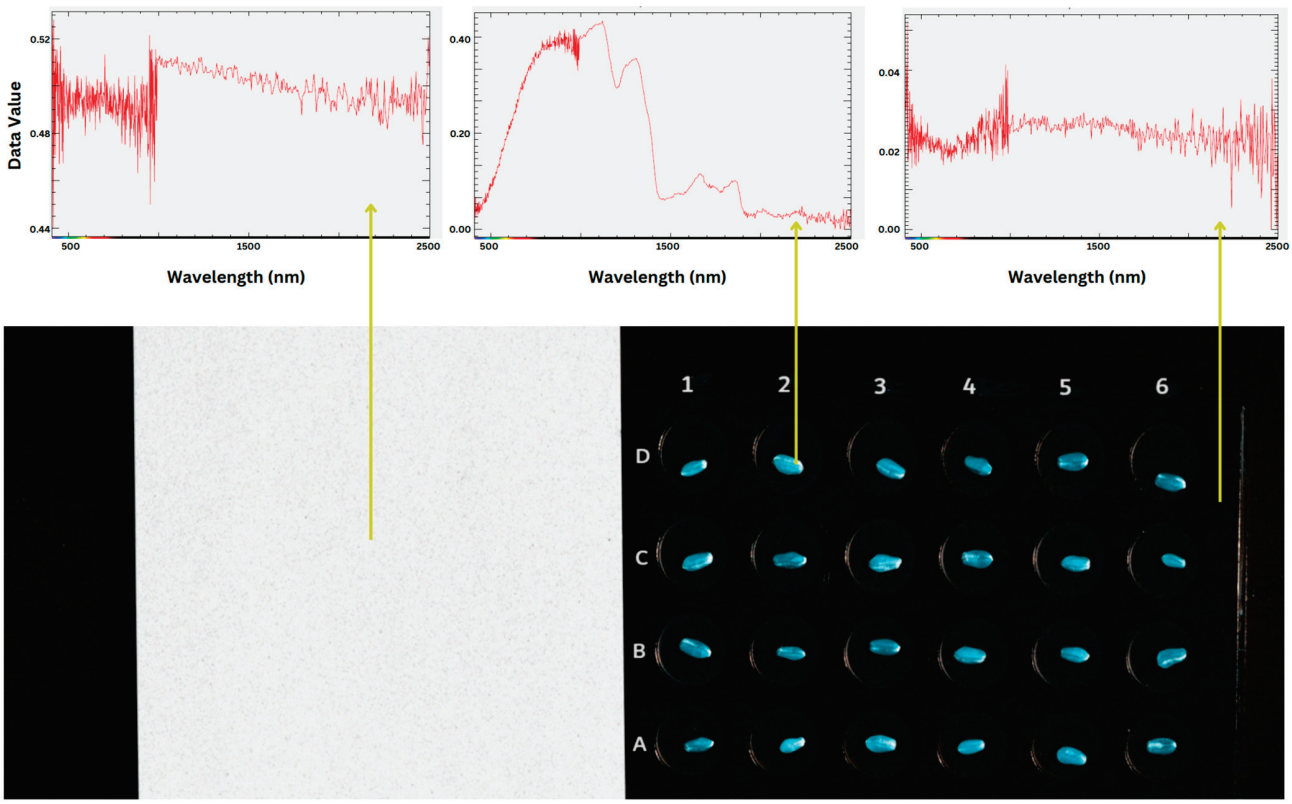


Figure 6. Differences in spectral profiles of the seeds from those of the tray and reflectance panel allow for creation of a rules-based segmentation mechanism to allow separation of seeds from the background. As per earlier explanation, the first batch of seeds were aligned in rows 4-seed-wide labeled A–D and 6-seed-long labeled 1–6.

The rule defines the expected ratio of values of a seed pixel between pairs of bands that distinguishes it from non-seed pixels and was determined experimentally by comparing initial spectral profiles of regions of interest on seed pixels against those of the background pixels. If a pixel satisfies all the defined conditions, then the pixel is considered a seed pixel and retained (assigned a value of 1). Otherwise, the pixel is considered part of the background or shadow and discarded (assigned a value of 0).

For the second step, morphological operations, a technique widely used in image processing for noise removal and boundary smoothing [53], were employed to address persisting background pixels and missed seed pixels, as illustrated in the second and third images of the provided sequence in Figure 7a. A morphological opening operation was performed using a 3×3 structuring element kernel using OpenCV's provided method, morphologyEx [54].

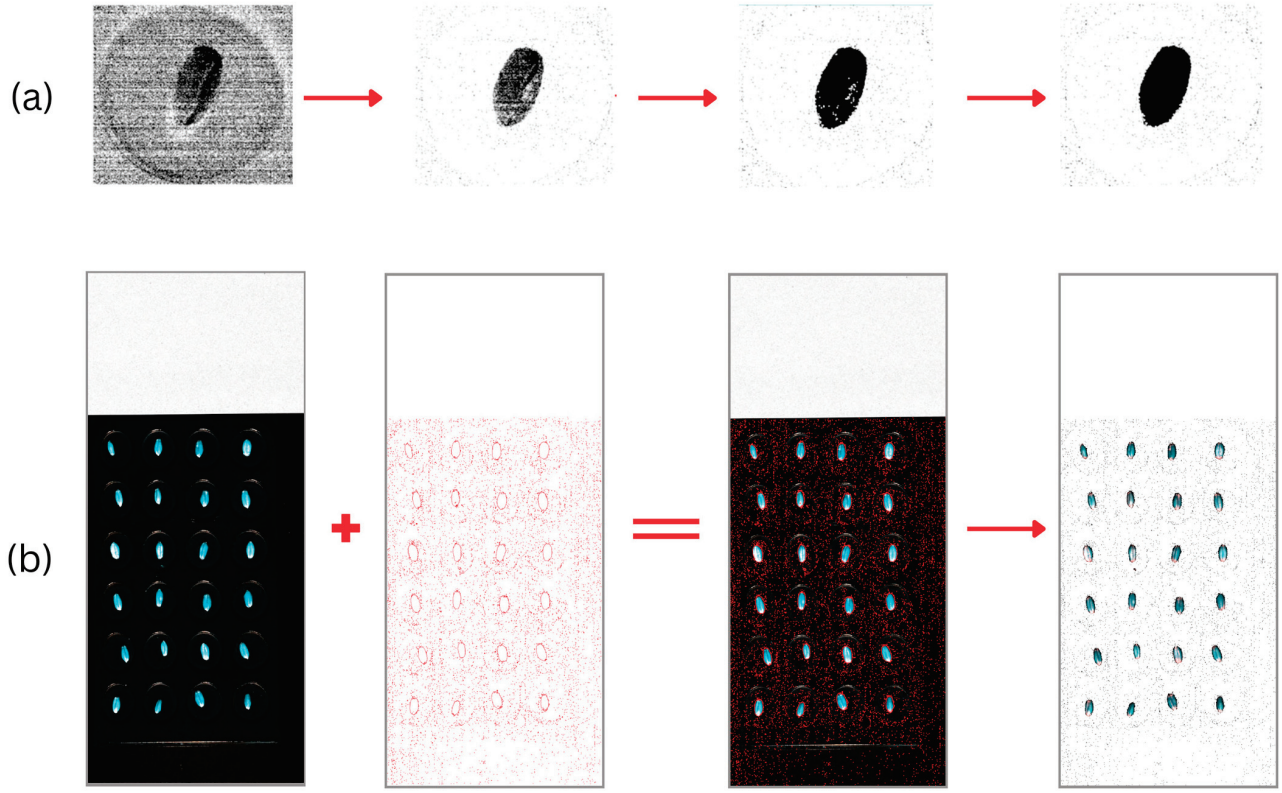


Figure 7. Image segmentation process, showing the process of obtaining and applying masks to segment the seeds from the background. This was achievable due to the spectral differences and color differences between the seeds and background material. (**a**) A zoomed-in single image highlighting the refinement of criteria for the rule-based segmentation approach which, results in the formation of 'holes' within the seed pixel regions and the subsequent result of using flood fill to help fill the 'holes'. (**b**) The application of the generated mask to extract seeds from the background.

Since the full seed images would be used to train a predictive vision model, all the pixels within the established boundary of the seed mask needed to be retained. As such, for the third step, a flood-fill algorithm was implemented, a region-based approach commonly used for image segmentation [55–57]. Starting from the top-left corner (0, 0) of the image, the flood-fill operation provided by OpenCV [58] effectively filled the background region with a new value (2), while preserving the seed pixels (value 1) and any remaining noise pixels (value 0).

In the post-processing stage, any remaining "noise" pixels which were not filled by the flood-fill operation were addressed by assigning them to the seed class (value 1), as we can safely assume these to be "holes" within the seed pixel regions, as shown in Figure 7a.

The filled background region was relabeled as class 0. Finally, the border added during the flood-fill operation was removed to obtain the final segmented image, containing only the seed pixels and the background.

To crop individual seeds from the image, an OpenCV-provided methodology was used, which specified three main steps. The first involves retrieving a three-band image consisting of the red, green, and blue channels. In our study, a single band from each section of the visual range of wavelengths was retrieved, including red (642 nm), green (546 nm), and blue (460 nm) bands, to compose our three-channel image. This image then had to be converted to grayscale, and the result was transformed into a binary format to distinguish the seeds from the background.

Following this, a connected-component analysis using OpenCV's connectedComponentsWithStats method [59] was applied to identify and isolate discrete seed regions, achieved by filtering for regions with areas that satisfied the following threshold: 250 < area < 1500. The isolated-component areas were then labeled according to a scheme facilitated by the spatial ordering of the identified seeds based on their original positions in a grid-like structure. The decision to take this systematic approach to labeling of the seeds was to enable us to correlate each seed scan to its correct protein content labels. This approach would also enable a comparative analysis against and potential fusion with other modes of data captured of the same seeds. A sample of the final single seed images is illustrated in Figure 8.

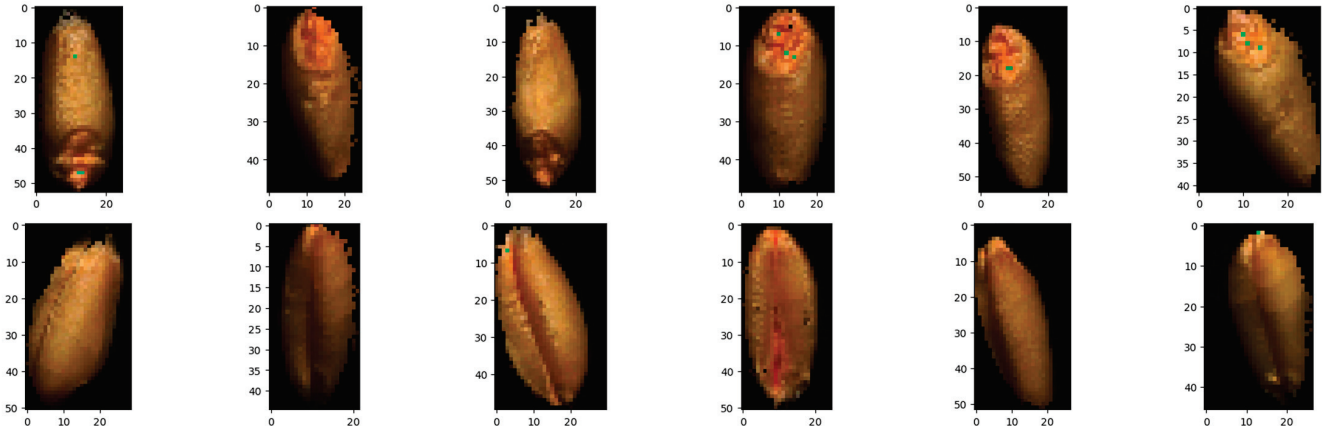


Figure 8. Sample of final cropped single-seed images, with the top row showing the seeds with the crease down (**top** view) and the bottom row showing the seeds with the crease up (**bottom** view).

In summary, for the segmentation portion of the workflow, we applied a combination of a rule-based masking algorithm, morphological operations, a flood-fill algorithm, and post-processing steps to isolate seed pixels from the background. For the subsequent cropping phase, we extracted a three-band image, converted it to grayscale and binary format, performed connected-component analysis, and applied area-based filtering to isolate individual seeds.

2.9. Comparison of Average Spectra per Protein Content

We were able to delineate a correlation between the amount of protein content in the seeds and the intensity of their reflectivity, as demonstrated in the graph in Figure 9, with seeds of higher protein content demonstrating lower reflectivity properties across all the bands, while lower-protein-content seeds indicated the highest intensity but also showed the largest dips at wavelengths 1400–1500 nm and 1900–2000 nm, which could suggest a relatively higher moisture content in the seeds with the lowest protein content, as these bands are the most affected by moisture [60].

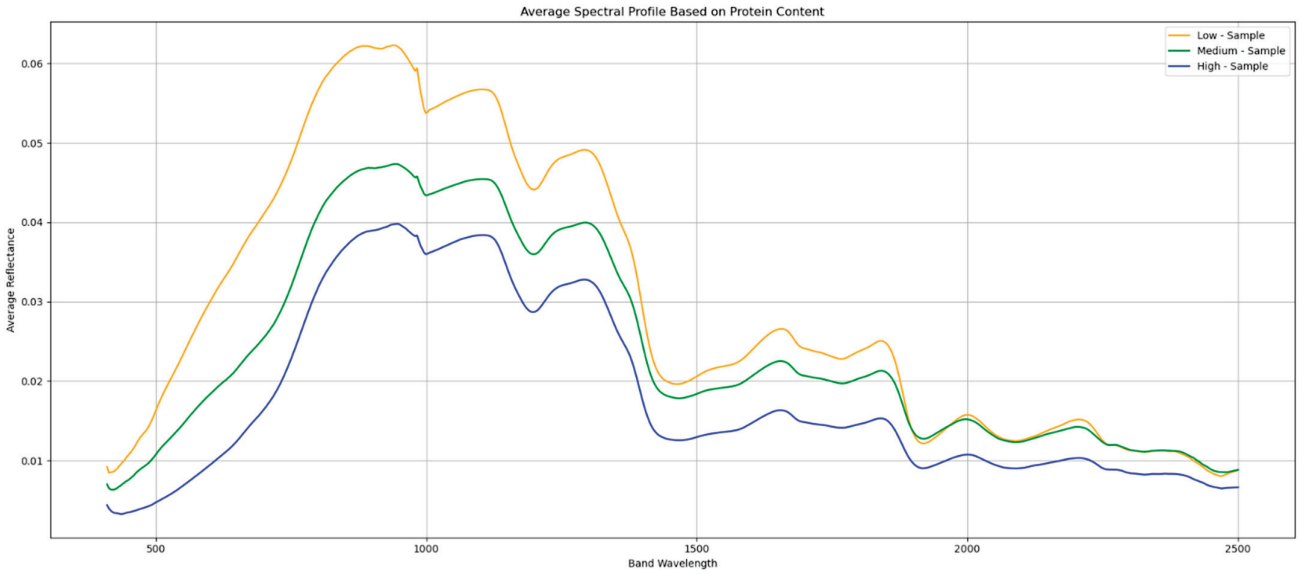


Figure 9. Comparison of average spectra of wheat seed based on the protein content.

2.10. Feature-Based Machine Learning for Protein Content Estimation

The average spectra of each seed scan were calculated and stored for further analysis. In HSI applications, signal smoothing is often necessitated by the presence of noise caused by the scanning instruments [61,62]. We discovered that the signal was mostly free of noise except for the last few bands. We therefore applied minimally intrusive smoothing on the spectra using the Savitzky–Golay [63] method provided by the SciPy library [64], with a window of 7 and polyorder of 3.

Spectral indices have been used previously in HSI studies to successfully predict different properties of numerous study subjects, including soil properties [65,66] and crop yield [67,68]. While there are currently no spectral indices developed specifically for bench-top scans of seeds, we attempted a simple band ratio approach in attempt to improve the predictive capabilities of our models. Given a sample i and for each unique pair of bands j and k (where $j < k$), this approach can be summarized as

$$R_{ijk} = \frac{X_{ij}}{X_{ik}} \tag{9}$$

where R_{ijk} represents the ratio of the spectral values between bands j and k for sample i, j iterates from 1 to the total number of bands, and k iterates from $j + 1$ to the total number of bands.

These data were then normalized to ensure that all features contributed proportionately to the objective function, preventing those with larger scales from dominating the optimization process of our predictive models [69]. Furthermore, by scaling our features to a common scale, normalization mitigates the issues of convergences in algorithms sensitive to the scale of the input data, such as logistic regression and Support Vector Machines [70].

We processed 5057 samples for the top orientation and 4746 for the bottom orientation. These samples were then split into train, test, and validation, with a test to train ratio of 3:1 and a validation to train ratio of 4:1. This yielded train, test, and validation datasets of 3034, 1264, and 759, respectively, for top orientation and 2847, 1187, and 712, respectively, for bottom orientation.

To extract the most informative bands for the regression modeling, we performed permutation feature importance analysis on each model and extracted the top 100 wavelengths. We then calculated the ratios of these bands, combined them with the original 100 bands, and trained a fresh set of models from scratch. This essentially yielded three different datasets, as summarized in Table 3.

Table 3. Summary of datasets generated.

Dataset Number	**1**	**2**	**3**
Dataset Type	All spectral bands	Top 100 spectral bands + band ratios of top 100 bands	All spectral bands + ratios of all bands

2.10.1. Random Forest Regression

Random Forest is an ensemble learning method which operates by constructing multiple decision trees and harmonizes the outputs of the individual trees to produce the final result [71]. At the individual tree level, random splitting of the features occurs at each node, resulting in increased variation among the trees and thus improving the robustness of the final model [71]. This is supplemented by a technique called bagging or bootstrap aggregating, a technique that involves building each tree on a bootstrap sample of the data, which helps reduce overfitting [72]. Random Forest can handle high-dimension spaces as well as large sample sets very effectively too, making it a good choice for our regression task.

2.10.2. Support Vector Regression

Support vector regression is a type of Support Vector Machine used for regression tasks but based on the same principles used for SVM classification, which is to find a best-fit line within a threshold value ε where distances from the actual targets y_i to the line are minimized and at as many as possible points lie within the ε-tube without penalty [73]. To determine this line, a combination of the points closest to the fit (support vectors) and a regularization parameter, C, are used to balance between minimizing model complexity and achieving a lower error on the training data. To solve nonlinear regression tasks, SVR uses kernel functions to map input data into higher-dimensional space where a linear regression can be fitted [74]. SVR is known for its strong performance in smaller datasets and its capability to manage high-dimensional spaces and has shown promising results in previous similar research [75,76].

2.11. Image-Based CNN Approaches

2.11.1. Data Preprocessing Methods

Data need to be scaled and normalized for CNN training pipelines as they facilitate faster convergence of the model [77] and improve the stability of the neural network [78]. Meanwhile, other research shows that normalizing within the neural network architecture itself through batch normalization can facilitate a more stable gradient flow and acceleration of the training process [78], is instrumental in preventing vanishing and exploding gradients [79], and in some cases has been shown to have a regularization effect that can prevent overfitting in neural networks [80]. For our study, we applied minmax normalization.

The final co-registered images consist of 662 contiguous bands over a wavelength range between 412 and 2500 nm. While this high dimensionality facilitates capturing intricate spectral profiles, it unfortunately suffers from a few intractable limitations, including the 'curse of dimensionality' phenomenon, where the complexity of data analysis tasks grows exponentially with increases in dimensionality [81]; data sparsity in higher dimensional spaces [82]; and increasing the effects of noise and irrelevant features [83]. We therefore applied dimensionality reduction to our data using PCA (Principal Component Analysis), a commonly used dimensionality reduction method for HSI applications [62,84,85].

Starting with our four-dimensional tensor with shape (H, W, C), where H and W are our spatial dimensions for each seed scan and C is the number of channels or spectral bands, we transform individual seed scans into a 2D vector of shape ($H \times W$, C), thus presenting each pixel as a distinct observation within the channel space. Applying PCA to our reshaped array reduces our channel dimension C to our predetermined number of principal components K followed by a reshaping of our data into (H, W, K), which

preserves our spatial dimensions. The number of principals K was determined via empirical performance in our specific application with the aim of balancing between achieving higher performance metrics while keeping computational complexity low.

The final step is to convert each three-dimensional tensor into a four-dimensional tensor. Here, it is necessary to apply 3D CNNs as they typically deal with volumetric data as they expect the input shape to be (N, H, W, D, C) [86], where N is the batch size and D is the new depth dimension that represents the number of channels in our image. In this case, this will correspond to K, specified above.

2.11.2. Model Architecture

The protein content estimation task consisted of two phases. The first was based on regression modeling using acquired truth labels representing the percentage of seed mass which was taken up by proteins. The second was classification using the three classes mentioned earlier in this paper as either low, medium, or high depending on the criteria specified.

For image-based protein content estimation, we used a modified version of HybridSN [87], a hybrid convolutional neural network (CNN) that combines the capabilities of 3D convolution layers which have typically seen use with 3D imagery such as 3D lidar (Voxel R-CNN) and the 2D convolutional layers used for machine learning with the standard 2D imagery for classification, segmentation, and regression tasks. This architecture is summarized in Figure 10a. The nature of this structure allows the model to leverage the depth of spectral analysis provided by the 3D convolution layers and the spatial detail recognition capabilities of the 2D convolution layers.

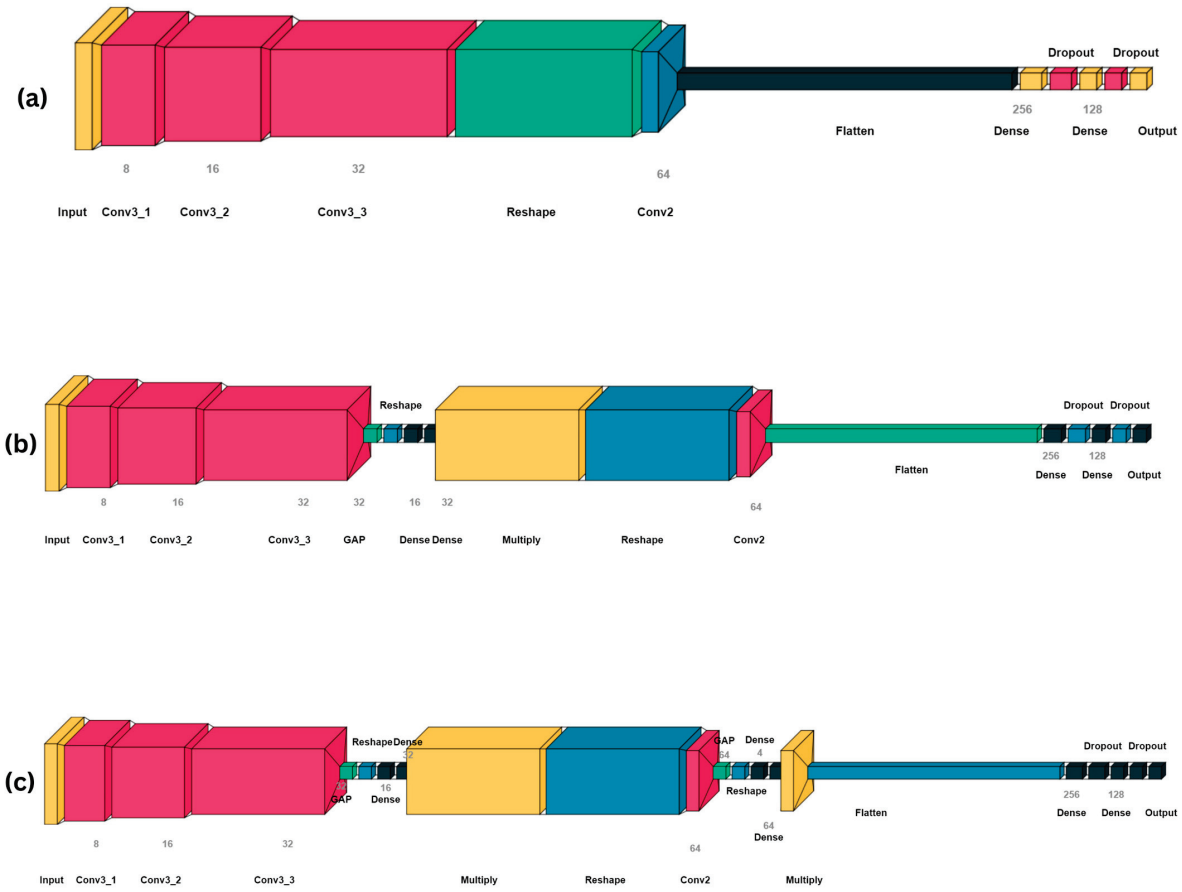


Figure 10. Visualization of the architectures of the different variations of HybridSN used in this study, including (**a**) the original architecture, (**b**) the modified architecture with global attention, and (**c**) the final modified model architecture with an additional attention mechanism of the Squeeze-and-Excitation network.

Assuming that our individual HSI data cubes can be denoted by $Xi \in R^{(H\times W)\times D}$, with D as the number of bands, then for each input X_i, a corresponding real-valued number $y_i \in R$ represents the target variable, which is represented as the percentage of the protein content's contribution to the overall mass of a seed.

The 3D convolution layer applies a 3D filter across the spatial and spectral dimensions of our input image, obtaining a dot product of kernel weights and local regions of the input volume covered and producing a 3D feature map in an operation that can be summarized by the following expression:

$$F(x,y,z) = \sum_{i=0}^{a-1}\sum_{j=0}^{b-1}\sum_{k=0}^{c-1} V(x+i,\ y+j,\ z+k)\cdot K(i,j,k) \tag{10}$$

assuming a kernel of size $a \times b \times c$ moving over input volume V at position (x, y, z) on the output feature map.

This approach ensures that we capture both spatial and spectral feature information of our input hypercube, and by stacking multiple 3D convolution layers, the model can learn a hierarchy of features, from the simplest patterns in the earlier layers to more complex features in the deeper layers.

This can be further condensed into

$$F_l = \sigma(B_l + \sum_{i,j,k} X_{i,j,k} * K_l) \tag{11}$$

where F_l, B_l, and K_l represent the feature map, bias term, and the 3D kernel, respectively, at the l-th layer, and σ is the activation function.

The 3D convolutions are subsequently followed by a 2D convolution operation as shown below:

$$F'_m = \sigma(B'_m + \sum_{i,j} F'_{Lij} * K'_m) \tag{12}$$

with F'_{Lij} being the output feature map from the previous 3D convolution layers. Finally, we flatten the feature map into a vector and pass through a vector $V\epsilon\ R^n$, which is passed through a series of dense layers, with the final output being

$$D_p = \sigma(W_p \cdot D_{p-1} + b_p) \tag{13}$$

where D_p, W_p, and b_p are the output, weight, and bias, respectively, of the p-th layer, and D_{p-1} is the output from the previous dense layer or the flattened vector $V\epsilon\ R^n$ in the case of the first dense layer. The final regression output thus becomes

$$Y = W_k \cdot D_{k-1} + b_k \tag{14}$$

With Y being the predicted continuous values corresponding to our regression targets. To train the regression model, we used the mean squared error (MSE) loss function, which minimizes the average squared difference between the predicted and true protein content values, ensuring accurate predictions of the continuous target variable.

As for the classification task, where if we assume the same notations for our input vectors as before, for each Xi, there exists a corresponding $y_i = [y_{i,1}, y_{i,2}, \ldots, y_{i,C}]$, such that $y_{i,j} \in \{0,1\}$, $\sum_{j=1}^{C} y_{ij} = 1$, and $C = 3$, which corresponds to our number of protein content classes.

Consequently, our final output becomes

$$Y = softmax(W_k \cdot D_{k-1} + b_k) \tag{15}$$

where Y is the probability distribution over classes. To train this classification model, we used the categorical cross-entropy loss function, which measures the divergence between the predicted class probabilities and the true one-hot encoded labels.

In the HybridSN documentation [87], the authors noted that 25 × 25 spatial dimensions worked best for their application. For our case, we also experimented with spatial dimensions of 32 × 32 and found that they produced the best results for our dataset. This resulted in an input size of $n \times 32 \times 32 \times 662$ for our model. We also experimented with different kernel sizes for the different 3D layers and documented the results. The first set of kernel sizes was the original kernel size specified from layer 1 to layer 3 as $3 \times 3 \times 7$, $3 \times 3 \times 5$, and $3 \times 3 \times 3$. The second set was $3 \times 3 \times 11$, $3 \times 3 \times 9$, and $3 \times 3 \times 7$ and the third set was $5 \times 5 \times 15$, $5 \times 5 \times 11$, and $5 \times 5 \times 9$. We also experimented with different combinations of batch sizes and learning rates. The empirical performances of each combination set were then documented.

For validation of the models, we calculated the errors and accuracies based on several metrics. The first was the coefficient of determination, calculated as

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y}_i)^2} \tag{16}$$

where y_i is the observed values, $\hat{y}$ is the predicted values, $\overline{y}$ is the mean of the observed values, and n is the number of observations.

We also evaluated the model using mean absolute error, which calculates the average prediction error in the units of variable being predicted and is denoted as

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{17}$$

Mean squared error is more useful for detecting outliers, as it penalizes larger errors more severely than it does smaller ones, and is calculated as

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{18}$$

Meanwhile, root mean square provides the standard deviation of the residuals (prediction errors), where lower values indicate a better fit, and is calculated simply as the square root of the MSE:

$$RMSE = \sqrt{MSE} \tag{19}$$

For the classification task, we used four different metrics: accuracy, precision, recall, and F1 score. Accuracy measures the overall correctness of the model and works well when the classes are balanced. It is defined as

$$Accuracy = \frac{True\ Positives\ (TP) + True\ Negatives\ (TN)}{Total\ Number\ of\ Samples} \tag{20}$$

Precision is a measure of accuracy of positive predictions and is useful when the cost of a false positive is high. It is defined as

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \tag{21}$$

Recall measures the model's ability to detect all relevant cases within a dataset and is used then the cost of a false negative is high. It is defined as

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)} \tag{22}$$

Finally, F1 score is the harmonic mean of precision and recall and can be expressed as

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{23}$$

The HybridSN model was modified through the addition of an attention mechanism to process the feature map output of the 3D convolution block. The global attention block [88] calculates the importance weights of each channel in the feature map. These weights are then used to scale the original channel values, effectively amplifying important channels and attenuating the less relevant ones.

A Global Average Pooling layer takes input tensor I with dimensions $H \times W \times D \times C$, with $H, W,\ D,$ and C as the height, width, spectral depth, and number of channels from the immediately preceding 3D convolution layer, reducing I to a 1D tensor with an equal number of elements to the number of channels, C. A dense layer applies a reduction in dimensionality followed by ReLU activation, and a second layer projects back to the original number of channels with a sigmoid activation to create attention map A:

$$A = \sigma(W_2 \cdot (ReLU(W_1 \cdot G + b_1)) + b_2) \tag{24}$$

where $W_1 \in \mathrm{R}^{C \times \frac{C}{2}}$ and $b_1 \in \mathrm{R}^{\frac{C}{2}}$ are the weight matrix and bias of the first layer, $W_2 \in \mathrm{R}^{\frac{C}{2} \times C}$, and $b_2 \in \mathrm{R}^{C}$ are the weight matrix and bias of the second layer, and our attention map has dimensions of $1 \times 1 \times 1 \times C$.

The final operation involves scaling each channel of the input feature by its corresponding attention weight:

$$O = \mathrm{I} \odot A \tag{25}$$

Here, O retains the dimensions $H \times W \times D \times C$.

In essence, the attention mechanism works by learning a set of weights from the input tensor itself, which is then applied to modulate the input tensor's channels, allowing the network to prioritize some channels over others. The final architecture is summarized in Figure 10b.

An additional architectural modification was implemented on the HybridSN network through the addition of a Squeeze-and-Excitation network after the 2D convolution layer to recalibrate channel-wise feature importances. First introduced by Jie [89], the Squeeze-And-Excitation block involves initially compressing each feature map into a single number by averaging over all the spatial dimensions, producing a vector of size equal to the number of channels. Given dimensions $H \times W \times C$ of our feature map, this can be expressed as

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{ijc} \tag{26}$$

This is followed by a bottleneck layer which reduces the dimensionality of the output from GPA, compressing channel-wise features to a lower-dimensional representation, and an expansion layer which reverses the representation back to its original channel-wise representation, going through a sigmoid transformation first. The two can be expressed as follows:

$$v = ReLU(W_1 \cdot z + b_1) \tag{27}$$

$$s = Sigmoid(W_2 \cdot v + b_2) \tag{28}$$

Finally, channel-wise multiplication is applied and used to scale the original feature map F:

$$F' = s \odot F \tag{29}$$

3. Results and Discussion

3.1. Co-Registration Results

As explained in our methodology, co-registration was a necessary and crucial step in the pipeline that made it possible to combine the full range of bands from VNIR and SWIR together and thus create a new method to determine protein content in wheat seeds. As such, the goal was to minimize errors as much as possible to ensure we had correctly aligned multi-sensor seed scans. We compared the errors of stacking the images using a simple resampling approach against the errors after using our method. The results are summarized in Table 4 and illustrated in Figure 11.

Table 4. Comparison of errors of co-registration when using simple resampling methodology vs. using our custom methodology.

	Average RMSE	Max RMSE	Min RMSE	Median RMSE
Resampling	0.1246	0.8846	0.0804	0.1211
Our method	0.0487	0.1381	0.0094	0.0205

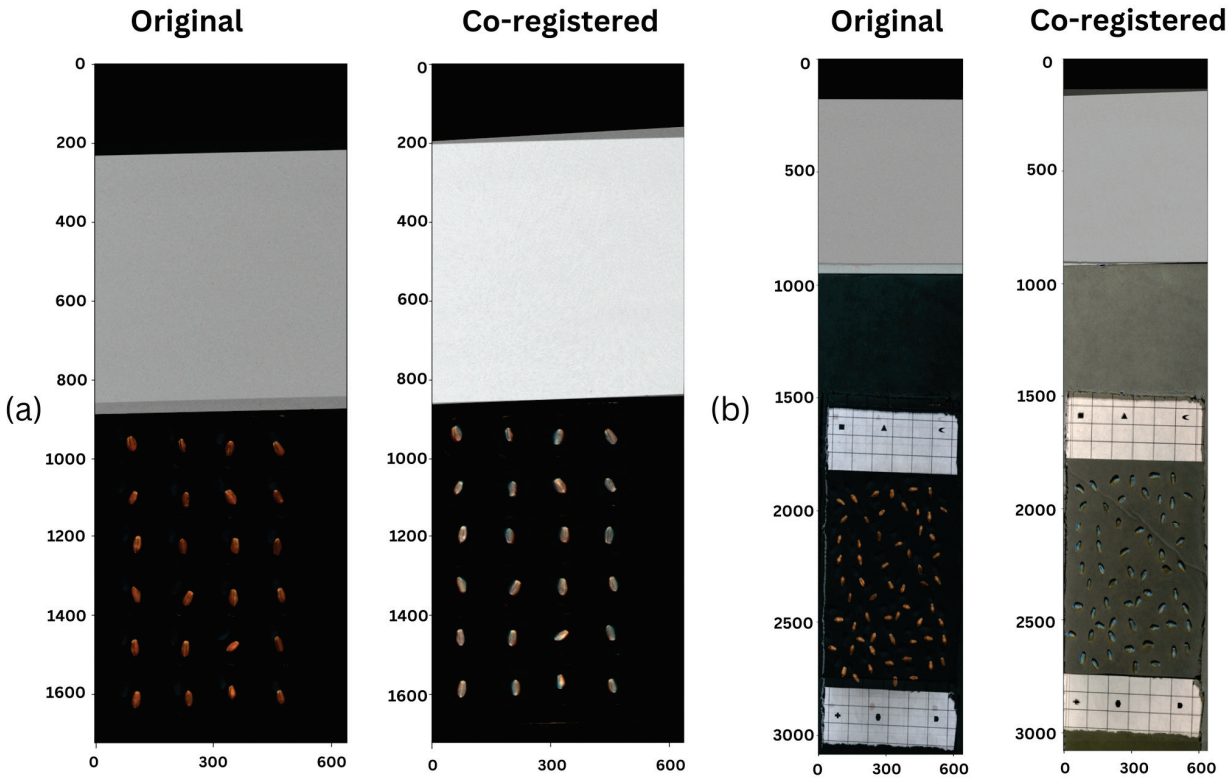


Figure 11. Analysis of the visual differences between simple resampling strategy and our semi-automated method to co-register VNIR and SWIR scans for (**a**) the first batch of scans and (**b**) the second batch of scans.

Overall, our method produced significantly less errors compared to the more simplified resampling method.

3.2. Protein Content Estimation

3.2.1. Feature-Based Machine Learning for Protein Content Estimation

The best overall model was Random Forest, with an R^2 of 0.77, comparable to when Caporaso et al. [11,15] used a standard normal variate and first derivatives as features. In our case, we found that the model trained on ratios of all the bands fared better than the ones trained on ratios of the most informative bands alone, with a coefficient of determination

score of 0.77 and 0.75 for Random Forest for bottom and top orientations, respectively, and 0.66 and 0.67 with SVR for bottom and top orientations, respectively. This could be an indication that even though some wavelengths, individually, are not as discriminative for seed protein estimation as those in the NIR region, they still contribute to the model training if their information is extracted and processed in a different way, rather than used directly. The results of these experiments are summarized in Supplementary Table S1.

Feature permutation analysis indicated that the wavelengths 2456 nm, 426 nm, 564 nm, 1399 nm, and 1409 nm appeared to be the most important input features for Random Forest modeling for the bottom side of the wheat scans and that 426 nm, 1399 nm, 2456 nm, 2501 nm, and 1409 nm were most important for top orientation, as seen in Figure 12. Meanwhile, the best wavelengths for SVR were 2203 nm, 2206 nm, 2212 nm, 2200 nm, and 986 nm for the bottom orientation and 1678 nm, 1681 nm, 1684 nm, 1674 nm, and 1722 nm for top orientation, as shown in Figure 13. These findings suggest the dominance of data in the NIR region over those of the visible region of the spectrum for studying the chemical characteristics of wheat seeds and corroborate the findings of previous research.

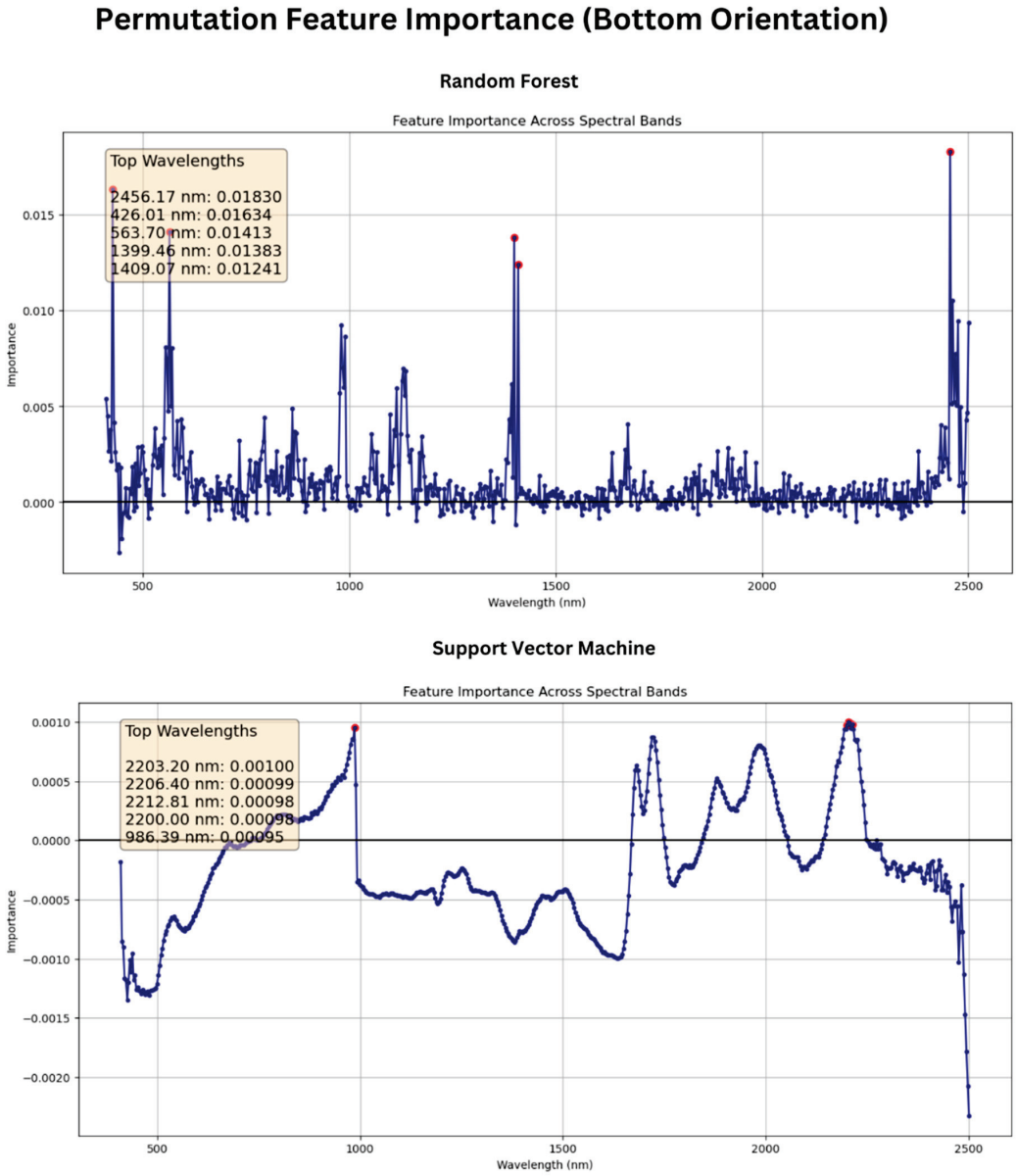


Figure 12. Permutation feature importance analysis for wheat seeds with crease facing up.

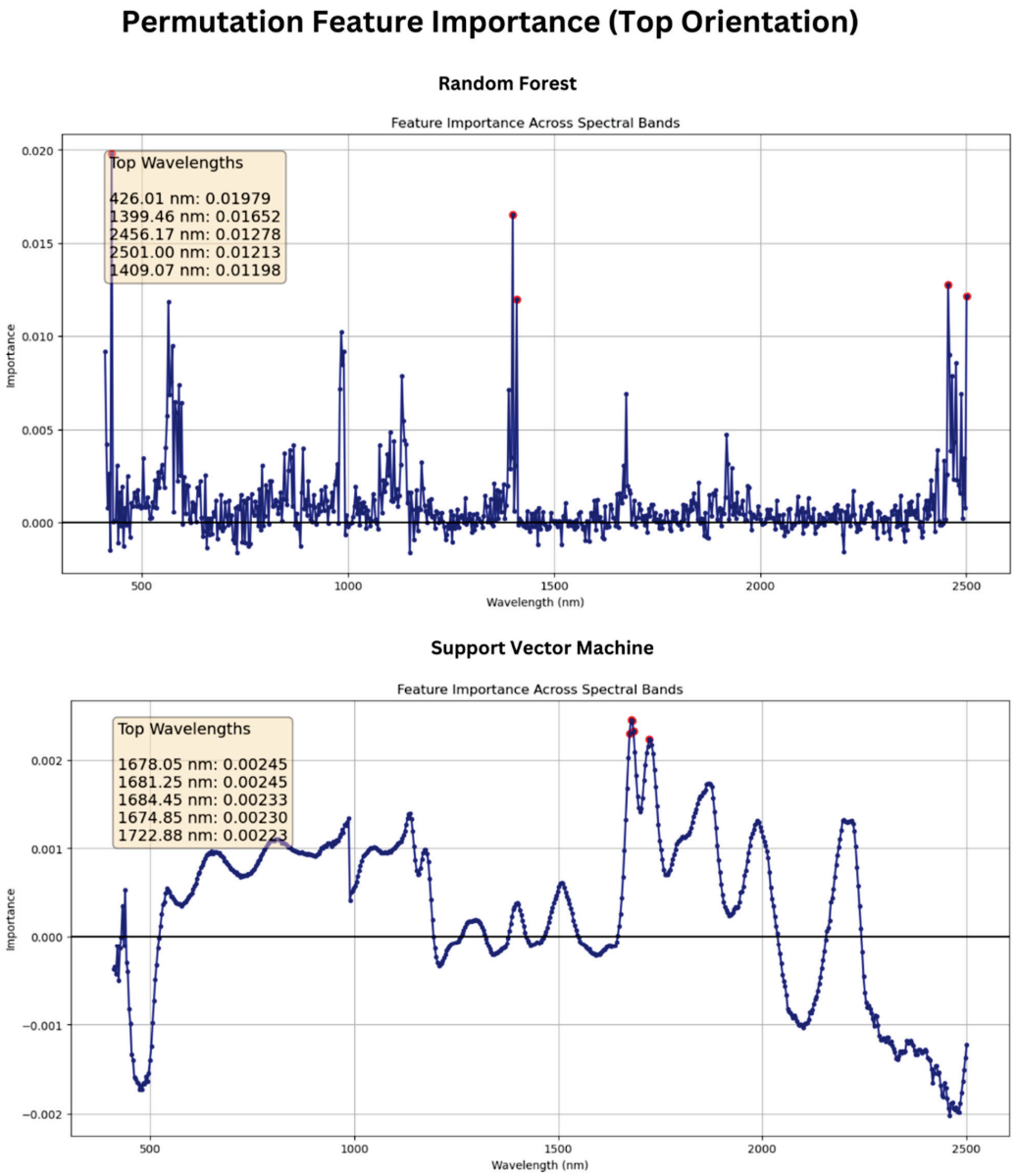


Figure 13. Permutation feature importance analysis for wheat seeds with crease facing down.

3.2.2. Image-Based Machine Learning for Protein Content Estimation

For image-based modeling, we found the best model to have an R^2 of 0.70 for the ventral side (bottom orientation) and 0.65 for the dorsal side (top orientation), with comparably similar performance across the different kernel size sets and model types. These are summarized in Supplementary Tables S2 and S3.

However, when compared with fixed values of batch size and kernel sizes, we found that the addition of attention would improve the performance of the model significantly and reduce errors, as indicated in Supplementary Tables S4 and S5 for the performance of the bottom-oriented and top-oriented seeds, respectively.

Increasing the kernel depth translated to a better performance of the model in some cases, with the best overall model working well with the second kernel size set, but this was not always the case. We found that increasing the spatial dimensions of the kernel to 5×5 made the performance of the model much worse before we ceased experimenting with higher spatial kernel dimensions.

A cross-examination of the models using scatterplots, as shown in Figure 14, demonstrates that the original model struggles with instances on all ends and those in the middle,

and while our models with attention achieve a slight improvement with the middle-placed values, they still also struggle with instance on the ends, especially the higher ends. This observation was made on both the training and test datasets. Overall, all models were able to achieve an RMSE of less than 2%.

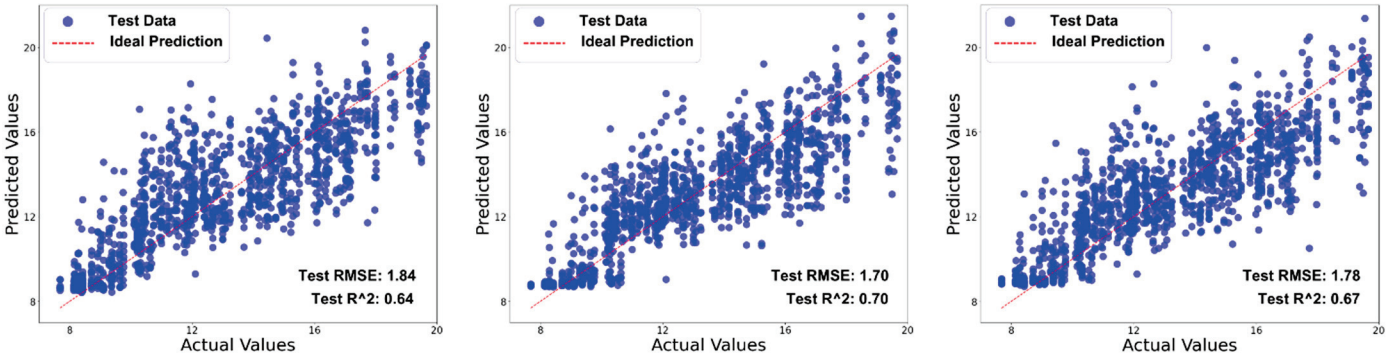


Figure 14. Scatterplot of predictions for seed protein content using the different variations of HybridSN. From left to right: original HybridSN architecture, HybridSN with global attention block, and HybridSN with global attention and Squeeze-and-Excite network blocks. The tighter clustering of predicted values around the ideal prediction line for the second model suggests better predictions with less variance for that model.

These results demonstrate that our CNN model can produce results that match or improve upon the performance of feature-based models, producing R^2 scores at least similar to or higher (0.65 and 0.70) than those produced by SVR (0.65 and 0.66) in experiments while being outperformed by the best feature-based model, Random Forest, by no more than 9%. This is encouraging because developing end-to-end, rapid image-based AI-powered approaches is a prerequisite for determining seed composition in operational settings.

The inclusion of the 3D CNN block in the HybridSN architecture allowed the model to learn and retain valuable spectral features from the HSI, in addition to the spatial features. This suggests that relying solely on spatial features from seed scans may be insufficient, and that retaining spectral information is crucial for accurate prediction of protein content.

Furthermore, the inclusion of a global attention layer after the 3D CNN block and a Squeeze-and-Excitation block improved the model's performance. The global attention layer allows the model to focus on the most relevant spatial–spectral features across the entire input, while the SE network adaptively recalibrates channel-wise feature responses in the spatial domain.

The performance gains achieved by incorporating these attention mechanisms suggest that the model benefits from selectively emphasizing informative spectral and spatial features. Notably, the global attention layer, which operates on the spectral dimension, likely contributes to the model's ability to better utilize the spectral information present in the HSI data. This further supports the argument that spectral features play a crucial role in accurately predicting protein content from wheat seed HSI scans.

When classifying the wheat samples according to protein content in the three classes, we found that our models were able to more easily discriminate between the different wheat seed samples based on protein content, with a high accuracy of 0.82 for both orientations on our test dataset. Additionally, we found high precision, recall, and F1 scores for most of the models, indicating the well-balanced performance of the models in identifying true positives while minimizing false positives. These results are available in Supplementary Tables S6 and S7. However, we found that the addition of attention mechanisms did not improve results enough to justify their use and in some cases affected the performance

negatively across all metrics. This appears in line with findings from another study using global attention [90].

For our study, this could be because the increased complexity of using attention mechanisms may have led to further overfitting, especially since our dataset was small (about 5000 per orientation). These mechanisms inherently add more parameters and layers to the model, and this often requires substantial data to help the models generalize more.

Although feature-based approaches that utilized spectral data (indies optimized for extracting features of interest) provided good results, they require manual selection of indices and feature engineering, which are often challenging without subject matter expertise. Spatial and texture information of seed kernels, which has been proven to be important in previous studies, is not considered in feature-based approaches. The image-based approaches proposed in this paper leverage both spatial and spectral features from hyperspectral images and can be implemented in automated seed sorting and breeding processes.

3.3. Sources of Error

The rich information contained in full-spectrum HSI data necessitates a discussion on the possible contemporary use of the technology in real-world applications. However, before this idea can be realized, we need to address and nullify the effects of errors that impact negatively on the performance of our models. A few sources we identified include the following:

- The optical and surface topography properties of wheat seeds potentially play a big role in the performance of the model, as we observed in our own results and others [91]. The top sides of the wheat seeds (crease-down) were observed to have a smoother surface with higher reflectivity scores than those of the bottom orientation of the same seed. The bottom side of the seeds has an uneven surface caused by the crease, which can result in scattering and incident reflection. A switch to using absorbance data, like Caporaso et al. [11], or harmonization of both can provide more insight into how much impact these effects have on the final output of our modeling efforts and result in better performances.
- We also theorize that the sample size affected the accuracy of our predictive models. Here, this limitation manifests in two ways. One is the actual number of wheat samples that were scanned and assessed for protein content via combustion analysis. We had a total of nearly 10,000 samples to use for training, which limited us from experimenting with advanced models such as transformers, which require massive datasets for training [92,93]. Secondly, we acquired our ground-truth protein content using Dumas combustion, which has a minimum weight limit of samples in order to produce protein content labels. Consequently, the protein content values were acquired per sample of at least 60 wheat kernels rather than per individual wheat kernel. It is plausible that this disparity in granularity between the training data and the corresponding ground-truth labels prevented a more precise mapping of the spatial–spectral characteristics of wheat seeds with protein content.

3.4. Future Direction

The main challenge of using HSI for single-wheat-kernel prediction of protein is the large sizes of the datasets. The average file size of a single scanning session was easily between 36 GB and 72 GB. Handling such large datasets requires a large amount of storage space and computing power and can prove to be challenging for most research efforts. One possible solution for this is to use spectral and spatial binning while performing image correction during the processing stage. Binning involves downsampling the spatial or

spectral properties of our image scans to a specified degree and will reduce the file sizes significantly at the cost of losing some information.

Secondly, to the best of our knowledge, no commercial tools exist that can co-register VNIR and SWIR images collected with bench-top scanners at the time of writing. This can prove to be cumbersome when you must process hundreds of datasets. We developed a semi-automated co-registration tool that allowed for batch processing of seed scans, resulting in faster processing of huge datasets.

Future research and development should focus on the scalability of the co-registration and image calibration process by introducing an accurate, fully automatic co-registration workflow. Additionally, foundational models need to be developed with increased sample data for training large language models so that the methods developed in this paper can be deployed in industrial settings.

4. Conclusions

The objectives of this paper were to (1) develop robust methods for seed trait prediction using a bench-top hyperspectral imaging scanner in the VNIR and SWIR region; (2) determine optimal wavelengths that are useful for the prediction of seed traits such as protein content; and (3) investigate how the orientation of wheat kernels during scanning affected the viability and usefulness of the data for predictive modeling.

Our results show the following:

1. The end-to-end CNN with attention mechanisms explained wheat kernel protein content with up to 0.7 R^2 values. R^2 values were 0.70 and 0.65 for ventral and dorsal seed orientations, respectively, indicating that seed orientation for scanning has a significant impact on the predictive power of CNN models.
2. Consistent with many previous studies, feature-based RF outperformed the direct imagery-based CNN with a higher R^2 of 0.77. However, feature engineering requires expert knowledge of the features (spectral indices in our case) and is time-consuming, not being scalable for operational scenarios. End-to-end prediction capabilities with CNNs are preferred for the automation of wheat protein estimation for breeding.
3. The CNN model predicted low, medium, and high protein concentrations with an R^2 of 0.82. The finding is significant because this type of classification provides more meaningful means (compared to estimating the exact amount of protein content) for crop breeding and seed profiling. Classification-based seed profiling with hyperspectral imaging and machine learning techniques plays a pivotal role in advancing precision breeding practices, optimizing seed sorting processes, and enabling targeted agricultural input applications.
4. Sensitivity analysis showed that NIR bands outperform all other hyperspectral bands for seed protein content estimation and have higher importance scores for models. However, when combined with information from other bands in the hyperspectral range, it was possible to produce more robust models.

Our study highlights significant potential for advancing agricultural practices in a targeted and efficient way. By moving beyond traditional NIR spectroscopy and incorporating a broader spectrum of data, we can deepen our understanding of factors that influence seed quality, with far-reaching implications including the following:

1. Enabling more precise breeding practices by enabling breeders to accurately and rapidly predict desirable qualities of seeds, with the knock-on effect of improved yield and higher quality crops.
2. Additionally, this technology could also lead to a more targeted use of agricultural inputs, such as fertilizers and pesticides, based on the genetic information of the seed material, as identified by our model.

3. Further, the orientation-dependent findings in our study suggest that a customized approach to agricultural machinery and processes is necessitated and should be geared towards optimizing seed handling in a manner that ensures high accuracy.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/s25020303/s1: Table S1: Best performing ML models for protein content regression; Table S2: Best performing deep learning models for protein content regression (bottom orientation); Table S3: Best performing deep learning models for protein content regression (top orientation); Table S4: Comparison of effects of different hyperparameter combinations on final coefficient of determination scores and reported errors for the different CNN architectures (bottom orientation); Table S5: Comparison of effects of different hyperparameter combinations on final coefficient of determination scores and reported errors for the different CNN architectures (top orientation); Table S6: Evaluation of protein content classification models for the different CNN architectures and different combinations of hyperparameters (bottom orientation); Table S7: Evaluation of protein content classification models for the different CNN architectures and different combinations of hyperparameters (top orientation).

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

Funding: This work was financially supported by Bayer Crop Science and the Saint Louis University Research Innovation Fund.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are not available for public release.

Conflicts of Interest: Author Kyle T. Peterson was employed by the company Bayer Crop Science. 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. Shewry, P.R.; Hey, S.J. The contribution of wheat to human diet and health. *Food Energy Secur.* **2015**, *4*, 178–202. [CrossRef] [PubMed]
2. Farouk, M.M.; Yoo, M.J.Y.; Hamid, N.S.A.; Staincliffe, M.; Davies, B.; Knowles, S.O. Novel meat-enriched foods for older consumers. *Food Res. Int.* **2018**, *104*, 134–142. [CrossRef] [PubMed]
3. Kniskern, M.A.; Johnston, C.S. Protein dietary reference intakes may be inadequate for vegetarians if low amounts of animal protein are consumed. *Nutrition* **2011**, *27*, 727–730. [CrossRef] [PubMed]
4. Smith, F.; Pan, X.Y.; Bellido, V.; Toole, G.A.; Gates, F.K.; Wickham, M.S.J.; Shewry, P.R.; Bakalis, S.; Padfield, P.; Mills, E.N.C. Digestibility of gluten proteins is reduced by baking and enhanced by starch digestion. *Mol. Nutr. Food Res.* **2015**, *59*, 2034–2043. [CrossRef]
5. Gorissen, S.H.M.; Horstman, A.M.H.; Franssen, R.; Crombag, J.J.R.; Langer, H.; Bierau, J.; Respondek, F.; van Loon, L.J.C. Ingestion of Wheat Protein Increases In Vivo Muscle Protein Synthesis Rates in Healthy Older Men in a Randomized Trial. *J. Nutr.* **2016**, *146*, 1651–1659. [CrossRef]
6. Shewry, P. What Is Gluten-Why Is It Special? *Front. Nutr.* **2019**, *6*, 101. [CrossRef]
7. Wieser, H.; Koehler, P.; Scherf, K.A. Chemistry of wheat gluten proteins: Qualitative composition. *Cereal Chem.* **2023**, *100*, 23–35. [CrossRef]
8. Nicolaï, B.M.; Beullens, K.; Bobelyn, E.; Peirs, A.; Saeys, W.; Theron, K.I.; Lammertyn, J. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biol. Technol.* **2007**, *46*, 99–118. [CrossRef]

9. Kokaly, R.F.; Couvillion, B.R.; Holloway, J.M.; Roberts, D.A.; Ustin, S.L.; Peterson, S.H.; Khanna, S.; Piazza, S.C. Spectroscopic remote sensing of the distribution and persistence of oil from the Deepwater Horizon spill in Barataria Bay marshes. *Remote Sens. Environ.* **2013**, *129*, 210–230. [CrossRef]
10. Baye, T.M.; Pearson, T.C.; Settles, A.M. Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. *J. Cereal Sci.* **2006**, *43*, 236–243. [CrossRef]
11. Caporaso, N.; Whitworth, M.B.; Fisk, I.D. Protein content prediction in single wheat kernels using hyperspectral imaging. *Food Chem.* **2018**, *240*, 32–42. [CrossRef] [PubMed]
12. Fu, D.D.; Zhou, J.F.; Scaboo, A.M.; Niu, X.F. Nondestructive phenotyping fatty acid trait of single soybean seeds using reflective hyperspectral imagery. *J. Food Process Eng.* **2021**, *44*, e13759. [CrossRef]
13. Zhang, L.; Wang, Y.Q.; Wei, Y.G.; An, D. Near-infrared hyperspectral imaging technology combined with deep convolutional generative adversarial network to predict oil content of single maize kernel. *Food Chem.* **2022**, *370*, 131047. [CrossRef] [PubMed]
14. Zhang, L.; Zhang, Q.; Wu, J.Z.; Liu, Y.T.; Yu, L.; Chen, Y.Y. Moisture detection of single corn seed based on hyperspectral imaging and deep learning. *Infrared Phys. Technol.* **2022**, *125*, 104279. [CrossRef]
15. Caporaso, N.; Whitworth, M.B.; Fisk, I.D. Near-Infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains. *Appl. Spectrosc. Rev.* **2018**, *53*, 667–687. [CrossRef]
16. Saha, D.; Senthilkumar, T.; Sharma, S.; Singh, C.B.; Manickavasagan, A. Application of near-infrared hyperspectral imaging coupled with chemometrics for rapid and non-destructive prediction of protein content in single chickpea seed. *J. Food Compos. Anal.* **2023**, *115*, 104938. [CrossRef]
17. Wang, L.; Tian, Y.; Yao, X.; Zhu, Y.; Cao, W. Predicting grain yield and protein content in wheat by fusing multi-sensor and multi-temporal remote-sensing images. *Field Crops Res.* **2014**, *164*, 178–188. [CrossRef]
18. Wang, Z.; Huang, W.; Tian, X.; Long, Y.; Li, L.; Fan, S. Rapid and Non-destructive Classification of New and Aged Maize Seeds Using Hyperspectral Image and Chemometric Methods. *Front. Plant Sci.* **2022**, *13*, 849495. [CrossRef]
19. Yang, S.; He, D.; Ning, J. Predicting wheat kernels' protein content by near infrared hyperspectral imaging. *Int. J. Agric. Biol. Eng.* **2016**, *9*, 163–170. [CrossRef]
20. Pasikatan, M.C.; Dowell, F.E. High-speed NIR segregation of high- and low-protein single wheat seeds. *Cereal Chem.* **2004**, *81*, 145–150. [CrossRef]
21. Mishra, P.; Lohumi, S. Improved prediction of protein content in wheat kernels with a fusion of scatter correction methods in NIR data modelling. *Biosyst. Eng.* **2021**, *203*, 93–97. [CrossRef]
22. Delwiche, S.R.; Hruschka, W.R. Protein content of bulk wheat from near-infrared reflectance of individual kernels. *Cereal Chem.* **2000**, *77*, 86–88. [CrossRef]
23. Liu, Y.; Delwiche, S.R.; Graybosch, R.A. Two-dimensional correlation analysis of near infrared spectral intensity variations of ground wheat. *J. Near Infrared Spectrosc.* **2009**, *17*, 41–50. [CrossRef]
24. Pasquini, C. Near Infrared Spectroscopy: Fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc.* **2003**, *14*, 198–219. [CrossRef]
25. Eldin, A.B.; Akyar, I. Near infra red spectroscopy. In *Wide Spectra of Quality Control*; InTech: Rijeka, Croatia, 2011; pp. 237–248.
26. Hyspex. HySpex VNIR-3000N. Available online: https://www.hyspex.com/media/ehbhuyge/hyspex_vnir-3000n.pdf (accessed on 1 July 2024).
27. Hyspex. HySpex SWIR-640. Available online: https://www.hyspex.com/hyspex-products/hyspex-classic/hyspex-swir-640/ (accessed on 1 July 2024).
28. Hyspex. Full Range Light Source. Available online: https://www.hyspex.com/hyspex-products/accessories/light-sources/full-range-light-source/ (accessed on 1 July 2024).
29. Muñoz-Huerta, R.F.; Guevara-Gonzalez, R.G.; Contreras-Medina, L.M.; Torres-Pacheco, I.; Prado-Olivarez, J.; Ocampo-Velazquez, R.V. A Review of Methods for Sensing the Nitrogen Status in Plants: Advantages, Disadvantages and Recent Advances. *Sensors* **2013**, *13*, 10823–10843. [CrossRef] [PubMed]
30. Serrano, S.; Rincón, F.; García-Olmo, J. Cereal protein analysis via Dumas method: Standardization of a micro-method using the EuroVector Elemental Analyser. *J. Cereal Sci.* **2013**, *58*, 31–36. [CrossRef]
31. Beljkas, B.; Matic, J.; Milovanovic, I.; Jovanov, P.; Misan, A.; Saric, L. Rapid method for determination of protein content in cereals and oilseeds: Validation, measurement uncertainty and comparison with the Kjeldahl method. *Accredit. Qual. Assur.* **2010**, *15*, 555–561. [CrossRef]
32. Jung, S.; Rickert, D.A.; Deak, N.A.; Aldin, E.D.; Recknor, J.; Johnson, L.A.; Murphy, P.A. Comparison of Kjeldahl and Dumas methods for determining protein contents of soybean products. *J. Am. Oil Chem. Soc.* **2003**, *80*, 1169–1173. [CrossRef]
33. Wiles, P.G.; Gray, I.K.; Kissling, R.C. Routine analysis of proteins by Kjeldahl and Dumas methods: Review and interlaboratory study using dairy products. *J. AOAC Int.* **1998**, *81*, 620–632. [CrossRef]
34. Agelet, L.E.; Hurburgh, C.R. Limitations and current applications of Near Infrared Spectroscopy for single seed analysis. *Talanta* **2014**, *121*, 288–299. [CrossRef]

35. Schläpfer, D.; Richter, R.; Feingersh, T. Operational BRDF Effects Correction for Wide-Field-of-View Optical Scanners (BREFCOR). *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1855–1864. [CrossRef]
36. Bhadra, S.; Sagan, V.; Sarkar, S.; Braud, M.; Mockler, T.C.; Eveland, A.L. PROSAIL-Net: A transfer learning-based dual stream neural network to estimate leaf chlorophyll and leaf angle of crops from UAV hyperspectral images. *ISPRS J. Photogramm. Remote Sens.* **2024**, *210*, 1–24. [CrossRef]
37. Wang, Y.K.; Gu, Y.F.; Li, X.M. A Novel Low Rank Smooth Flat-Field Correction Algorithm for Hyperspectral Microscopy Imaging. *IEEE Trans. Med. Imaging* **2022**, *41*, 3862–3872. [CrossRef]
38. Kask, P.; Palo, K.; Hinnah, C.; Pommerencke, T. Flat field correction for high-throughput imaging of fluorescent samples. *J. Microsc.* **2016**, *263*, 328–340. [CrossRef]
39. Sagan, V.; Maimaitijiang, M.; Paheding, S.; Bhadra, S.; Gosselin, N.; Burnette, M.; Demieville, J.; Hartling, S.; LeBauer, D.; Newcomb, M.; et al. Data-Driven Artificial Intelligence for Calibration of Hyperspectral Big Data. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5510320. [CrossRef]
40. Jiao, Z.T.; Schaaf, C.B.; Dong, Y.D.; Román, M.; Hill, M.J.; Chen, J.M.; Wang, Z.S.; Zhang, H.; Saenz, E.; Poudyal, R.; et al. A method for improving hotspot directional signatures in BRDF models used for MODIS. *Remote Sens. Environ.* **2016**, *186*, 135–151. [CrossRef]
41. Lenhard, K.; Baumgartner, A.; Gege, P.; Nevas, S.; Nowy, S.; Sperling, A. Impact of Improved Calibration of a NEO HySpex VNIR-1600 Sensor on Remote Sensing of Water Depth. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6085–6098. [CrossRef]
42. OpenCV. Geometric Image Transformations. Available online: https://docs.opencv.org/4.x/da/d54/group__imgproc__transform.html (accessed on 14 July 2024).
43. Reddy, B.S.; Chatterji, B.N. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Trans. Image Process.* **1996**, *5*, 1266–1271. [CrossRef]
44. Gottesfeld, L. A Survey of Image Registration. *ACM Comput. Surv.* **1992**, *24*, 326–376.
45. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
46. Bay, H.; Tuytelaars, T.; Van Gool, L. SURF: Speeded Up Robust Features. *Comput. Vis. Image Underst.* **2006**, *110*, 346–359. [CrossRef]
47. OpenCV. Camera Calibration and 3D Reconstruction. Available online: https://docs.opencv.org/4.x/d9/d0c/group__calib3d.html#ggaf63d4c617c61e03ab502b3c836966150a724159df258a5d7e29410a6a2f4e6c87 (accessed on 15 July 2024).
48. Maimaitijiang, M.; Sagan, V.; Bhadra, S.; Nguyen, C.; Mockler, T.C.; Shakoor, N. A fully automated and fast approach for canopy cover estimation using super high-resolution remote sensing imagery. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *3*, 219–226. [CrossRef]
49. Shu, M.Y.; Zuo, J.Y.; Shen, M.Y.; Yin, P.F.; Wang, M.; Yang, X.H.; Tang, J.H.; Li, B.G.; Ma, Y.T. Improving the estimation accuracy of SPAD values for maize leaves by removing UAV hyperspectral image backgrounds. *Int. J. Remote Sens.* **2021**, *42*, 5864–5883. [CrossRef]
50. Xu, L.; Jia, H.M.; Lang, C.B.; Peng, X.X.; Sun, K.J. A Novel Method for Multilevel Color Image Segmentation Based on Dragonfly Algorithm and Differential Evolution. *IEEE Access* **2019**, *7*, 19502–19538. [CrossRef]
51. Su, Y.; Xiao, L.-T. 3D Visualization and Volume-Based Quantification of Rice Chalkiness In Vivo by Using High Resolution Micro-CT. *Rice* **2020**, *13*, 69. [CrossRef]
52. Mittal, S.; Dutta, M.K.; Issac, A. Non-destructive image processing based system for assessment of rice quality and defects for classification according to inferred commercial value. *Measurement* **2019**, *148*, 106969. [CrossRef]
53. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 4th ed.; Pearson Education Limited: Harlow, Essex, UK, 2018.
54. OpenCV. Morphological Transformations. Available online: https://docs.opencv.org/4.x/d9/d61/tutorial_py_morphological_ops.html (accessed on 15 July 2024).
55. Roychoudhury, A.; Missura, M.; Bennewitz, M. Plane Segmentation Using Depth-Dependent Flood Fill. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021. [CrossRef]
56. Essamlali, A.; Millot-Maysounabe, V.; Chartier, M.; Salin, G.; Becq, A.; Arrivé, L.; Camus, M.D.; Szewczyk, J.; Claude, I. Bile Duct Segmentation Methods Under 3D Slicer Applied to ERCP: Advantages and Disadvantages. *Int. J. Biomed. Eng. Clin. Sci.* **2023**, *9*, 66–74. [CrossRef]
57. Januszewski, M.; Maitin-Shepard, J.; Li, P.; Kornfeld, J.; Denk, W.; Jain, V. Flood-Filling Networks. *arXiv* **2016**, arXiv:1611.00421.
58. OpenCV. Miscellaneous Image Transformations. Available online: https://docs.opencv.org/4.x/d7/d1b/group__imgproc__misc.html (accessed on 15 July 2024).
59. OpenCV. Structural Analysis and Shape Descriptors: connectedComponentsWithStats. Available online: https://docs.opencv.org/4.x/d3/dc0/group__imgproc__shape.html#ga107a78bf7cd25dec05fb4dfc5c9e765f (accessed on 15 July 2024).
60. Manley, M. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chem. Soc. Rev.* **2014**, *43*, 8200–8214. [CrossRef]

61. Lowe, A.; Harrison, N.; French, A.P. Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* **2017**, *13*, 80. [CrossRef]
62. Li, Y.; Zhang, H.K.; Shen, Q. Spectral-Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network. *Remote Sens.* **2017**, *9*, 67. [CrossRef]
63. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **2002**, *36*, 1627–1639. [CrossRef]
64. Scipy. scipy.signal.savgol_filter. Available online: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.savgol_filter.html (accessed on 15 July 2024).
65. Ben-Dor, E.; Chabrillat, S.; Demattê, J.A.M.; Taylor, G.R.; Hill, J.; Whiting, M.L.; Sommer, S. Using Imaging Spectroscopy to study soil properties. *Remote Sens. Environ.* **2009**, *113*, S38–S55. [CrossRef]
66. Stevens, A.; Udelhoven, T.; Denis, A.; Tychon, B.; Lioy, R.; Hoffmann, L.; van Wesemael, B. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. *Geoderma* **2010**, *158*, 32–45. [CrossRef]
67. Dong, X.; Feng, G.; Zemach, I. Using normalized difference vegetation index to estimate sesame drydown and seed yield. *J. Crop Improv.* **2021**, *35*, 508–521. [CrossRef]
68. Hoyos-Villegas, V.; Fritschi, F.B. Relationships Among Vegetation Indices Derived from Aerial Photographs and Soybean Growth and Yield. *Crop Sci.* **2013**, *53*, 2631–2642. [CrossRef]
69. Jollife, I.T. *Principal Component Analysis*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2002.
70. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2009.
71. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
72. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
73. Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.; Vapnik, V. Support Vector Regression Machines. In Proceedings of the Advances in Neural Information Processing Systems 9 (NIPS 1996), Denver, CO, USA, 2–5 December 1996.
74. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]
75. Chen, J.; Zhu, S.; Zhao, G. Rapid determination of total protein and wet gluten in commercial wheat flour using siSVR-NIR. *Food Chem.* **2017**, *221*, 1939–1946. [CrossRef]
76. Singh, N.; Kaur, S.; Philanim, W.S.; Kumar, A.; Sharma, P.; Ananthan, R.; Bhardwaj, R. Integrating NIR spectroscopy with machine learning and heuristic algorithm-assisted wavelength selection algorithms for protein content prediction in rice bean (*Vigna umbellata* L.). *Food Humanit.* **2024**, *3*, 100399. [CrossRef]
77. LeCun, Y.A.; Bottou, L.; Orr, G.B.; Müller, K.R. Efficient BackProp. In *Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 1998.
78. Sergey Ioffe, C.S. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
79. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
80. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
81. Donoho, D. High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality. *AMS Math Chall. Lect.* **2000**, *1*, 32.
82. Aggarwal, C.; Hinneburg, A.; Keim, D. On the Surprising Behavior of Distance Metric in High-Dimensional Space. In *Database Theory—ICDT 2001, Proceedings of the 8th International Conference, London, UK, 4–6 January 2001*; Lecture Notes in Computer Science; Van den Bussche, J., Vianu, V., Eds.; Springer: Berlin/Heidelberg, Germany, 2001; Volume 1973, pp. 420–434.
83. Trunk, G.V. A Problem of Dimensionality: A Simple Example. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI-1*, 306–307. [CrossRef]
84. Sikakollu, P.; Dash, R. Ensemble of multiple CNN classifiers for HSI classification with Superpixel Smoothing. *Comput. Geosci.* **2021**, *154*, 104806. [CrossRef]
85. Paoletti, M.E.; Haut, J.M.; Plaza, J.; Plaza, A. A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 120–147. [CrossRef]
86. Ben Hamida, A.; Benoît, A.; Lambert, P.; Ben Amar, C. 3-D Deep Learning Approach for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4420–4434. [CrossRef]
87. 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. [CrossRef]
88. Liu, Y.; Shao, Z.; Hoffmann, N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021.
89. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.

90. Hu, F.; Cui, Z.; Wu, S.; Liu, Q.; Wu, J.; Wang, L.; Tan, T. Second-Order Global Attention Networks for Graph Classification and Regression (Abstract). In *Artificial Intelligence, Proceedings of the Second CAAI International Conference, CICAI 2022, Beijing, China, 27–28 August 2022*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 496–507.
91. Wesley, I.J.; Osborne, B.G.; Larroque, O.; Bekes, F. Measurement of the Protein Composition of Single Wheat Kernels Using near Infrared Spectroscopy. *J. Near Infrared Spectrosc.* **2008**, *16*, 505–516. [CrossRef]
92. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in vision: A survey. *ACM Comput. Surv. (CSUR)* **2022**, *54*, 1–41. [CrossRef]
93. Wang, W.; Zhang, J.; Cao, Y.; Shen, Y.; Tao, D. Towards Data-Efficient Detection Transformers. In *Computer Vision—ECCV 2022, Proceedings of the 17th European Conference, Tel Aviv, Israel, 23–27 October 2022*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 88–105.

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.

sensors

MDPI

Article

High-Resolution Single-Pixel Imaging of Spatially Sparse Objects: Real-Time Imaging in the Near-Infrared and Visible Wavelength Ranges Enhanced with Iterative Processing or Deep Learning

Rafał Stojek [1,2], Anna Pastuszczak [1], Piotr Wróbel [1], Magdalena Cwojdzińska [1], Kacper Sobczak [1] and Rafał Kotyński [1,*]

[1] Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland; rafal.stojek@fuw.edu.pl (R.S.); anna.pastuszczak@fuw.edu.pl (A.P.); piotr.wrobel@fuw.edu.pl (P.W.); m.cwojdzinska@student.uw.edu.pl (M.C.); k.sobczak4@student.uw.edu.pl (K.S.)
[2] VIGO Photonics, Poznańska 129/133, 05-850 Ożarów Mazowiecki, Poland
* Correspondence: rafal.kotynski@fuw.edu.pl

Abstract: We demonstrate high-resolution single-pixel imaging (SPI) in the visible and near-infrared wavelength ranges using an SPI framework that incorporates a novel, dedicated sampling scheme and a reconstruction algorithm optimized for the rapid imaging of highly sparse scenes at the native digital micromirror device (DMD) resolution of 1024 × 768. The reconstruction algorithm consists of two stages. In the first stage, the vector of SPI measurements is multiplied by the generalized inverse of the measurement matrix. In the second stage, we compare two reconstruction approaches: one based on an iterative algorithm and the other on a trained neural network. The neural network outperforms the iterative method when the object resembles the training set, though it lacks the generality of the iterative approach. For images captured at a compression of 0.41 percent, corresponding to a measurement rate of 6.8 Hz with a DMD operating at 22 kHz, the typical reconstruction time on a desktop with a medium-performance GPU is comparable to the image acquisition rate. This allows the proposed SPI method to support high-resolution dynamic SPI in a variety of applications, using a standard SPI architecture with a DMD modulator operating at its native resolution and bandwidth, and enabling the real-time processing of the measured data with no additional delay on a standard desktop PC.

Keywords: single-pixel imaging; infrared imaging; compressive imaging; computational imaging; image reconstruction algorithms; deep learning; signal processing

1. Introduction

Single-pixel imaging (SPI) [1,2] has led to a multiplicity of novel ideas about image measurement at various wavelength ranges, spectral imaging, imaging through scattering media, 3D imaging, etc. [3–5]. SPI is designed to capture images using a single detector, instead of the millions of detectors (pixels) included in traditional cameras. This enables imaging in wavelength ranges where conventional multi-pixel sensor arrays are expensive or even unavailable (e.g., infrared [6–10], ultraviolet [11], terahertz [12,13], or X-ray [14]). Additionally, SPI makes it possible to capture images through scattering or opaque media, including fog or biological objects, where conventional imaging techniques often fail due to light scattering.

High-resolution SPI (in the meaning of the number of pixels) puts immense demands on the image modulation bandwidth, and the popular DMD modulators operating at frequencies on the order of 20 kHz are often used at actual resolutions between 32 × 32 and 256 × 256, which is a fraction of their capability. In fact, with no compression and in

the most straightforward differential imaging mode with a doubled number of expositions, SPI imaging at the resolution of 32×32 pixels is possible at only 10 Hz (since displaying $32 \times 32 \times 2 = 2048$ patterns at 20 kHz takes 0.1 s). The acquisition of a single high-quality 1024×768 image takes a lot more time. For instance, in a recently proposed fast Fourier SPI method [15], it required as much as 1180 s, which is close to four orders of magnitude more than in our present paper. Long acquisition times may be hard to accept in practical applications. However, compressed SPI imaging has been shown many times at resolutions up to 256×256 and at frequencies between 5 and 15 Hz [10,16–20]. Unfortunately, with high compression, image reconstruction becomes computationally demanding. In this paper, we push SPI imaging with DMD modulator to its limits, going for the native 1024×768 binary mode resolution, with efficiently implemented differential sampling and image reconstruction. The cost of this approach is accepting extreme compression, and only sparse images can be reconstructed accurately.

In recent years, there has been growing interest in applying deep learning methods to compressive SPI [21]. In [17], an end-to-end solution is proposed, where a convolutional autoencoder generates sensing patterns and reconstructs images from sparse SPI measurements in real time. Recent studies have explored the use of neural networks (NNs) in SPI systems, either for end-to-end image reconstruction [22–24] or as denoising tools that complement other reconstruction methods [25,26]. The latter approach offers more flexibility, as the NN can be easily adjusted to work with various sampling schemes, compression levels, and changing measurement conditions. Additionally, there is growing interest in combining convolutional and recurrent NN to reconstruct temporal sequences of images from dynamic compressive measurements [27,28]. Although NNs provide flexible, fast, and increasingly accurate solutions to inverse problems, they also have notable disadvantages. The training process is typically time-consuming and demands significant GPU power and memory. Moreover, to achieve high-quality reconstructions and sufficient generality, an NN must be trained on large, diverse datasets, which can pose a challenge in some scenarios. To address these issues, there has been interest in exploring the possibility of applying untrained NNs to SPI image reconstruction [29–31].

In this paper, we propose a framework for fast, high-resolution single-pixel imaging applicable to spatially sparse objects. This work extends our previous study [32,33], where we introduced a sampling scheme based on image maps and a two-stage image reconstruction approach. The sampling is now modified to distribute information about the mean value of the object across all sampling patterns. The iterative reconstruction algorithm is reformulated with minor adjustments in matrix form for more straightforward implementation on GPUs. Additionally, we have developed Python code using the PyTorch module, enabling GPU execution. We also introduce an alternative second-stage reconstruction method based on a proposed NN, implemented in PyTorch. This latter approach is significantly faster but lacks the generality of the iterative algorithm. Finally, the proposed algorithms are validated in an optical setup with two parallel channels, with detectors operating in the VIS and NIR wavelength ranges, respectively.

The paper is organized as follows. In the Materials and Methods section, we provide separate subsections to explain the consecutive elements of the SPI framework. In Sections 2.1–2.3, we describe the development of sampling patterns and the motivation behind this approach. In Section 2.4, we explain the initial stage of the reconstruction algorithm, which relies on the generalized inverse of the measurement matrix. Sections 2.5 and 2.6 detail the two alternative approaches for the second reconstruction stage, and Section 2.7 describes the optical setup. In the Results section, in Section 3.1, we outline the computational and bandwidth requirements and limitations. In Section 3.2, we discuss various noise sources, and we demonstrate cases where the proposed SPI framework produces high-resolution results. Finally, in Section 3.3, we present our experimental results. The paper concludes with the discussion in Section 4.

2. Materials and Methods

Sampling patterns and the image reconstruction algorithm are the two components that define any SPI framework. They are often independent, and various image reconstruction algorithms may be used in conjunction with different sampling patterns [4,12,34,35]. The block diagram of the SPI framework proposed in this paper is shown in Figure 1, and the corresponding SPI optical setup is shown in Figure 2. The details of the SPI operation are presented in the following subsections.

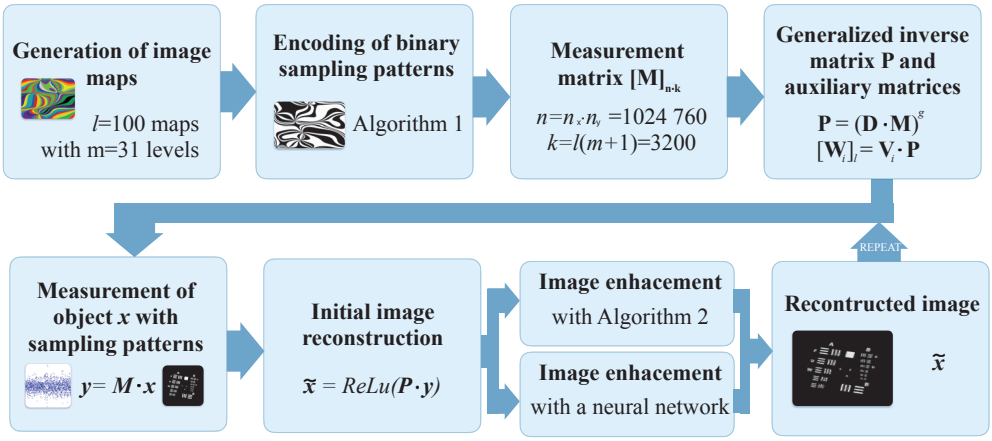


Figure 1. Block diagram of the proposed high-resolution SPI framework. The blocks in the upper row show the preparation stage, involving the calculation of large matrices needed for object sampling and for the reconstruction of the SPI measurement afterward. The blocks in the bottom row describe the actual SPI measurement and reconstruction. In continuous SPI imaging, the bottom blocks are executed cyclically.

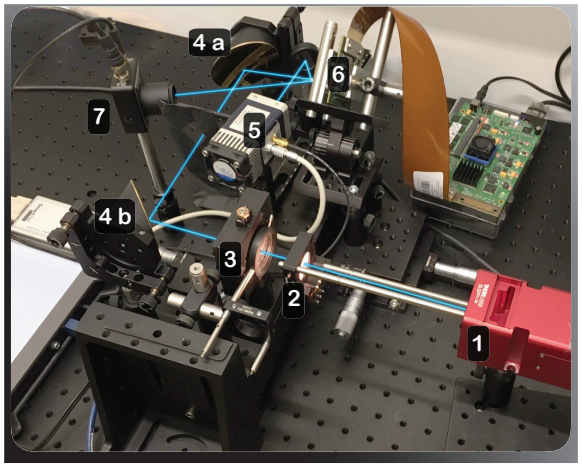


Figure 2. SPI optical setup with a broadband light source, a DMD modulator operating in the native 1024 × 768 binary mode at up to 22 kHz, and two detectors—a VIS/NIR range Si amplified detector and a thermoelectrically cooled MCT detector for the NIR/MIR range. The detectors are placed in a complementary architecture, measuring the signal reflected by the DMD mirrors in either the "zero" or "one" position.

2.1. Constraints and Objectives for the Proposed Sampling Framework: SPI Pattern Requirements for DMD Modulators

The type of image sampling is tied to the hardware used for spatial modulation. If DMD technology is used for this purpose, the sampling patterns should be binary, as the DMD mirrors inherently have only two states. Using grayscale modulation requires introducing spatial or temporal multiplexing, which occurs at the cost of the native DMD resolution or bandwidth [36]. The DMD bandwidth is also harnessed by utilizing dif-

ferential projection [36], a commonly used technique that improves the SNR, removes background light from the detector signal, and enables the display of both the positive and negative patterns present in, for example, Walsh–Hadamard matrices. However, these objectives can largely be achieved by adding a single pattern to those displayed, rather than doubling their number [37].

There are other important challenges introduced in a realistic optical setup that are not present in a purely digital experiment. These include the need for noise-mitigating techniques such as differential or complementary detection [38–40], as well as maintaining a consistent detection signal range throughout the measurement [10]. The latter suggests that it is inadvisable to combine, in a single measurement, sampling patterns where all pixels are in a single state with those where approximately half of the pixels share the same state. Otherwise, the experiment results in the suboptimal usage of the detector's dynamic range and A/D converter's bit depth for signal acquisition. Additionally, a signal obtained with a small number of pixels in the on state tends to be weak in low-light environments and does not take advantage of Fellgett's [41,42] multiplex advantage, which is particularly important in the infrared wavelength range.

To address these hardware-driven requirements, the sampling patterns need to satisfy the following properties:

- Sampling patterns should be binary (with values 0 and 1) because DMD mirrors have two states.
- Sampling patterns should include approximately half of the pixels in the on state (due to the arguments related to light efficiency, Fellgett's advantage, level of quantization errors induced by DAQ, and signal entropy considerations).
- Sampling should be differential, meaning that image reconstruction is based on the differences between measurements with consecutive patterns. This eliminates any constant bias from stray light, the background, or inactive areas of the DMD.
- Sampling patterns should have a spatial spectrum dominated by low spatial frequency contents because most real-world images have their Fourier representation concentrated at low spatial frequencies (this is the usual approach in compressive SPI with, for example, the selection of a subset of Walsh–Hadamard, DCT or wavelet patterns).
- Sampling should provide easily accessible information on the locations of sparse areas of the probed image. Identifying sparse image regions makes it possible to reconstruct dense regions with better accuracy also at a low sampling rate. For instance, random binary patterns would measure mostly the mean value of the entire image. The best would be patterns in the forms of small figures with various shapes.

The last postulate contradicts some of the previous ones. For instance, any small figure certainly includes less than half of the image pixels. This is why we propose a sampling scheme that involves differential multiplexing and is based on image maps. The measurement with these patterns provides data equivalent to a different measurement that would include patterns with small binary figures.

2.2. Sampling Patterns

The process of preparing the sampling patterns and storing them in the measurement matrix (M) is described in Algorithm 1. The entry point to the algorithm is a set of arbitrary image maps. Every image map is a mapping between the image pixels and a subset of integer numbers, thereby defining a partition of the image plane into subareas or regions that have been assigned equal numbers. A single region may include more than one group of adjacent pixels or even isolated pixels. These partitions are introduced in the form of image maps. Here, we use uniformly distributed maps with varying characteristic sizes. The Python code used to generate the maps is the same as the one included in the supplementary materials of Ref. [32]. The maps could be defined in other ways as well; for instance, by analogy to foveated SPI imaging [19,43,44], some regions of interest could be enhanced at the expense of other areas [33]. The typical feature size of the regions can also be aligned with specific object feature sizes and shapes. Sample maps are shown in

Figure 3. We usually take $l = 100$ maps, each with $m = 31$ regions. Then, the algorithm encodes a sequence of $k = l \cdot (m + 1) = 3200$ binary full-resolution 1024×768 patterns. This corresponds to the sampling rate (compression rate, i.e., the ratio of the number of sampling patterns to the number of pixels) of 0.41.

Samples of encoded patterns are shown in Figure 4. Encoding or translating image maps into patterns involves using a lookup table, which is described in the next subsection. The lookup table used throughout this paper is shown graphically in Figure 5.

Algorithm 1 Construction of a binary measurement matrix from image maps.

Input: **m**—$l \times n$ matrix containing l maps with $n = n_x \cdot n_y$ pixels
Input: **r**—vector of length n with random integer values in the range $[1, l]$
Input: **A**—a look-up table; $(m + 1) \times m$ binary matrix , such that $\mathbf{D} \cdot \mathbf{A}$ is full rank. Here **D** is finite difference operator that subtracts matrix rows.
Output: **M**—$k \times n$ measurement matrix with rows containing binary patterns with n pixels each

```
function MAKE PATTERNS(m, r, A)
    M ← initialize and fill with zeros
    p ← 0
    for i = 1, l do                                          ▷ Iterate maps
        m_i ← i-th row of m                                  ▷ i-th map
        for q = 1, (m + 1) do                                ▷ Iterate rows of A
            for j = 1, m do                                  ▷ Iterate sectors of map m_i
                p ← p + 1                                    ▷ Index of the next sampling pattern
                v ← where((m_i = j) and (r ≠ i))             ▷ Get indices of all pixels of j-th sector
of i-th map
                M(p, v) ← A(q, j)                            ▷ Assign binary values to pixels of a sector
            end for
        end for
    end for
    return M̃                                                 ▷ Return the measurement matrix
end function
```

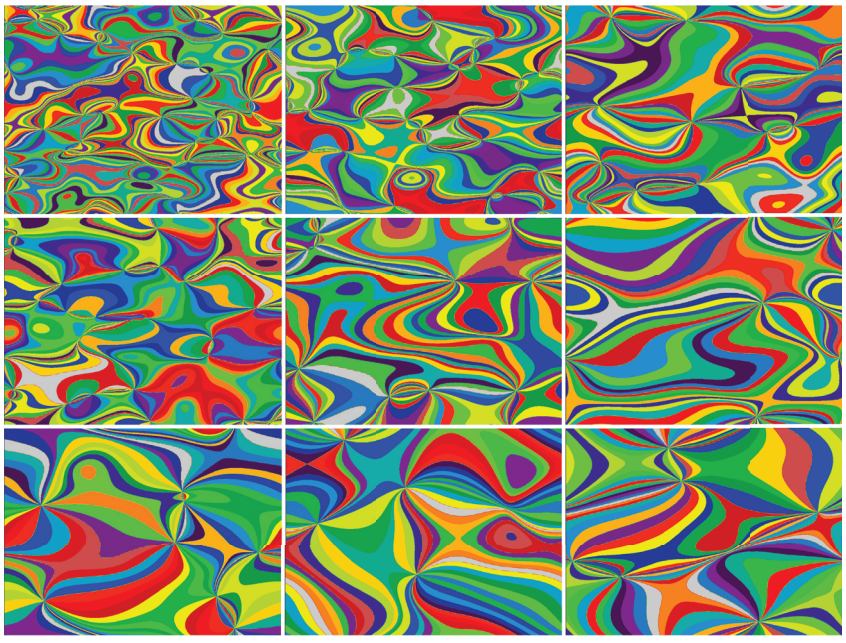


Figure 3. Examples of high-resolution ($n = 1024 \times 768$) image maps. The above 9-element sample is a subset taken from the $l = 100$-element set of maps. Every map includes $m = 31$ different regions shown in different (false) colors. The characteristic feature size is different in each map. Each of the maps is used to create $m + 1$ sampling patterns (see Algorithm 1).

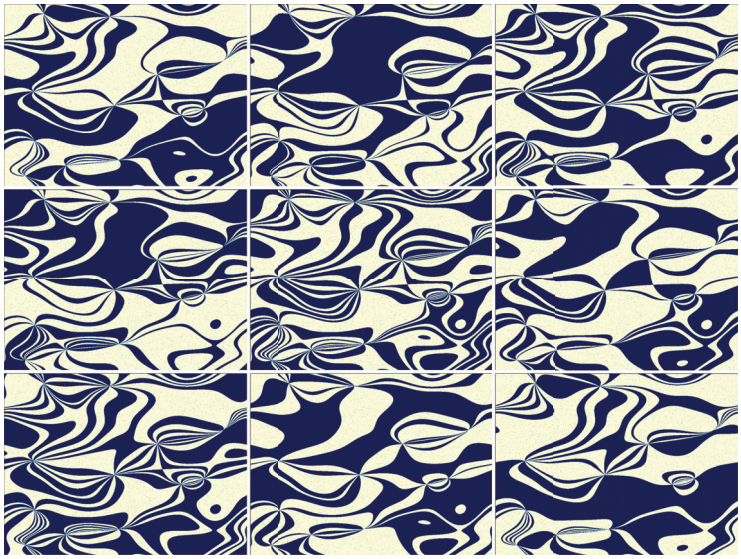


Figure 4. Examples of 1024×768 binary sampling patterns constructed from a single image map (that map is shown in the top-right subplot of Figure 3). The patterns contain approximately the same number of pixels in the on and off states. In total, we calculate $l \cdot m = 3200$ binary sampling patterns, which are then stacked in the measurement matrix $\mathbf{M}$. The sampling patterns are projected onto the DMD in the native binary full-resolution mode. The measurement is differential in the sense that only the differences between subsequent intensity measurements are processed, eliminating the influence of background illumination, yet information about the intensity in each region of every image map is guaranteed to be preserved.

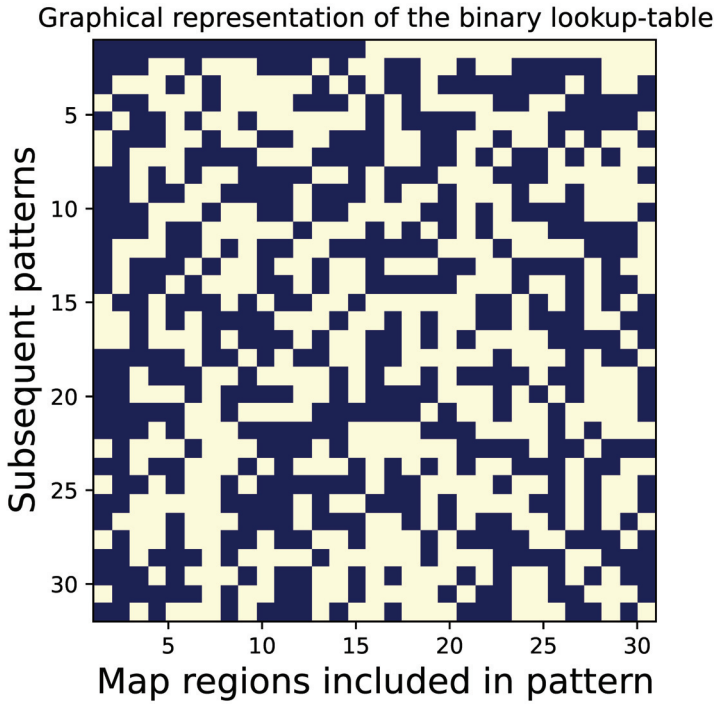


Figure 5. Graphical representation of the binary lookup table $\mathbf{A}$ required by Algorithm 1. $\mathbf{A}$ is a $(m+1) \times m$ matrix with binary values (here, $m = 31$) obtained through numerical optimization [32]. Algorithm 1 uses this matrix to construct $m+1$ subsequent rows of the measurement matrix $\mathbf{M}$ by taking consecutive rows of $\mathbf{A}$ and treating its columns as an indication of which regions from the map should be included in the sampling pattern.

Algorithm 1 is similar, but not identical, to the one we proposed previously [32,33]. There is a small change that affects the way the mean value of the object is encoded

in the sampling patterns. Previously, we dropped some rows from the lookup table when translating the second and subsequent maps into patterns. This approach removes redundancy in the sampling. However, the disadvantage is that it allows the recovery of the mean value of the object from the first $m + 1$ sampling patterns. Since the mean value of the object influences all pixels in the reconstructed image, this type of sampling resulted in a very non-uniform distribution of values in the generalized inverse matrix.

Currently, Algorithm 1 processes all the maps in the same way, distributing the information on the mean value of the object across all sampling patterns. As a result, the inverse matrix has improved uniformity.

2.3. Differential Multiplexing

The mechanism of differential multiplexing is governed by the multiplexed pattern encoding algorithm, with a crucial role played by the dedicated binary lookup table. The translation of image maps into sampling patterns is described by Algorithm 1. Pattern encoding is performed using a lookup table, which is a binary $(m + 1) \times m$ matrix $\mathbf{A}$. The table is scanned row by row, with its columns indicating which map regions should be included in the encoded sampling pattern. The encoding mechanism should ensure that the differences between consecutively created binary patterns convey the same information on the sampled objects as would be obtained through sampling using individual map regions. Therefore, the lookup table must satisfy certain mathematical properties. Specifically, its rows should contain approximately equal numbers of zeros and ones, and a matrix formed by subtracting its rows should be a full-rank square matrix. These conditions cannot be achieved for even m, so the matrix has an odd number of columns. A gallery of lookup tables of different sizes is shown in Figure A1. Despite these conditions, there is still room for further optimization, and we search for matrices with the smallest possible determinant $det(\mathbf{D} \cdot \mathbf{A})$. As a result, the lookup tables we use are derived from brute-force numerical optimization. It is worth noting that the columns of the lookup tables can be freely reordered, but not the rows. In fact, the regions of each map can be reordered as desired.

Finally, we note that in a non-differential measurement, we could use a square Hadamard matrix instead of our lookup tables, as Hadamard matrices are inherently full-rank and have an equal number of the two binary values (with the first row different). In fact, Hadamard matrices have a well-established position in multiplexed measurements. However, Hadamard matrices include both positive and negative values (e.g., ± 1), which introduces additional challenges when displaying them on a DMD. To address this, one would need to double the number of exposures, thus reducing the effective DMD bandwidth. Therefore, the proposed matrices are more suitable in designing an inherently differential multiplex measurement that is equivalent to binary $(0, 1)$ sampling with patterns occupying small parts of the image (see Figure 6) and together forming l independent partitions of the image surface. In many cases, they may be better suited for conducting a multiplexed measurement than Hadamard matrices.

2.4. Initial Image Reconstruction

The measurement of an object in SPI imaging is usually expressed mathematically using a concise measurement equation, which in a noise-free situation may be written as

$$\mathbf{y} = \mathbf{M} \cdot \mathbf{x}. \tag{1}$$

Here, $\mathbf{x}$ denotes the object, such as a 2D image, with pixels gathered into a single vector. $\mathbf{M}$ is the measurement matrix whose rows contain the sampling patterns. Finally, $\mathbf{y}$ is the measurement vector. A refined version of this equation would also include signal-dependent and signal-independent noise. There are various ways to conduct a differential measurement. In the present work, we calculate the difference of subsequent elements in the measurement vector, i.e., $\mathbf{y}' = \mathbf{D} \cdot \mathbf{y}$, and use only $\mathbf{y}'$ for image reconstruction. Here, $\mathbf{D}$ denotes a finite difference operator that subtracts the subsequent rows of a matrix or the

subsequent elements of a vector. Any constant DC component, such as one originating from background light, is eliminated from $\mathbf{y}'$. The SPI reconstruction algorithm should then be able to recover $\mathbf{x}$ from $\mathbf{D} \cdot \mathbf{y}$. In a compressive measurement, where $\mathbf{M}$ is a rectangular matrix with more columns than rows, this recovery is usually ambiguous and approximate.

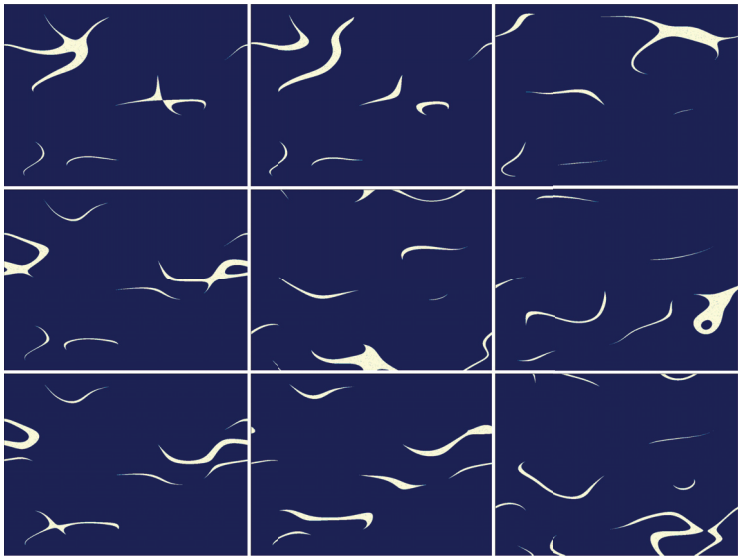


Figure 6. Examples of binary regions extracted from a single image map (that map is shown in the top right subplot of Figure 3, and the sampling patterns shown in Figure 4 include superpositions of the binary, non-overlapping regions of the same map). In a mathematical SPI experiment with no necessity of eliminating a DC background bias and with no limiting effect of the small region size on SNR, the binary regions shown in this figure could be used equivalently to the patterns from Figure 4, and both measurements would acquire exactly the same information on the sampled object. In an optical experiment, the patterns from Figure 4 have the advantage of occupying approximately half of the DMD surface and of the inherent possibility of a differential measurement.

Most SPI image reconstruction algorithms are either iterative or based on applying the pseudoinverse of the measurement matrix to the measured signal, i.e., $\tilde{\mathbf{x}} = \mathbf{M}^{+} \cdot \mathbf{y}$. This is particularly straightforward when the measurement matrix is semi-orthogonal. An even simpler situation arises when it corresponds to a fast linear transform such as DCT, Fourier, or Walsh–Hadamard transform. Our approach is slightly different. We work with the full form of the measurement matrix $\mathbf{M}$ and find its generalized inverse $\mathbf{M}^{g}$. The Moore–Penrose pseudoinverse is an example of a matrix's generalized inverse, but it is not the only one. The calculation of the generalized inverse of $\mathbf{M}$, which is a large matrix, is computationally demanding. It becomes even more difficult as the sampling rate is increased.

$\mathbf{M}^{g}$ is calculated before SPI imaging, and it may require a significant amount of memory (we typically work with 10 GB inverse matrices). However, the image reconstruction is fast because it is based on a simple matrix–vector multiplication. Since we assume a differential measurement, we need an inverse matrix of the form $\mathbf{P} = (\mathbf{D} \cdot \mathbf{M})^{g}$.

For the generalized inversion (^{g}), one could use the pseudoinverse, but it is preferable to use a regularized pseudoinverse. By choosing various generalized inverses of the measurement matrix, we could calculate different solutions of the underdetermined measurement equation. The role of regularization is to select such a solution that additionally minimizes a combined quadratic criterion. In our case, the criterion is defined as a weighted sum of squared norms of the convolutions with selected spatial filters, including discrete gradient filters, a penalty filter for high spatial frequencies, and an identity filter [16]. We use the name Fourier-Domain Regularized Inversion (FDRI) for this kind of matrix inversion because the quadratic criteria are conveniently written in the Fourier space as positive definite diagonal quadratic forms. With FDRI, it is possible to reconstruct images from a

compressive SPI measurement with fewer artifacts than with a simple pseudoinverse. In Ref. [16], FDRI was applied for real-time SPI imaging. The quality of the reconstructed images was comparable to that obtained by the minimization of the total-variation (TV) norm, while the image reconstruction time was almost two orders of magnitude smaller. The form of the generalized inverse used here is the same as in Ref. [16] with the only difference being that we calculate the generalized inverse of $\mathbf{D} \cdot \mathbf{M}$ [10,32] and not just of $\mathbf{M}$:

$$\mathbf{P} = \mathbf{F}^* \cdot \hat{\mathbf{\Gamma}} \cdot \mathbf{F} \cdot (\mathbf{D} \cdot \mathbf{M} \cdot \mathbf{F}^* \cdot \hat{\mathbf{\Gamma}} \cdot \mathbf{F})^+ \cdot \mathbf{D}. \tag{2}$$

Here, $\mathbf{M}$ is the measurement matrix, $\mathbf{F}$ is the 2D Fourier transform, $\mathbf{D}$ is the finite difference operator, and $\hat{\mathbf{\Gamma}}$ is a diagonal non-negative regularizing term [16,32]. Matrix $\hat{\mathbf{\Gamma}}$ is defined as in Ref. [16],

$$\hat{\Gamma}_{i,j} = \frac{\delta_{i,j}}{\sqrt{(1-\mu)^2(sin^2(\omega_x) + sin^2(\omega_y)) + \mu^2 \frac{\omega_x^2+\omega_y^2}{2\pi^2}} + \epsilon}, \tag{3}$$

where μ and ϵ are used to weight the criteria included in the regularization, and $\omega_{x,y}$ are the spatial frequencies. We assume that $\mu = 0.5$ and $\epsilon = 10^{-7}$.

The reconstructed image is then calculated as

$$\tilde{\mathbf{x}} = ReLu(\tilde{\mathbf{x_0}}), \tag{4}$$

where

$$\tilde{\mathbf{x_0}} = \mathbf{P} \cdot \mathbf{y}. \tag{5}$$

The *ReLu* function removes the negative part from the reconstructed image, which we consider to be a reconstruction artifact. At lower resolutions, $\tilde{\mathbf{x}}$ could be the final result of SPI imaging. Since we are interested in high-resolution SPI, this first-stage reconstruction serves as the starting point for further enhancement with Algorithm 2 or an NN.

Algorithm 2 Image reconstruction from a compressive measurement.

Input: $\mathbf{y}$—measurement vector of length k (it i assumed that $k \ll n$, where $n = n_x \cdot n_y$ is the image size in pixels, and that the measurement equation is $\mathbf{y} = \mathbf{M} \cdot \mathbf{x}$, where $\mathbf{x}$ is the measured image, and $\mathbf{M}$ is the binary measurement matrix)
Input: $\mathbf{V}$—$l \times m \times n$ binary array such that $V_{i,j,p} = 1$ at pixels p belonging to region j of map i unless $r_p = i$. In further notation, $\mathbf{V_i}$ ($i \in \{1, 2, ..., l\}$) represents a $m \times n$ slice of $\mathbf{V}$ which corresponds to the i-th map.
Input: $\mathbf{W}$—array of l matrices $\mathbf{W_i} = \mathbf{V}_i \cdot \mathbf{P}$ ($i \in \{1, 2, ..., l\}$) with dimensions $m \times k$. Here $\mathbf{P} = (\mathbf{D} \cdot \mathbf{M})^g$ is the generalized matrix inverse ^{g} applied to the measurement matrix $\mathbf{M}$ after taking the differences of its rows with operator $\mathbf{D}$
Input: $\tilde{\mathbf{x}}$—the initial image reconstruction vector of size $n = n_x \cdot n_y$ (set by us to the initial reconstruction result $ReLu(\mathbf{P} \cdot \mathbf{y})$ but may also be filled with constant positive values)
Input: $f \in (0, 1)$—learning rate (we took $f = 0.75$)
Output: $\tilde{\mathbf{x}}$—vector of size n with the reconstructed image

function IMAGE RECONSTRUCT($\mathbf{y}$, $\mathbf{W}$, $\mathbf{V}$, $\tilde{\mathbf{x}}$, f)
  for $i = 1, maxiter$ **do**
    for $j = 1, l$ **do** ▷ Loop over maps
      $\mathbf{w}_j \leftarrow \mathbf{W}_j \cdot \mathbf{y}$ ▷ Expected pixel sums in sectors of map j
      $\mathbf{v_j} \leftarrow \mathbf{V}_j \cdot \tilde{\mathbf{x}}$ ▷ Current pixel sums in sectors of map j
      $\mathbf{c} \leftarrow \mathbf{w}_j / \mathbf{v_j}$ where $v_j \neq 0$ else 0
      $\tilde{\mathbf{x}} \leftarrow (1 - f)\tilde{\mathbf{x}} + f \cdot diag(\mathbf{c} \cdot \mathbf{V_j}) \cdot \tilde{\mathbf{x}}$
    end for
  end for
  return $\tilde{\mathbf{x}}$ ▷ Return the reconstructed image
end function

2.5. Reconstruction Enhancement with an Iterative Algorithm

The proposed Algorithm 2 enhances the initially reconstructed image by leveraging two key elements. First, since the initial reconstruction is based on the generalized inverse of the measurement matrix, $\mathbf{P} \cdot \mathbf{y}$ exactly fulfills the measurement Equation (1). Second, the composition of the binary measurement matrix $\mathbf{M}$ ensures that differential sampling with patterns from this matrix is mathematically equivalent to using patterns that correspond to all individual regions of all image maps. In other words, in the absence of measurement noises, the reconstructed image is guaranteed to have the correct mean values within every region of every map. Empty regions will have the mean value of zero. Because the images have only non-negative pixel values, a mean value of zero (or near zero) signifies that the region is empty.

Algorithm 2 iteratively processes all image maps, applying corrections to the mean values of pixels within the regions of each map (this is the "Loop over maps" in Algorithm 2). Because these regions are non-overlapping, they can be processed in parallel. A similar reconstruction method was used in Ref. [32]; however, here we introduce a parameter, the learning rate $0 < f < 1$, which typically improves convergence in the presence of noise. In this paper, we assume $f = 0.75$.

An important property of this algorithm is its efficient elimination of empty (zero-valued) regions from the reconstructed image. As a result, for spatially sparse images, it improves the image quality in regions containing objects. This improvement is achieved because the limited information in the compressed measurement $\mathbf{y}$ is directed toward reconstructing only the non-empty regions of the image. This approach can be viewed as complementary to adaptive sampling; however, adaptive sampling is technically more complex, requiring additional logic in the measurement stage and posing challenges in generating adaptively high-resolution patterns at the DMD frame rate. In contrast, our approach relies on predefined patterns for measurement and separates the measurement process from the reconstruction phase.

2.6. Reconstruction Enhancement with a Neural Network

In an alternative approach to Algorithm 2, we explore the application of an NN to enhance the reconstruction quality (see Figure 7). The proposed network bears some resemblance to the U-Net architecture; however, instead of concatenation in the skip connections, we employ weighted data summation, where the weights are learned parameters of the network. This modification enables the network to balance the influence of input data and extracted features on the final output, placing it in a functional space between a U-Net and an autoencoder.

The NN is trained simultaneously on two handcrafted image datasets. The first dataset comprises binary, high-resolution images of line art, while the second dataset contains images of randomly placed and scaled handwritten digits drawn from the MNIST dataset [45]. The training dataset consists of 30,000 images, drawn equally from both datasets, and is randomly regenerated in each training epoch to enhance the variety of the training data. The evaluation dataset, which is precalculated, contains 6000 images, also drawn equally from both datasets. Notably, the handwritten digits included in the evaluation dataset are not used for generating the training images. All images in both datasets have a resolution of 768×1024 pixels and are normalized to pixel values in the range [0, 1]. The generated images contain 96% of zero-valued pixels on average (ranging between 89% and 99.7%).

The NN is designed to perform the second-stage reconstruction as an alternative to using Algorithm 2. The input data consist of initially reconstructed images obtained by applying Equation (5) to simulated SPI measurements of the training or testing images. To model realistic experimental conditions, we introduce additive Gaussian noise to the measurement. The network is trained to optimize a joint loss function:

$$C = C_1 + \lambda \cdot C_2, \quad \text{where } C_1 = MSE(\hat{x}, x), \;\; C_2 = MSE(\hat{v}, v), \tag{6}$$

where *MSE* denotes the mean squared error, x represents the pixel values of the ground truth image, $\hat{x}$ is the output of the NN, v and $\hat{v}$ are the sums of pixel values within all sectors of the maps calculated for images x and $\hat{x}$, respectively, and λ is a weight parameter empirically adjusted during training.

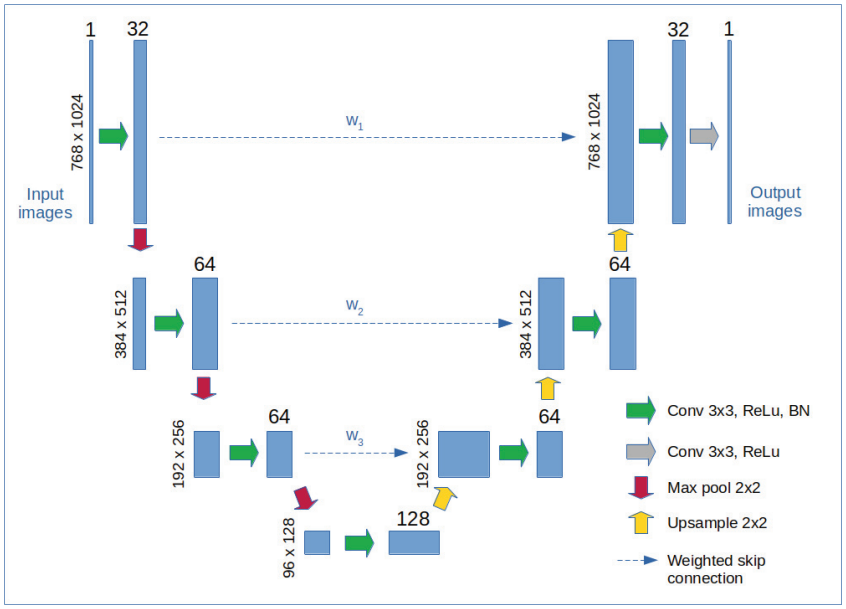


Figure 7. Block diagram of the NN architecture. The network consists of 8 convolutional blocks arranged in a U-Net architecture, with the first 4 blocks forming the contracting path and the remaining 4 blocks forming the expanding path. Each convolutional block comprises a 3×3 convolutional layer, a rectified linear unit (ReLU) nonlinear activation, and a batch normalization (BN) layer. Two-dimensional max pooling is applied in the contracting path to reduce the spatial size of the data, while nearest-neighbor interpolation is used in the expanding path to upsample the data. Additionally, weighted skip connections, with learned weights w_1, w_2, w_3, combine features from matching blocks in both paths to improve data fidelity. The skip connections are applied directly after the convolutional layers and before the ReLU activations.

2.7. Optical Set-Up

We validate the performance of the proposed algorithm by implementing the SPI optical setup using a 1024×768 pixel DMD (Vialux V-7001 XGA with DLP7000), which splits the light beam into two separate detectors: an uncooled Thorlabs PDA100A2 silicon photodiode for the 300–1100 nm range, and a thermoelectrically cooled VIGO Photonics PVI-4TE-3 detector for the 1100–3000 nm range. The schematic is shown in Figure 2. Light from a thermal source passes through a test pattern composed of a series of holes in a nontransparent metal foil, with the object observed in transmission mode.

To minimize chromatic aberration, we first use a set of gold-coated parabolic mirrors as the imaging system. We also test an alternative setup using fused silica lenses. The parabolic mirrors provide better signal-to-noise ratio (SNR) for the infrared (IR) optical channel, while the glass lenses offer higher image resolution in the visible light range. The signal is digitized using a 14-bit digital oscilloscope (Picoscope 5000) with a sampling at 128 ns. Since the DMD frequency is lower, the measurement vector $\mathbf{y}$ is obtained from these samples by averaging.

3. Results

3.1. Image Acquisition and Reconstruction Times and Computational Requirements

Throughout this paper, we assume an image resolution of 1024×768, with the object being sampled at a compression rate of 0.41%, requiring 3200 binary patterns for sampling. The DMD operates at 22 kHz, resulting in an image acquisition frequency of 6.9 Hz.

Our image reconstruction algorithm is implemented in Python, utilizing the NumPy and PyTorch libraries, and is capable of running on both CPU and GPU processors. Due to the high memory demands, we use a GPU with 24 GB of memory (NVIDIA GeForce RTX 3090). Specifically, the measurement matrix **M** and the inverse matrix **P** each occupy 10 GB in 32-bit floating-point format. This suggests that further increasing the number of sampling patterns at the same resolution may present challenges for this framework and could require either a scanning strategy or parallel SPI measurements in lower-resolution blocks.

The initial reconstruction takes 16 ms and can be performed on the fly without a noticeable impact on the DMD acquisition rate, reducing the rate from 6.9 Hz to 6.2 Hz. The second-stage reconstruction using the NN takes 26 ms (including the initial reconstruction stage), resulting in an SPI rate of 5.8 Hz. The iterative algorithm is slower, requiring approximately 50 ms per iteration, which limits real-time applications to one or two iterations; performing 50 iterations would take 2.5 s. Additionally, due to the GPU architecture, images from two or more channels can be reconstructed in parallel without additional time penalty, leveraging the batch processing implemented in PyTorch. Without a GPU, the processing times are significantly longer. On an Intel i7-11700K CPU, the initial reconstruction takes 350 ms, and the iterative algorithm is much slower, taking 2.2 s per iteration (however, the computation speed could improve by up to tenfold with a Numba-based implementation [32]). In summary, for applications that are not time critical, Algorithm 2 may be applied with multiple iterations. At the same time, only a GPU-based approach—either using the second-stage reconstruction with the NN or a single iteration of Algorithm 2—enables image acquisition and reconstruction at over 6 Hz at the resolution of 1024×768 without hindering the DMD acquisition rate.

3.2. Effect of Compression, Background Noise, and Detector Noises on Imaging

Eliminating additive background noise is critical in optical SPI. This noise may include a constant or slowly varying bias in the detector signal due to the dark current, background illumination, or reflections from inactive parts of the DMD, as well as variations in the light source intensity. In infrared imaging, the changing temperature of the setup, and particularly of the DMD, also contributes to a slowly varying additive background signal. Background noise is usually reduced through differential or interferometric techniques, such as projecting negated DMD patterns or using the complementary detection of the two beams reflected by the DMD in the directions corresponding to the on and off states of the mirrors. In our case, we incorporate the differential mechanism in the way we construct sampling patterns and measurements. Thus, the reconstructed image is invariant with respect to adding any constant value to the detection signal. This approach also allows us to use the AC coupling of the DAQ, although this type of high-pass filtering is not entirely transparent to the image reconstruction outcome.

A variety of other noise sources also affect SPI imaging, and the impact of one source may often become dominant. We now discuss the effect of compression and detector additive noise. The impact of compression on SPI imaging quality depends on the structure of the object. In our work, the compression rate is extremely small, and the reconstruction quality depends on the presence of empty (zero-valued) areas within the object image. To our knowledge, there are no existing test image datasets categorized by spatial sparsity, making it difficult or impossible to use standard image archives for the analysis in this paper. Consequently, we first select three images with different properties to illustrate the typical behavior of the presented SPI method in the presence of different sparsity levels. We then create synthetic images based on the MNIST database with added Gaussian noise for further comparisons.

In Figures 8 and 9, we demonstrate the role of spatial sparsity in SPI imaging with three example images: the Siemens star, which is a dense grayscale test object without empty regions (with less than 0.1% zero-valued pixels); a limited-field-of-view Siemens star (with 96.3% of zero-valued pixels); and a sparse line-art image (with 98.9% of zero-valued pixels). In this analysis, due to the absence of other noises, compression is the only factor responsible for imperfect imaging. An approximate reconstruction result is obtained with the initial reconstruction method, similar to what could be achieved with low-resolution SPI imaging

(e.g., 128 × 128 or 256 × 256 at most). For the grayscale image, no significant improvement is achieved with the iterative algorithm. However, Algorithm 2 significantly improves the quality of the other two images, achieving up to the DMD resolution (1024 × 768). In Figure 8, we also measure the reconstruction quality using two standard criteria: peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). The notable differences in these criteria reveal that averaging these measures across a range of standard database images would obscure information about the properties of imaging sparse objects.

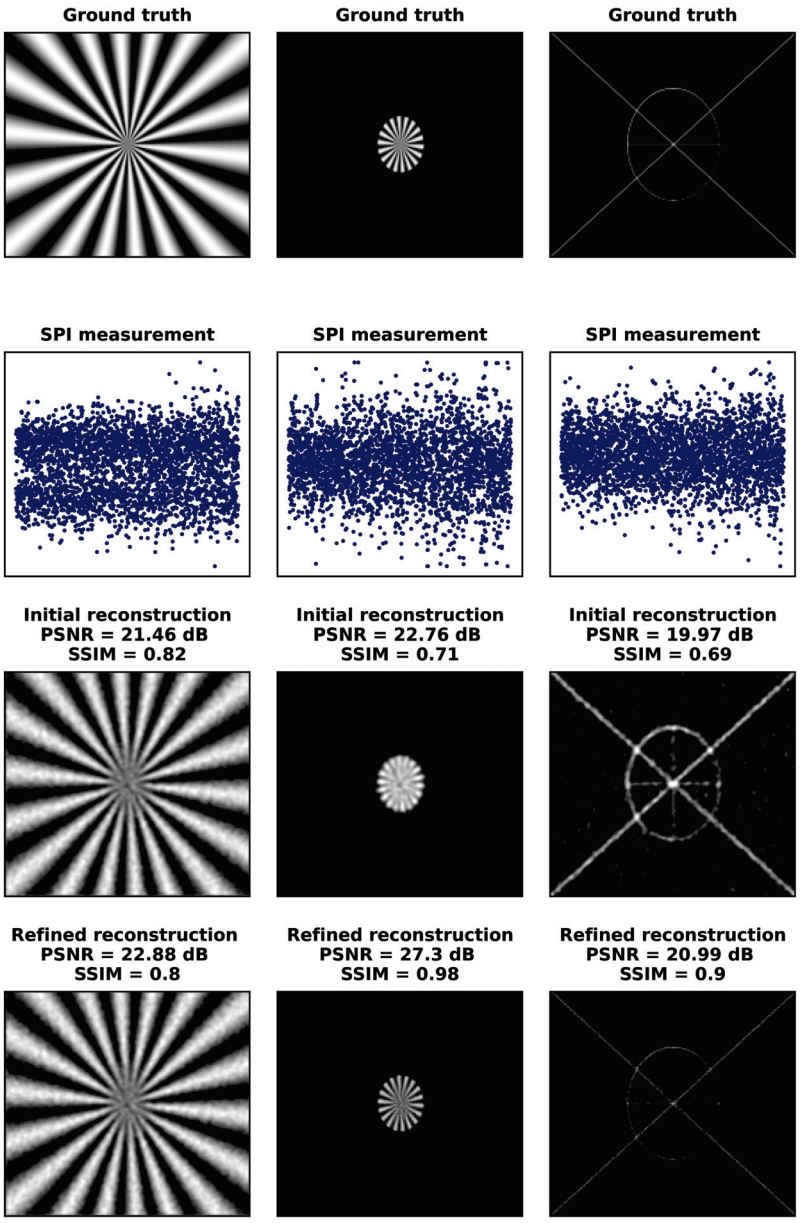


Figure 8. High-resolution (1024 × 768) SPI imaging at very high compression ($k/n = 0.41\%$). The three columns show examples of three different types of objects: **left**—grayscale image (Siemens star) occupying the entire image area, **center**—grayscale image with a limited support region, and **right**—binary, high-resolution, highly sparse line-art image. The rows, starting from the top, show the ground truth, the SPI measurement data, the first-stage reconstruction result, and the second-stage improved reconstruction result (50 iterations). The reconstruction quality of sparse images is significantly improved in the second reconstruction stage with Algorithm 2.

In Figure 9, we show the convergence of PSNR and SSIM for the same three objects as a function of the number of iterations in Algorithm 2. In addition to the data presented

with solid lines, we show the same convergence when the initial reconstruction stage is omitted, which slightly improves the computation time and significantly reduces the memory requirements due to the large size of the inverse matrix **P**. Results from this plot indicate that typically the first 1–5 iterations are sufficient to nearly reach the final image quality. The importance of the first stage reconstruction largely depends on the type of image; it is especially crucial for the grayscale Siemens star, although it also plays a role in achieving convergence of PSNR and SSIM for the other two objects.

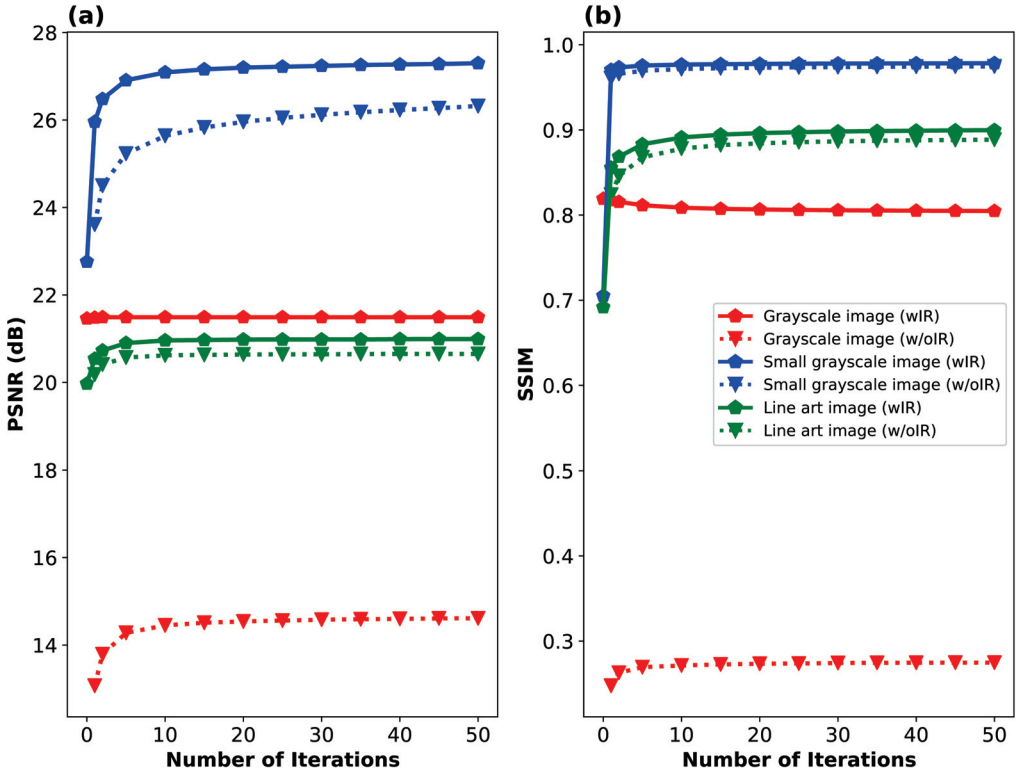


Figure 9. Convergence of image reconstruction quality in terms of two metrics: (**a**) PSNR and (**b**) SSIM as a function of the number of iterations in Algorithm 2. Two variants of the reconstruction algorithm are considered: one with an initial reconstruction calculated using Equation (4) (wIR), and the other without an initial reconstruction, where the vector $\tilde{x}$ is initialized with constant positive values (w/oIR). The performance is evaluated on three distinct images illustrated in Figure 8: a full-area grayscale image (Siemens star), a smaller grayscale image with limited field of view, and a high-resolution sparse line-art image.

The role of additive detector noise is depicted in Figures 10–12. An often overlooked source of additive noise is the limited bit resolution of the detection system introduced by the DAQ. Depending on the sampling patterns, the required dynamic range of the detector may be utilized more or less efficiently. A common issue in SPI is that, aside from patterns containing approximately half of the pixels in either the on or off states, there also appear frames with all pixels switched on or off. As a result, the quality of the measurement depends on fine signal changes between consecutive projected patterns, yet some signal points deviate significantly from the typical values, which imposes unnecessary constraints on the DAQ bit resolution. We avoid this issue by properly constructing the measurement matrix. Nevertheless, the DAQ bit resolution does influence the final SPI quality as shown in Figure 10, which illustrates the relationship between PSNR and SSIM versus the DAQ bit resolution.

We note that due to its randomness, DAQ quantization error exhibits properties similar to those of additive Gaussian noise, unless very few bits are effectively used in the signal's quantization. According to the results in Figure 10, at resolutions below 12 bits, the limited

bit resolution begins to impact the image quality more than compression. For this reason, in the optical experiments, we use at least 14-bit quantization in the measurement.

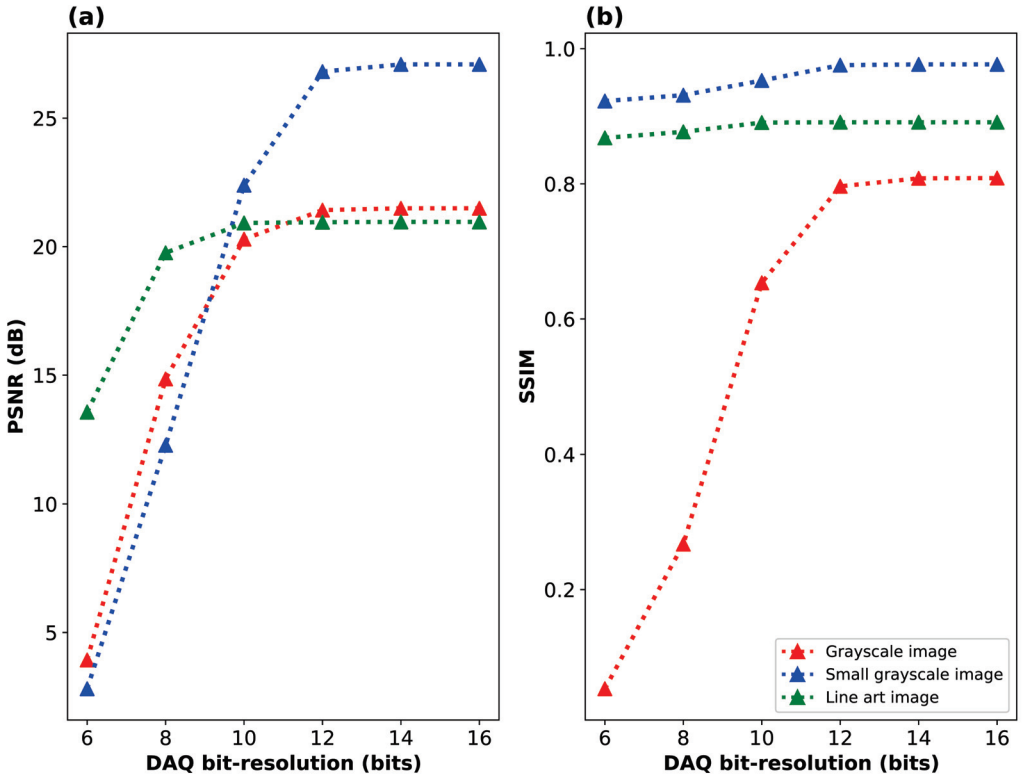


Figure 10. Performance of Algorithm 2 as a function of the bit resolution of the data acquisition device in terms of two metrics: (**a**) PSNR and (**b**) SSIM. The evaluation is performed on three distinct images illustrated in Figure 8: a full-area grayscale image (Siemens star), a smaller grayscale image with a limited field of view, and a high-resolution sparse line-art image.

A comparison of the performance of the two second-stage reconstruction methods (Algorithm 2 and NN) in the presence of additive Gaussian detector noise is shown in Figure 11. This comparison is performed on a set of images similar (but not identical) to those included in the NN training set. At an SNR of 45 dB, both methods perform well and achieve high-resolution SPI imaging, with the NN method being significantly faster and yielding better quantitative results in terms of SSIM and PSNR. We use the standard definition of SNR, but since the measurement is differential, we subtract the mean value from the signal before calculating SNR. In fact, adding any constant value to the measurement vector does not affect the reconstructed image.

For images that differ from the training data, the NN often introduces significant artifacts, while Algorithm 2 maintains good performance even for arbitrarily designed images. In Figure 12, we compare these two methods more systematically, using PSNR, SSIM, and the C_1 and C_2 loss function components, averaged over a set of 1000 images and plotted against the measurement SNR.

3.3. Experimental Results

Real-world visible and infrared scenes often include dense grayscale objects. While the SPI method proposed in this paper can be applied to dense images, high-resolution imaging is only achievable for sparse objects. For dense objects, an alternative measurement matrix—such as one based on binarized DCT patterns—may be more suitable and could improve the overall quality of reconstruction. Sparse objects, in this context, include those obtained through phase contrast or polarization contrast, as well as objects observed over a

known background that can be subtracted from the measurement, or scenes with the field of view reduced by vignetting to a selected region of interest.

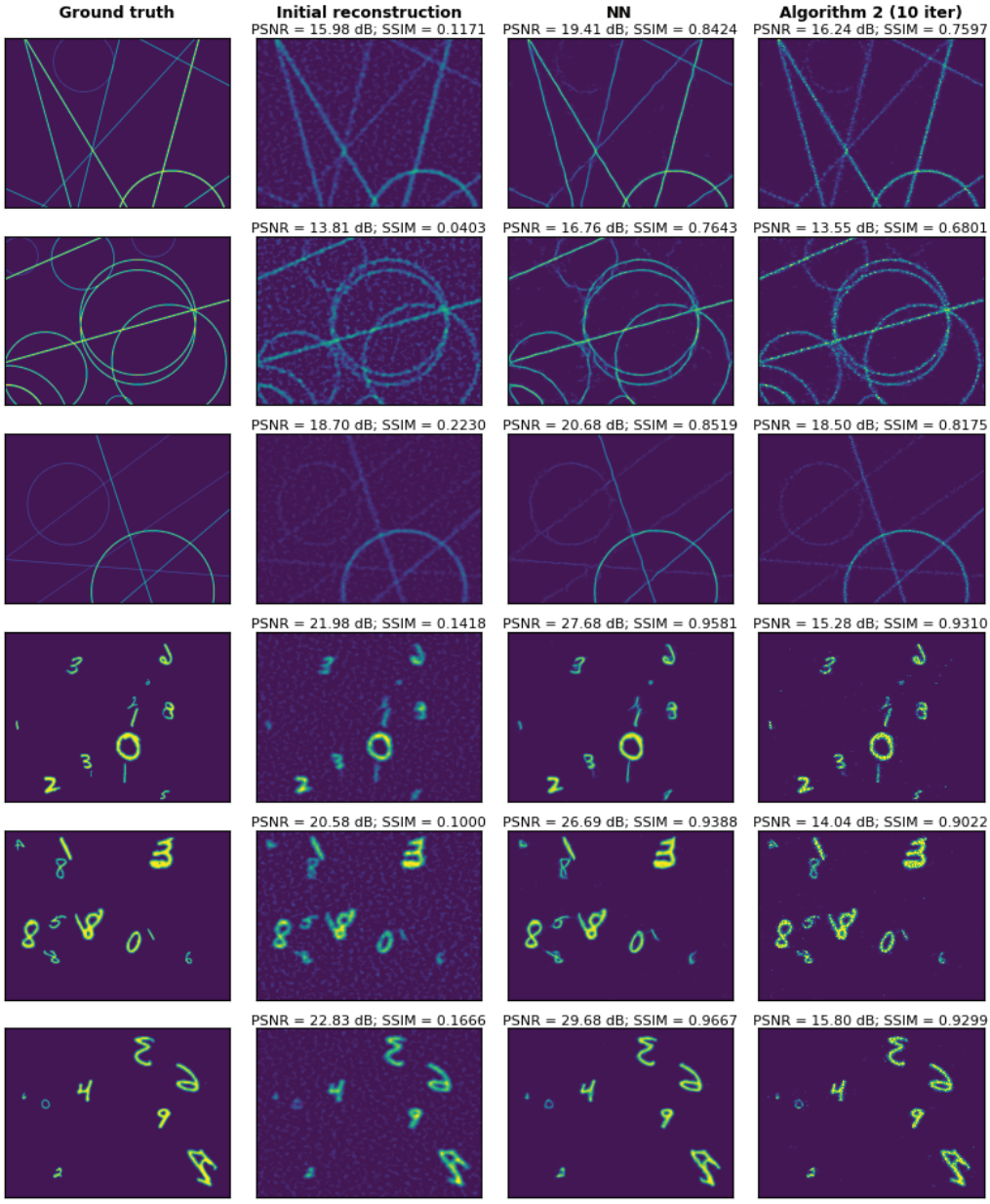


Figure 11. Examples of image reconstructions obtained using either Algorithm 2 or the NN as the second-stage reconstruction method. In this case, a simulated compressive SPI measurement is performed using several exemplary images from the NN evaluation dataset as objects. Additive Gaussian detector noise (with SNR = 45 dB) is included in the measurement. Columns, starting from the left, display the following: (1) the ground truth images, (2) the initial reconstructions obtained using Equation (5), (3) reconstructions enhanced with the NN, and (4) reconstructions enhanced by applying 10 iterations of Algorithm 2.

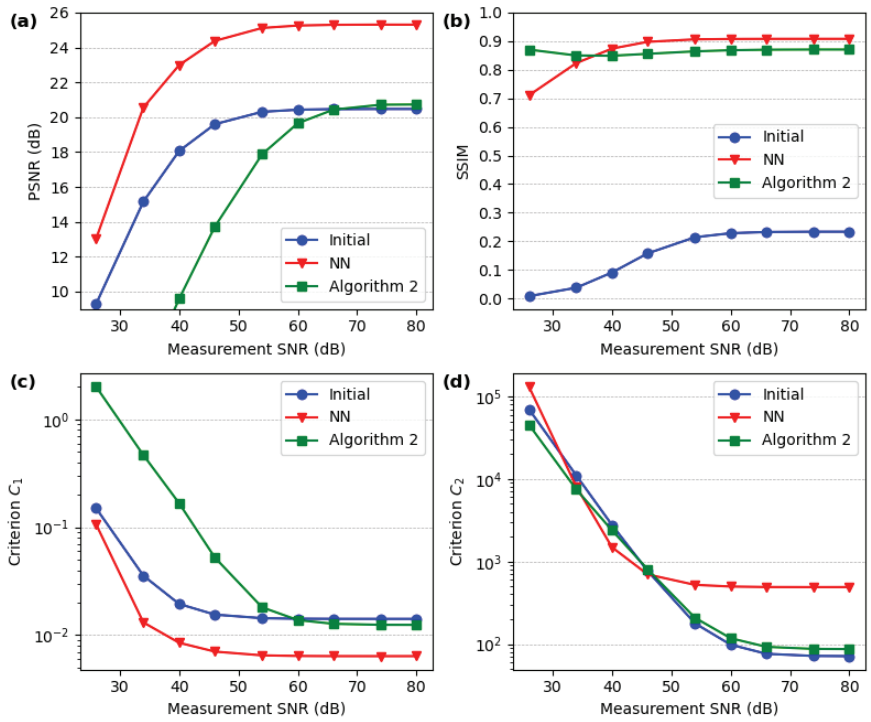


Figure 12. Sensitivity of the proposed image reconstruction methods to additive SPI measurement noise. We evaluate the quality of the initial reconstructions and the second-stage reconstructions, obtained with either Algorithm 2 or the NN, based on four criteria: (**a**) the peak signal-to-noise ratio (PSNR), (**b**) the structural similarity index (SSIM), and (**c**,**d**) the optimization criteria C_1 and C_2 from Equation (6). The metrics are averaged over a subset of 1000 images from the evaluation dataset, which consists of an equal mix of line art and handwritten digit images.

In our example, a sparse object is represented by a metal foil with small holes arranged in the shape of the faculty logo. In our experimental setup, two measurement vectors are captured simultaneously by two detectors operating in different wavelength ranges. These detectors measure the light intensities reflected from the DMD in two directions corresponding to the on and off states of the micromirrors. Such an arrangement could support complementary detection—a technique for enhancing the signal-to-noise ratio (SNR) and removing the DC component from the signal. Since our measurement is inherently differential, we can utilize the two channels independently. Given that the image reconstruction method is invariant to adding a constant to the signal, data from both channels are processed identically, with the only difference being that signals obtained with mirrors in the negated state (in relation to the measurement matrix, in a boolean sense) also need to be negated (in the sense of multiplication by -1). Grayscale images from the two channels are then used to construct a false-color image.

Figure 13 shows the reconstruction result of the logo image. Figure 14 presents the same image measured during movement. Because SPI is a time-multiplexing technique, non-static objects tend to produce artifacts. Although the resolution is affected, the contents remain clearly visible.

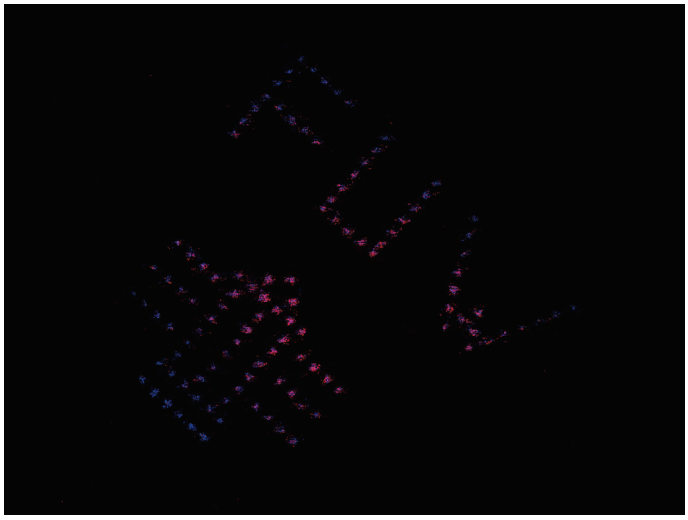


Figure 13. Experimental high-resolution SPI imaging. The object consists of a metal foil with holes arranged in the Faculty of Physics logo of the Univ. of Warsaw. The false colors, red and blue, correspond to the NIR and VIS channels, respectively. The object is static.

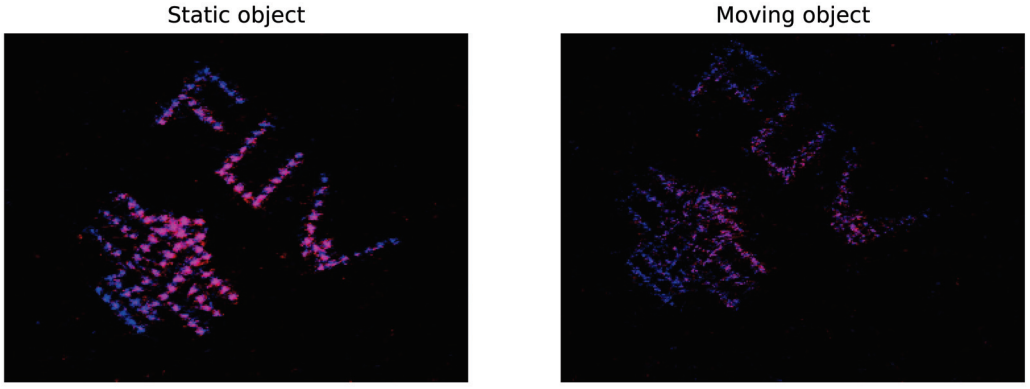


Figure 14. Experimental high-resolution SPI imaging of a static object (**left**) and of a moving object (**right**). For better visibility, the contrast of the false color images is enhanced. Object movement during the acquisition of the measurement vector results in deteriorated reconstruction quality and artifacts.

4. Discussion

We propose and compare two methods for SPI imaging at the native DMD resolution. Both methods use the same differential multiplexing sampling scheme and an initial reconstruction stage based on the dot product with the regularized generalized inverse of the measurement matrix (according to Equation (4)). The second reconstruction stage is either based on an iterative Algorithm 2 or an NN, inspired by the convolutional U-Net architecture as shown in Figure 7. The proposed SPI approaches rely on extremely high compression, which is dictated by the limited bandwidth of the DMD modulators. This restricts the possibility of high-resolution reconstruction to spatially sparse images with limited information content. Dense grayscale images, on the other hand, can only be reconstructed with lower accuracy.

For NN-based reconstruction, high-quality results are achieved only for images similar to those in the training set. The NN is trained on line-art images and transformed alphanumeric objects from the MNIST dataset, and may not perform well on other types of objects or images at different resolutions without further training or modifications. In contrast, Algorithm 2 can be directly applied to different resolutions and sparse images with various shapes of the zero-valued areas.

The proposed method is validated in an optical setup with two detectors operating in the VIS and NIR wavelength ranges. The object is sampled at a compression of 0.41%. Both the NN and Algorithm 2, when used with a single iteration, can reconstruct images at a rate that does not hinder the acquisition frame rate. However, Algorithm 2 with multiple iterations can still be applied in practical applications, with a time penalty of 50 ms per iteration on a GPU.

One advantage of Algorithm 2 is its ability to identify empty regions of the object, which helps in improving the reconstruction accuracy in areas containing actual content. In this way, the proposed high-resolution SPI framework offers an alternative to adaptive sampling methods, which are challenging to implement. Given the low resolution of most current SPI systems (typically between 32×32 and 256×256), which limits their potential to replace traditional high-resolution cameras, we consider this work an important step toward enabling practical SPI applications across a variety of fields. Applications involving the imaging of sparse objects, in particular, may benefit from this work. These include imaging with phase contrast or polarization contrast, as well as cases where objects are observed against a known background that can be subtracted from the measurement. Additionally, it has potential applications in the microscopic imaging of scenes where the field of view is reduced by vignetting to a selected region of interest.

Author Contributions: Methodology, R.K. and A.P.; experimental set-up, R.S. and P.W.; neural network A.P.; other software, R.K., A.P. and R.S.; validation, M.C., K.S. and A.P.; dataset resources, K.S.; writing, review and editing, R.K., A.P., R.S. and P.W.; visualization, M.C., A.P., K.S. and R.S.; supervision, R.K.; project administration, A.P.; funding acquisition, A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly funded by National Science Center, Poland UMO-2019/35 /D/ST7/03781 (A.P.).

Data Availability Statement: Datasets with image maps, sampling patterns, generalized inverse matrix, NN model, etc., may be downloaded from the repository of the University of Warsaw [46]. We will release the Python code in a short time after the paper is published. The link to the code will be put in the repository [46]. Prior to this, we may provide parts of the code upon reasonable request. The code for generating image maps is similar to the one included in Supplement 1 of Ref. [32].

Conflicts of Interest: Author Rafał Stojek was employed by the company VIGO Photonics. 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.

Abbreviations

The following abbreviations are used in this manuscript:

SPI	Single-Pixel Imaging
DMD	Digital Micromirror Device
NN	Neural Network
PSNR	Peak Signal to Noise Ratio
SSIM	Structural Similarity index
MSE	Mean Square Error
FDRI	Fourier Domain Regularized Inversion
MD-FDRI	Map-based Differential FDRI
SNR	Signal-to-Noise-Ratio
GPU	Graphical Processing Unit

Appendix A

In Figure A1 of this Appendix, we present a gallery of the optimized lookup matrices for binary differential multiplexed measurements. These matrices were originally introduced in [32]. In this presentation, in addition to displaying them graphically, we apply a systematic column ordering to the matrices to facilitate better comparison among

them and with Hadamard matrices. The role of the lookup tables in designing differential multiplexed measurements is discussed in Section 2.3.

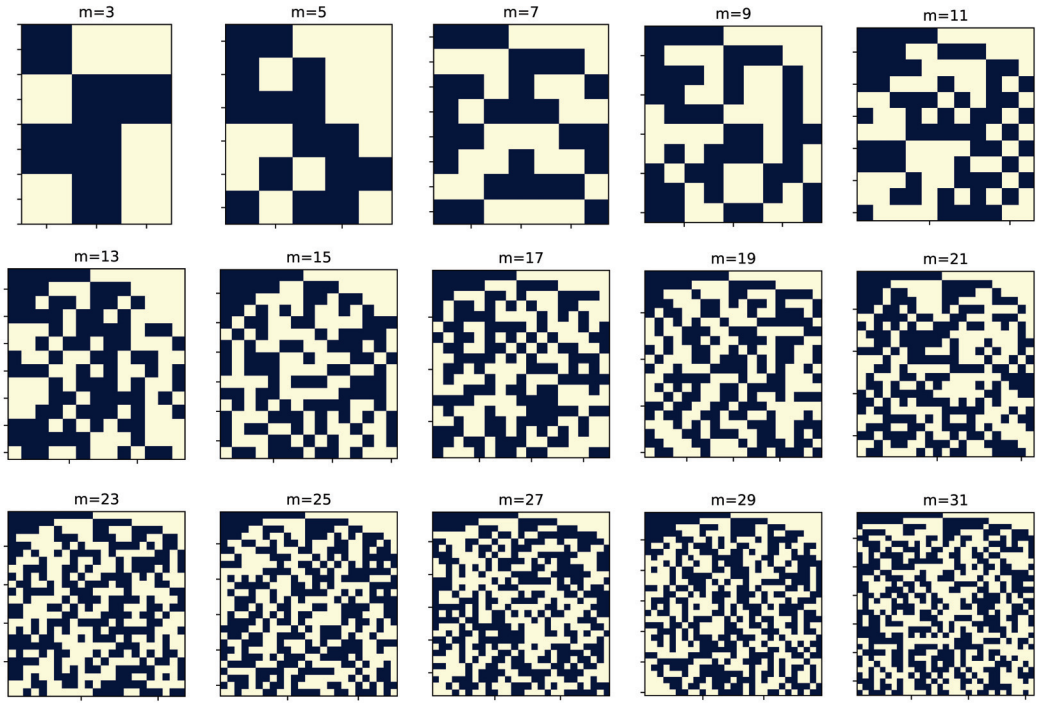


Figure A1. A gallery of lookup tables of different sizes for differential multiplexed measurements. Note that the last matrix, i.e., for $m = 31$, is equivalent to the one in Figure 5 but with a different column ordering. In this figure, the columns are ordered by sorting the values in rows starting from the top row (but preserving the rows sorted earlier.)

References

1. Duarte, M.F.; Davenport, M.A.; Takhar, D.; Laska, J.N.; Sun, T.; Kelly, K.F.; Baraniuk, R.G. Single-pixel imaging via compressive sampling. *IEEE Sig. Proc. Mag.* **2008**, *25*, 83–91. [CrossRef]
2. Candes, E.J.; Romberg, J.K.; Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **2006**, *59*, 1207–1223. [CrossRef]
3. Osorio Quero, C.A.; Durini, D.; Rangel-Magdaleno, J.; Martinez-Carranza, J. Single-pixel imaging: An overview of different methods to be used for 3D space reconstruction in harsh environments. *Rev. Sci. Instrum.* **2021**, *92*, 111501. [CrossRef] [PubMed]
4. Gibson, G.M.; Johnson, S.D.; Padgett, M.J. Single-pixel imaging 12 years on: A review. *Opt. Express* **2020**, *28*, 28190–28208. [CrossRef]
5. Edgar, M.P.; Gibson, G.M.; Padgett, M.J. Principles and prospects for single-pixel imaging. *Nat. Photonics* **2018**, *13*, 13–20. [CrossRef]
6. Mahalanobis, A.; Shilling, R.; Murphy, R.; Muise, R. Recent results of medium wave infrared compressive sensing. *Appl. Opt.* **2014**, *53*, 8060–8070. [CrossRef]
7. Radwell, N.; Mitchell, K.J.; Gibson, G.M.; Edgar, M.P.; Bowman, R.; Padgett, M.J. Single-pixel infrared and visible microscope. *Optica* **2014**, *1*, 285–289. [CrossRef]
8. Denk, O.; Musiienko, A.; Žídek, K. Differential single-pixel camera enabling low-cost microscopy in near-infrared spectral region. *Opt. Express* **2019**, *27*, 4562. [CrossRef]
9. Zhou, Q.; Ke, J.; Lam, E.Y. Near-infrared temporal compressive imaging for video. *Opt. Lett.* **2019**, *44*, 1702. [CrossRef]
10. Pastuszczak, A.; Stojek, R.; Wróbel, P.; Kotyński, R. Differential real-time single-pixel imaging with Fourier domain regularization—Applications to VIS-IR imaging and polarization imaging. *Opt. Express* **2021**, *29*, 2685–26700. [CrossRef]
11. Ye, J.T.; Yu, C.; Li, W.; Li, Z.P.; Lu, H.; Zhang, R.; Zhang, J.; Xu, F.; Pan, J.W. Ultraviolet photon-counting single-pixel imaging. *Appl. Phys. Lett.* **2023**, *123*, 024005. [CrossRef]
12. Stantchev, R.; Yu, X.; Blu, T.; Pickwell-MacPherson, E. Real-time terahertz imaging with a single-pixel detector. *Nat. Commun.* **2020**, *11*, 2535. [CrossRef] [PubMed]
13. Zanotto, L.; Piccoli, R.; Dong, J.; Morandotti, R.; Razzari, L. Single-pixel terahertz imaging: A review. *Opto-Electron Adv.* **2020**, *3*, 200012. [CrossRef]

14. Olbinado, M.P.; Paganin, D.M.; Cheng, Y.; Rack, A. X-ray phase-contrast ghost imaging using a single-pixel camera. *Optica* **2021**, *8*, 1538–1544. [CrossRef]
15. Li, J.; Cheng, K.; Qi, S.; Zhang, Z.; Zheng, G.; Zhong, J. Full-resolution, full-field-of-view, and high-quality fast Fourier single-pixel imaging. *Opt. Lett.* **2023**, *48*, 49–52. [CrossRef]
16. Czajkowski, K.; Pastuszczak, A.; Kotynski, R. Real-time single-pixel video imaging with Fourier domain regularization. *Opt. Express* **2018**, *26*, 20009–20022. [CrossRef]
17. Higham, C.; Murray-Smith, R.; Padgett, M.; Edgar, M. Deep learning for real-time single-pixel video. *Sci. Rep.* **2018**, *8*, 2369. [CrossRef]
18. Rizvi, S.; Cao, J.; Zhang, K.; Hao, Q. DeepGhost: Real-time computational ghost imaging via deep learning. *Sci. Rep.* **2020**, *10*, 11400. [CrossRef]
19. Wang, G.; Deng, H.; Ma, M.; Zhong, X.; Gong, X. A non-iterative foveated single-pixel imaging using fast transformation algorithm. *Appl. Phys. Lett.* **2023**, *123*, 081101. [CrossRef]
20. Zhu, X.; Li, Y.; Zhang, Z.; Zhong, J. Adaptive real-time single-pixel imaging. *Opt. Lett.* **2024**, *49*, 1065–1068. [CrossRef]
21. Zhang, Z.; Zheng, S.; Qiu, M.; Situ, G.; Brady, D.J.; Dai, Q.; Suo, J.; Yuan, X. A Decade Review of Video Compressive Sensing: A Roadmap to Practical Applications. *Engineering* **2024**, *In Press*. [CrossRef]
22. Ni, Y.; Zhou, D.; Yuan, S.; Bai, X.; Xu, Z.; Chen, J.; Li, C.; Zhou, X. Color computational ghost imaging based on a generative adversarial network. *Opt. Lett.* **2021**, *46*, 1840–1843. [CrossRef] [PubMed]
23. Wu, H.; Wang, R.; Zhao, G.; Xiao, H.; Wang, D.; Liang, J.; Tian, X.; Cheng, L.; Zhang, X. Sub-Nyquist computational ghost imaging with deep learning. *Opt. Express* **2020**, *28*, 3846–3853. [CrossRef] [PubMed]
24. Wang, F.; Wang, H.; Wang, H.; Li, G.; Situ, G. Learning from simulation: An end-to-end deep-learning approach for computational ghost imaging. *Opt. Express* **2019**, *27*, 25560–25572. [CrossRef] [PubMed]
25. Wang, F.; Wang, C.; Deng, C.; Han, S.; Situ, G. Single-pixel imaging using physics enhanced deep learning. *Photon. Res.* **2022**, *10*, 104–110. [CrossRef]
26. Tian, Y.; Fu, Y.; Zhang, J. Plug-and-play algorithms for single-pixel imaging. *Opt. Lasers Eng.* **2022**, *154*, 106970. [CrossRef]
27. Cheng, Z.; Lu, R.; Wang, Z.; Zhang, H.; Chen, B.; Meng, Z.; Yuan, X. BIRNAT: Bidirectional Recurrent Neural Networks with Adversarial Training for Video Snapshot Compressive Imaging. In Proceedings of the Computer Vision—ECCV 2020, Glasgow, UK, 23–28 August 2020; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Springer: Berlin/Heidelberg, Germany, 2020; pp. 258–275. [CrossRef]
28. Qin, C.; Schlemper, J.; Caballero, J.; Price, A.N.; Hajnal, J.V.; Rueckert, D. Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction. *IEEE Trans. Med. Imaging* **2019**, *38*, 280–290. [CrossRef]
29. Li, Z.; Huang, J.; Shi, D.; Chen, Y.; Yuan, K.; Hu, S.; Wang, Y. Single-pixel imaging with untrained convolutional autoencoder network. *Opt. Laser Technol.* **2023**, *167*, 109710. [CrossRef]
30. Liu, S.; Meng, X.; Yin, Y.; Wu, H.; Jiang, W. Computational ghost imaging based on an untrained neural network. *Opt. Lasers Eng.* **2021**, *147*, 106744. [CrossRef]
31. Wang, C.H.; Li, H.Z.; Bie, S.H.; Lv, R.B.; Chen, X.H. Single-Pixel Hyperspectral Imaging via an Untrained Convolutional Neural Network. *Photonics* **2023**, *10*, 224. [CrossRef]
32. Stojek, R.; Pastuszczak, A.; Wrobel, P.; Kotynski, R. Single pixel imaging at high pixel resolutions. *Opt. Express* **2022**, *30*, 22730–22745. [CrossRef] [PubMed]
33. Pastuszczak, A.; Stojek, R.; Wróbel, P.; Kotyński, R. Single pixel imaging at high resolution with sampling based on image maps. In Proceedings of the 2024 International Workshop on the Theory of Computational Sensing and its Applications to Radar, Multimodal Sensing and Imaging (CoSeRa), Santiago de Compostela, Spain, 18–20 September 2024; pp. 86–90. [CrossRef]
34. Zhao, W.; Gao, L.; Zhai, A.; Wang, D. Comparison of Common Algorithms for Single-Pixel Imaging via Compressed Sensing. *Sensors* **2023**, *23*, 4678. [CrossRef] [PubMed]
35. Bian, L.; Suo, J.; Dai, Q.; Chen, F. Experimental comparison of single-pixel imaging algorithms. *J. Opt. Soc. Am. A* **2018**, *35*, 78–87. [CrossRef] [PubMed]
36. Zhang, Z.; Ma, X.; Zhong, J. Single-pixel imaging by means of Fourier spectrum acquisition. *Nat. Commun.* **2015**, *6*, 6225. [CrossRef]
37. Czajkowski, K.M.; Pastuszczak, A.; Kotynski, R. Single-pixel imaging with sampling distributed over simplex vertices. *Opt. Lett.* **2019**, *44*, 1241–1244. [CrossRef]
38. Yu, W.K.; Liu, X.F.; Yao, X.R.; Wang, C.; Zhai, Y.; Zhai, G.J. Complementary compressive imaging for the telescopic system. *Sci. Rep.* **2014**, *4*, 5834. [CrossRef]
39. Gong, W. Disturbance-free single-pixel imaging camera via complementary detection. *Opt. Express* **2023**, *31*, 30505–30513. [CrossRef]
40. Yu, W.K.; Yao, X.R.; Liu, X.F.; Li, L.Z.; Zhai, G.J. Compressive moving target tracking with thermal light based on complementary sampling. *Appl. Opt.* **2015**, *54*, 4249–4254. [CrossRef]
41. Fellgett, P. Conclusions on multiplex methods. *J. Phys. Colloq.* **1967**, *28*, C2-165–C2-171. [CrossRef]
42. Scotté, C.; Galland, F.; Rigneault, H. Photon-noise: Is a single-pixel camera better than point scanning? A signal-to-noise ratio analysis for Hadamard and Cosine positive modulation. *J. Phys. Photonics* **2023**, *5*, 035003. [CrossRef]

43. Phillips, D.B.; Sun, M.J.; Taylor, J.M.; Edgar, M.P.; Barnett, S.M.; Gibson, G.M.; Padgett, M.J. Adaptive foveated single-pixel imaging with dynamic supersampling. *Sci. Adv.* **2017**, *3*, e1601782. [CrossRef] [PubMed]
44. Cui, H.; Cao, J.; Zhang, H.; Zhou, C.; Yao, H.; Hao, Q. Uniform-sampling foveated Fourier single-pixel imaging. *Opt. Laser Technol.* **2024**, *179*, 111249. [CrossRef]
45. Deng, L. The mnist database of handwritten digit images for machine learning research. *IEEE Signal Process. Mag.* **2012**, *29*, 141–142. [CrossRef]
46. Sobczak, K.; Kotynski, R.; Stojek, R.; Pastuszczak, A.; Cwojdzinska, M.; Wróbel, P. *Large Datasets for High Resolution Single Pixel Imaging*; Data Repository of the University of Warsaw: Warsaw, Poland, 2024. [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.

sensors

MDPI

Article

BSDR: A Data-Efficient Deep Learning-Based Hyperspectral Band Selection Algorithm Using Discrete Relaxation

Mohammad Rahman [1,2,*], Shyh Wei Teng [1], Manzur Murshed [3], Manoranjan Paul [4] and David Brennan [5]

[1] Institute of Innovation, Science and Sustainability, Federation University Australia, University Drive, Mt Helen, VIC 3350, Australia; shyh.wei.teng@federation.edu.au
[2] Cooperative Research Centre for High Performance Soils, Callaghan, NSW 2308, Australia
[3] School of Information Technology, Deakin University, 221 Burwood Hwy, Burwood, VIC 3125, Australia; m.murshed@deakin.edu.au
[4] School of Computing, Mathematics and Engineering, Charles Sturt University, Panorama Ave, Bathurst, NSW 2795, Australia; mpaul@csu.edu.au
[5] Wimmera Catchment Management Authority, 24 Darlot St, Horsham, VIC 3400, Australia; david.brennan@wcma.vic.gov.au
* Correspondence: aa.rahman@federation.edu.au

Abstract: Hyperspectral band selection algorithms are crucial for processing high-dimensional data, which enables dimensionality reduction, improves data analysis, and enhances computational efficiency. Among these, attention-based algorithms have gained prominence by ranking bands based on their discriminative capability. However, they require a large number of model parameters, which increases the need for extensive training data. To address this challenge, we propose Band Selection through Discrete Relaxation (BSDR), a novel deep learning-based algorithm. BSDR reduces the number of learnable parameters by focusing solely on the target bands, which are typically far fewer than the original bands, thus resulting in a data-efficient configuration that minimizes training data requirements and reduces training time. The algorithm employs discrete relaxation, transforming the discrete problem of band selection into a continuous optimization task, which enables gradient-based search across the spectral dimension. Through extensive evaluations on three benchmark datasets with varying spectral dimensions and characteristics, BSDR demonstrates superior performance for both regression and classification tasks, achieving up to 25% and 34.6% improvements in overall accuracy, compared to the latest attention-based and traditional algorithms, respectively, while reducing execution time by 96.8% and 97.18%. These findings highlight BSDR's effectiveness in addressing key challenges in hyperspectral band selection.

Keywords: band selection; discrete relaxation; gradient-based search; hyperspectral; data-efficient

1. Introduction

Hyperspectral data are high dimensional and comprise hundreds or even thousands of spectral bands. In various fields, including agriculture [1], environmental monitoring [2], mineralogy [3], health [4], and defense [5], obtaining a low-dimensional representation of high-dimensional hyperspectral data is important for purposes such as enhancing computational efficiency and identifying the most relevant features for specific tasks [6]. This dimensionality reduction is often accomplished using hyperspectral band selection algorithms, a type of feature selection technique [7] that removes highly correlated bands and selects a subset with greater discriminative power.

Hyperspectral band selection focuses on retaining the most informative spectral bands and preserving critical patterns and features in the data while enabling efficient processing and improving task performance across diverse applications such as land cover classification [8], soil property estimation [9], crop classification [10], target detection [11], and fault diagnosis in mechanical systems [12]. While traditional band selection methods,

such as the Monte Carlo Uninformative Variable Elimination (MCUVE) [13] and Successive Projections Algorithm (SPA) [14], rely on statistical metrics or heuristic approaches to select bands, modern approaches often employ advanced deep learning (DL) techniques to model complex dependencies and non-linear relationships within hyperspectral data [15]. Among these, attention-based algorithms [15–22] have gained prominence in recent years.

In DL, attention [23] is a neural network construct that assigns varying importance to different parts of the input, which allows the model to focus on relevant features for a given task. For band selection, this approach first appeared in BS-Nets [16,17], referred to as the Band Attention Module and Band Attention Network, respectively. In both cases, these modules generate weights for individual bands, based on which a number of top-ranked bands, corresponding to the given target size (t), are selected. This strategy has inspired several band selection algorithms [15,18–22], as well as other machine learning tasks [24–27].

A general architecture of attention-based band selection algorithms is depicted in Figure 1a. The Band Attention Module is a neural block—such as an FCNN, as seen in BS-Net-FC, which is one of the models under BS-Nets [16]—with an appropriate activation function at the last layer. Once the Band Attention Module generates a weight vector from the reflectance values of all bands (L, the total number of bands), a weighted reflectance vector is obtained through the element-wise multiplication of the reflectance vector and the weight vector. This weighted reflectance vector is then passed through another neural block, referred to here as the Inference Module, to assess the efficacy of the assigned weights through a downstream task. For instance, in BS-Nets, an autoencoder is used for unsupervised band selection, aiming to reconstruct the original hyperspectral data. In contrast, the model proposed by Zhang et al. [15], referred to hereafter as BS-Net-Classifier, employs a 1D convolutional neural network (1D-CNN) to classify maize seed varieties, selecting bands particularly relevant to this task. Some band selection algorithms inspired by BS-Nets incorporate advanced mechanisms beyond the general architecture depicted in Figure 1a to enhance band ranking. For instance, BSFormer [22] employs a transformer-based model as the Band Attention Module instead of an FCNN, while BS-Net-Classifier introduces a sparse constraint module to zero out the weights of unimportant bands below a specified threshold.

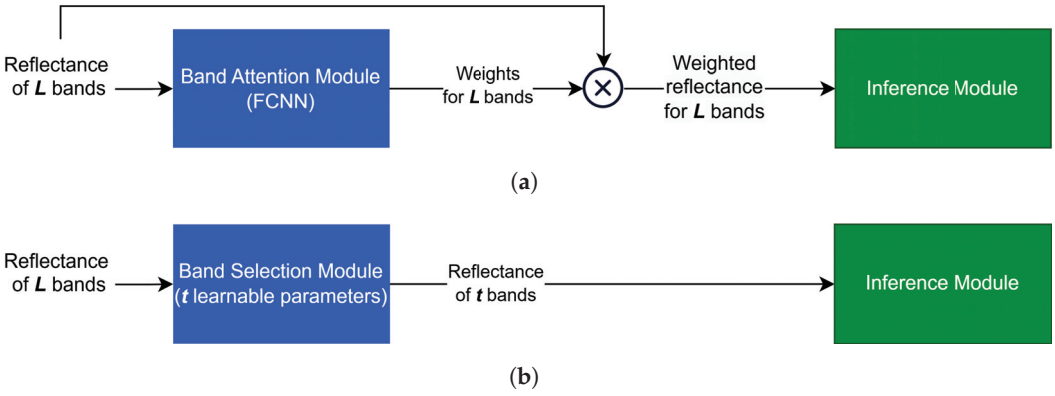


Figure 1. High-level architecture of the following: (**a**) Attention-based band selection algorithms; (**b**) The proposed algorithm. (**a**) Attention-based band selection algorithms: the number of model parameters depends on the original number of bands (L). (**b**) Proposed algorithm: the number of model parameters depends on the target size (t) and is independent of the original number of bands (L), as only the reflectance values of the t selected bands are processed during each training iteration.

As shown in Figure 1a, the input layer of the Band Attention Module must contain L nodes to attend to each individual band. Similarly, the output layer, which represents the weight vector of size L, must also have L nodes. Consequently, the number of learnable model parameters scales with L, regardless of the specific architecture of the Band Attention

Module, whether it is an FCNN, CNN, or transformer-based model. The Inference Module is similarly affected as its input is the weighted reflectance vector of size L. Since L is often large, this imposes significant computational demands and requires large training datasets, which are frequently scarce. Learning an excessive number of parameters from limited data can lead to the Hughes phenomenon, resulting in overfitting [28].

Through the effective use of model parameters for enhanced data efficiency as indicated by the ability to achieve strong generalization performance with the available training data, the number of learnable model parameters should be decoupled from the spectral dimension (L). To achieve this, we propose a novel band selection algorithm that focuses on selecting a subset of bands matching the target size (t) at each training iteration, ensuring that the number of required model parameters depends only on t rather than L. Figure 1b illustrates the high-level architecture of the proposed algorithm. As shown, the Band Selection Module includes only t learnable parameters, which represent the selected band indices. The module updates these parameters to capture the combination of important bands based on their relevance to the objective task, as evaluated by the Inference Module. Unlike attention-based algorithms, the proposed approach only feeds the reflectance values of the t selected bands into the Inference Module at each training iteration, thereby reducing the number of parameters in both the Band Selection Module and the Inference Module. This makes the algorithm inherently data-efficient, enabling effective training with a modest number of samples. The model is trained end-to-end, jointly optimizing the parameters of both modules through a unified optimization process.

While subset selection problems are generally NP-hard for optimal solutions [29], leveraging the quasi-continuous nature of hyperspectral data [30], where small changes in band index result in smooth changes in band values, allows us to apply gradient-based methods for a practical solution, which makes the subset selection task tractable. The primary challenge, however, is that band indices are discrete integers, which complicates the application of the gradient-based search methods that rely on a continuous search space. A common approach for addressing such challenges is discrete relaxation, an optimization technique that approximates discrete variables with continuous ones, enabling the use of continuous optimization methods [31]. By relaxing discrete constraints, the problem becomes differentiable, allowing for an efficient exploration of the solution space. Once optimized, the continuous values are mapped back to discrete ones to find a feasible solution to the original problem, making otherwise intractable problems solvable. In our proposed algorithm, we allow target indices to adopt non-integer values by applying discrete relaxation, which transforms the task into a continuous optimization problem suitable for gradient-based search. The rounding process has a minimal impact on the results since neighboring bands tend to be highly correlated [32]. This approach enables the solving of the problem in polynomial time with respect to the number of target indices, after which the indices are rounded to the nearest integer. Based on this, we name the proposed algorithm: Band Selection through Discrete Relaxation (BSDR).

Selecting a random subset and updating it based on its impact in predicting a target variable requires a subset selection mechanism. This mechanism must handle set inputs with properties such as permutation invariance and element uniqueness. We approximate the subset selection task using a learning mechanism that operates on vector inputs, where the vector represents a set. To ensure element uniqueness, the parameters are initialized by linearly spacing them across the spectral dimension.

We attempt to assess the efficacy of BSDR for two types of tasks: regression and classification. Since 3D spectral–spatial datasets for regression are limited, we focus on single-pixel samples. Our aim is to compare BSDR's performance against BS-Net-FC [16] and BS-Net-Classifier [15]. Additionally, we include two traditional band selection algorithms: MCUVE [13] and SPA [14].

To sum up, the main contributions of this paper are as follows:

1. To the best of our knowledge, BSDR is the first hyperspectral band selection algorithm that applies discrete relaxation to band indices, enabling gradient-based optimization

for band selection, thereby transforming the otherwise NP-hard subset selection problem into a tractable one for practical solutions.
2. Unlike modern the DL-based band selection algorithms that analyze the entire spectrum, BSDR only evaluates a limited number of bands at any given time. This approach requires significantly fewer model parameters than other DL-based models, making it not only inherently data-efficient but also enabling noticeable time efficiency without compromising predictive accuracy.
3. To the best of our knowledge, BSDR is the first band selection algorithm that enables the application of gradient-based search to a subset selection task by structuring a vector to emulate a set.

The remainder of the paper is structured as follows: Section 2 describes in detail our test datasets and the proposed algorithm. The results and relevant discussions are presented in Sections 3 and 4, respectively. Finally, Section 5 concludes the research and highlights future directions.

2. Materials and Methods

2.1. Test Datasets

To assess the efficacy of BSDR across datasets with diverse characteristics, including variations in sample size, and number of bands, we employed three different datasets in this study. For the classification task, we utilized the Global Hyperspectral Imaging Spectral-library of Agricultural crops for Conterminous United States (GHISACONUS) [33] and Indian Pines [34] datasets. For the regression task and to demonstrate the efficacy of BSDR on hyperspectral data with a very high spectral dimension, we utilized the Land Use/Cover Area frame Survey (LUCAS) [35] dataset. In the classification task, the goal is to predict land cover classifications for the Indian Pines dataset and crop identification for the GHISACONUS dataset, while in the regression task, our aim is to estimate the Soil Organic Carbon (SOC) content. A brief overview of these datasets is provided below.

GHISACONUS: GHISACONUS is formed through a collaboration between the United States Geological Survey and global partners, compiling hyperspectral signatures from major agricultural crops within the Conterminous United States. Data was acquired using the Earth Observing-1 (EO-1) Hyperion hyperspectral data and supported by the USDA Cropland Data Layer as reference data. It utilizes Earth Observing-1 (EO-1) Hyperion data, which spans 220 bands with a spectral range of 400 nm to 2500 nm; however, 198 bands ranging from 437 nm to 2395 nm are available in the dataset. This dataset includes 6988 samples of five different crops, collected from seven distinct agroecological zones.

Indian Pines: The Indian Pines dataset captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over Northwestern Indiana consists of 224 spectral bands. It primarily includes agricultural land and forest or natural vegetation. The ground truth for this dataset comprises 16 different classes, such as alfalfa, corn, grass-pasture, soybeans, wheat, woods, and other land cover types. The dataset is widely used for testing classification algorithms in hyperspectral imaging, given its challenging nature due to the high overlap in spectral signatures between some of the classes. The data cover wavelengths from 400 nm to 2500 nm, with 24 bands removed due to water absorption.

LUCAS: The LUCAS dataset is a large-scale, systematic survey conducted by Eurostat to monitor land use, land cover, and environmental conditions across the European Union. Launched in 2001, it gathers harmonized data to track changes in land cover over time, supporting policies related to agriculture, the environment, and climate. Through extensive field surveys, trained surveyors collect data on various land use types, soil characteristics, and vegetation from sample points across the EU. The dataset includes over 20,000 topsoil samples collected from 25 European countries, providing absorbance measurements in both the visible and near-infrared and short-wave infrared ranges, covering 400 nm to 2500 nm with a 0.5 nm spectral resolution, comprising 4200 bands.

To use these three datasets in our study, certain preprocessing steps are required. Among the 198 bands of the GHISACONUS dataset, we excluded 67 bands as information is missing for most of the samples for those bands. While the Indian Pines dataset originally provided 3D spectral–spatial data, Cai et al. [16] transformed it into single-pixel hyperspectral data to utilize it with BS-Net-FC. We have adopted the same transformation methodology in our study, excluding the background pixels. For LUCAS, we converted *absorbance* values into $reflectance$ using the equation discussed in [36]:

$$reflectance = \frac{1}{10^{absorbance}}. \tag{1}$$

Table 1 provides an overview of the three datasets. The data collection includes diverse geographic regions, encompassing a wide range of soil types with significant variability for both the Ghisaconus [33] and LUCAS [35] datasets. Additionally, the Indian Pines dataset exhibits a notable class imbalance [34], which poses unique challenges for classification tasks. Consistent performance by a band selection algorithm across these datasets would underscore its overall efficacy, demonstrating its capability to link spectral characteristics to their corresponding classes while effectively addressing complexities such as geographic diversity, soil variability, and class imbalance.

Table 1. Details of the datasets used in this study.

Dataset	Task	Target Variable	Samples	Bands
GHISACONUS	Classification	16 classes	6933	131
Indian Pines		5 crops	10,249	200
LUCAS	Regression	SOC content	21,782	4200

2.2. Proposed Algorithm

2.2.1. Mathematical Foundation

Hyperspectral data for N samples with L bands can be expressed as N spectral vectors v_i, where $v_i(j)$ represents the reflectance value of the j-th band of the i-th sample, for $i \in \{1, 2, \ldots, N\}$ and $j \in \{1, 2, \ldots, L\}$. To estimate a target variable y_i for $i \in \{1, 2, \ldots, N\}$, we determine an optimal combination of t bands, which requires deriving the subset that comprises the indices of the selected bands.

Selecting any t bands yields a subset s such that:

$$s = \{k_1, k_2, \ldots, k_t\} \tag{2}$$

where k_p represents a selected band index for all $p \in 1, 2, \ldots, t$. Considering $f_\theta(\cdot)$ approximates the target variable based on the reflectance values of the selected bands in s, where the function f is parameterized on θ, the set of learnable model parameters, the prediction for the i-th sample $\hat{y}_i$ is as follows:

$$\hat{y}_i = f_\theta(\{v_i(k_1), v_i(k_2), \ldots, v_i(k_t)\}). \tag{3}$$

Given that ϵ_i represents the error term in the prediction for the i-th sample,

$$y_i = \hat{y}_i + \epsilon_i \, . \tag{4}$$

The overall error ϵ can be minimized by minimizing the following expression:

$$\epsilon = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{5}$$

$$= \frac{1}{N} \sum_{i=1}^{N} (y_i - f_\theta(\{v_i(k_1), v_i(k_2), \ldots, v_i(k_t)\}))^2. \tag{6}$$

The discrete optimization problem of finding the optimal subset that minimizes ϵ can be formulated as the following objective function:

$$\underset{\theta,\{k_1,k_2,\ldots,k_t\}}{\arg\min} \frac{1}{N}\sum_{i=1}^{N}(y_i - f_\theta(\{v_i(k_1), v_i(k_2), \ldots, v_i(k_t)\}))^2. \tag{7}$$

In order to model the scenario with a DL-based architecture using a fully connected neural network, the first challenge is that typical fully connected neural networks generally learn from vector inputs. However, in this case, the input is a set. The state-of-the-art methods for such set learning tasks include Deep Sets [37] and Set Transformer [38]. However, since these methods process all elements uniformly and do not leverage sequential patterns, they are not suitable for our study, where we aim to optimize the number of model parameters and time complexity. Therefore, we introduce a vector $W_d = [k_1, k_2, \ldots, k_t]$, where the elements are initialized by linearly spacing them across the spectral dimension. W_d effectively emulates the set $\{k_1, k_2, \ldots, k_t\}$, particularly when there are no duplicate elements. As a result, we redefine the relation between $\hat{y_i}$ and f_θ as follows:

$$\hat{y_i} = f_\theta([v_i(k_1), v_i(k_2), \ldots, v_i(k_t)]). \tag{8}$$

The objective function is then redefined as follows:

$$\underset{\theta,W_d}{\arg\min} \frac{1}{N}\sum_{i=1}^{N}(y_i - f_\theta([v_i(k_1), v_i(k_2), \ldots, v_i(k_t)]))^2. \tag{9}$$

Equation (9) represents an NP-hard problem. In order to make the problem tractable for a practical solution by leveraging the sequential pattern in the hyperspectral data, we transform the discrete band indices to a continuous domain through discrete relaxation. In this approach, we allow the discrete band indices to take on continuous values. For this purpose, we define a function $d(.)$ that maps fractional indices into the original band indices. We restrict the domain of $d(\cdot)$ to $[0, 1]$ to ensure that the function bijectively maps to the original indices $[1, L]$. The domain of $d(\cdot)$ can be considered as normalized band indices between 0 and 1. $d(\cdot)$ can be considered the denormalization function that transforms the normalized band indices back to the scale and range of the original band indices. For a normalized band index q to derive the corresponding original integer band index I, $d(\cdot)$ can be defined as follows:

$$I = d(q) = \lfloor q \cdot (L-1) \rfloor + 1. \tag{10}$$

To approximate the reflectance value for the i-th sample at a fractional index q, we define a continuous function $B_i(.)$ that estimates this value through piecewise linear interpolation from the vector v_i. That is,

$$B_i(q) = v_i[d(q)] + (v_i[d(q)+1] - v_i[d(q)]) \cdot (q \cdot (L-1) - \lfloor q \cdot (L-1) \rfloor). \tag{11}$$

The continuous optimization formulation corresponding to the discrete optimization problem is expressed as follows:

$$\underset{\theta,W_c}{\arg\min} \frac{1}{N}\sum_{i=1}^{N}(y_i - f_\theta([B_i(r_1), B_i(r_2), \ldots, B_i(r_t)]))^2 \tag{12}$$

where $W_c = [r_1, r_2, \ldots, r_t]$ is the vector that emulates the selected subset of normalized band indices $\{r_1, r_2, \ldots, r_t\}$.

It is noteworthy that $B_i(\cdot)$ ensures that the objective function is smooth and differentiable with respect to the normalized band indices, which are the elements of W_c. This differentiability makes the problem amenable to DL-based models with gradient-based

optimization. In this case, gradient descent optimization updates not only θ but also W_c, simultaneously minimizing the objective function.

At the end of a fixed number of training iterations, the vector W^* containing the target integer indices is achieved by first denormalizing the normalized band indices in W_c and then rounding them to the nearest integer:

$$W^* = \{\text{round}(d(r_1)), \text{round}(d(r_2)), \ldots, \text{round}(d(r_t))\}. \quad (13)$$

However, it is possible that multiple elements of W_c may converge to a single value, leading to duplicate elements in W^*. Therefore, as a final step, by removing any duplicate values from W^*, we derive the resulting set S^*, which contains the unique target band indices in their original integer form: $I_1, I_2, \ldots, I_{t'}$, where $t' \leq t$.

The mathematical foundation discussed in this section is primarily based on regression problems. However, the same principles are applicable to classification problems, particularly by replacing the mean squared error function with the cross-entropy loss function as the objective function.

2.2.2. Time Complexity

The continuous functions $B_i(.)$ remain unchanged throughout the training process and, therefore, can be constructed only once. With a fixed number of training data, training iterations, and layers and nodes in the DL model, the training time depends only on the target size. Specifically, each unit of increase in target size results in an equivalent increase in the number of model parameters, thereby proportionally increasing training time.

With a fixed target size and number of training samples, the number of model parameters for regression tasks remains unchanged regardless of the dataset. However, for classification tasks, the number of weights between the last two layers increases proportionally with the number of output classes, slightly increasing the training time.

2.2.3. Architecture

Hyperspectral band selection algorithms can use either single pixels containing spectral information or 3D hyperspectral patches that include both spectral and spatial information as input. For clarity, we focus solely on the scenario where single pixels are used as input. The detailed architecture of BSDR is illustrated in Figure 2. First, in the Band Selection Module, we initialize a vector of learnable model parameters (c_1, c_2, …, c_t) corresponding to indices linearly spaced across the spectral dimension. These model parameters undergo a sigmoid transformation, restricting the transformed values (r_1, r_2, …, r_t) between 0 and 1, representing the normalized band indices. For the i-th sample, the Continuous Function Generator module produces a continuous function $B_i(\cdot)$, which approximates reflectance values for normalized band indices with piecewise linear interpolation as shown in the Equation (11).

During the training process, the approximated reflectance values corresponding to the normalized band indices, denoted as $B_i(r_1), B_i(r_2), \ldots, B_i(r_r)$, are passed as input into the Inference Module to predict the target variable. The structure of the Inference Module for classification and regression tasks has been illustrated in Figures 3a,b. Through a joint optimization process, all the learnable parameters in the entire architecture, including the weights and biases in the Inference Module and the learnable model parameters in the Band Selection Module corresponding to the target indices, are simultaneously updated.

The internal configuration of the Inference Module is designed such that the efficacy of the overall model depends more on the selected band indices than on the relational modeling capabilities of the module. Consequently, it has been deliberately simplified compared to the deeper layers used in previous studies [15,16]. Considering the available training data in the three datasets, we opted for a two-layer architecture with 128 and 64 nodes, respectively. To introduce non-linearity, we employed the Leaky ReLU activation function [39] after each hidden layer, as it is more favorable than the commonly used ReLU

due to its mitigation of the dying neuron issue [40], thereby improving learning dynamics. For datasets with fewer training samples, a simpler configuration may be adopted.

After training is complete, by denormalizing these normalized indices and then rounding them to the nearest integers, the vector W^* containing the target indices is achieved as shown in Equation (13). Finally, we derive the resulting set S^* by removing any duplicate values from W^*.

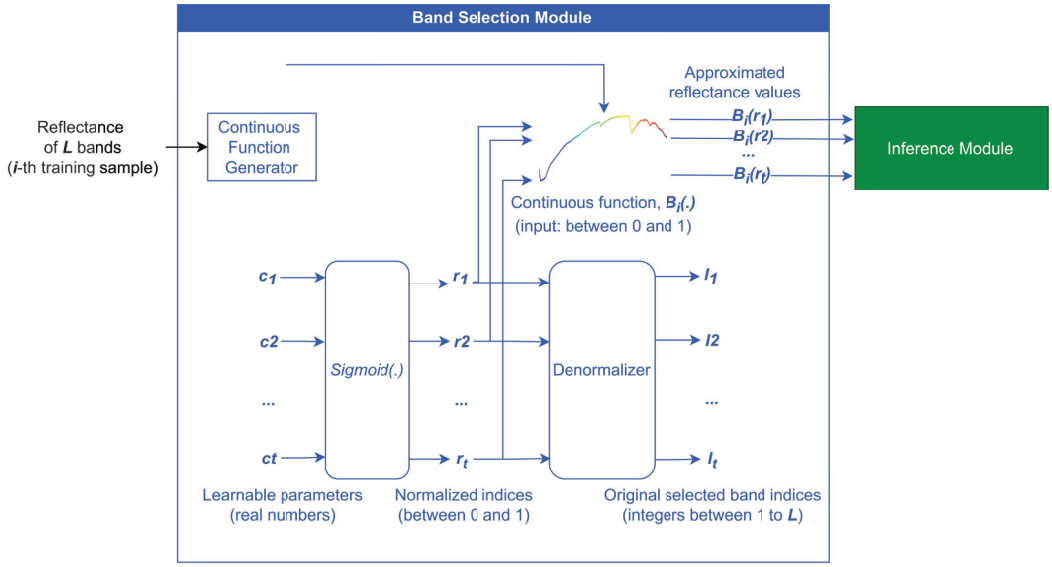
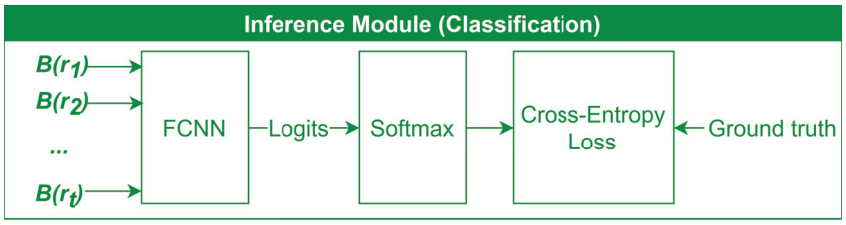
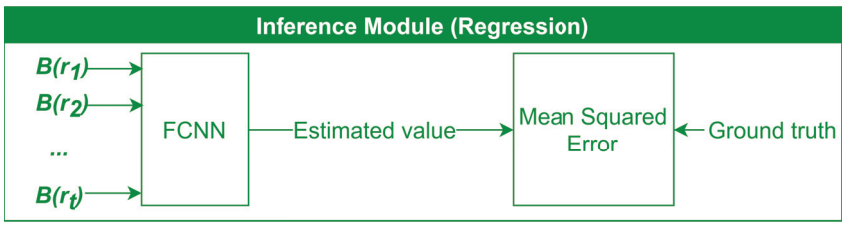


Figure 2. Detailed architecture of BSDR; the model parameters within the Band Selection Module ($c_1, c_2, \ldots, c_t$) are initialized to correspond to indices linearly spaced across the spectral dimension, which are updated during the training process based on their predictive capability, as measured by the Inference Module.



(**a**)



(**b**)

Figure 3. Inference Module for the following: (**a**) Classification; (**b**) Regression.

While mini-batch training is more common in DL, in this study, we utilized full-batch training due to the significantly low number of model parameters. Training with the entire dataset to compute the gradient in each epoch ensures more accurate and stable gradient calculations. This approach mitigates the noise and variance in gradient estimates that typically occur with smaller batches.

A low number of training epochs results in suboptimal band selection, whereas excessive epochs can lead to overfitting within the Inference Module. Consequently,

the selected bands become overly dependent on DL-based algorithms, diminishing their effectiveness when applied to alternative algorithms, such as SVM, within the context of this study. Following empirical observations, a balanced number of 500 training epochs has been determined.

Since the target bands are represented by learnable parameters, which are updated through the gradient descent algorithm, BSDR is highly sensitive to the learning rate. An aggressive learning rate can easily cause overshooting in band selection, while a low learning rate may fail to select an effective set of bands before training concludes. A learning rate of 0.001 has been found to be effective, offering a balanced approach.

2.3. Experimental Design and Objectives

For the three datasets, we aim to identify the target feature set for target sizes of 5, 10, 15, 20, 25, and 30 using all algorithms. Similar to previous studies [15,16], the efficacy of the bands selected by each algorithm is evaluated using a support vector machine (SVM) with a radial basis function kernel. As BS-Net-Classifier [15] is focused on classification, we limit its comparison to classification tasks only. For Cai et al.'s algorithm [16], we use BS-Net-FC since BS-Net-Conv is designed for 3D spectral–spatial data.

To ensure a fair and unbiased comparison of the algorithms, we apply 10-fold cross-validation. In each each cross-validation iteration, 90% of the data are randomly sampled for training the band selection algorithm, while the remaining 10% is used to evaluate the selected bands using an SVM classifier. From the training data, a randomly sampled 10% subset is allocated for model validation, with the rest used for training. Additionally, the evaluation split is randomly divided into two equal parts, one for training and the other for testing the SVM model on the selected bands. Figure 4 illustrates the process during each iteration. The prediction performance is also evaluated using all bands using SVM to assess the band selection efficacy of each algorithm relative to the performance without any data reduction. Due to the extensive time required by SPA for datasets with a large number of features, SPA was excluded from the 10-fold cross-validation for the LUCAS dataset, and only a single run was executed.

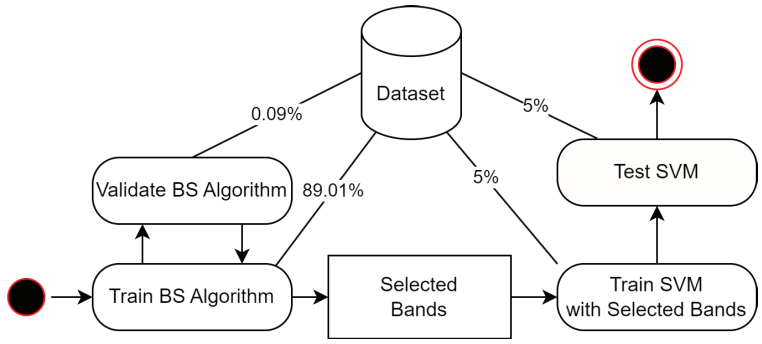


Figure 4. Activity diagram illustrating the experimental design, including data-split percentages for each activity: training and validating the band selection (BS) algorithm and training and testing SVM with the selected bands.

For hyperparameter selection in SVM, the regularization parameter C was set to 100 for regression and to 10^5 for classification, respectively, while the kernel coefficient γ was set to 1 for both, all determined through grid search.

2.3.1. Evaluation Metrics

To compare the performance of BSDR with other algorithms, we measure Overall Accuracy (OA) and Cohen's Kappa (κ) for classification tasks, as performed in previous studies [15,16]. While OA provides a straightforward measure of the proportion of correctly classified instances to the total, κ offers a more nuanced assessment by considering the agreement beyond chance between predicted and actual class labels. OA is intuitive and

easy to interpret but may overlook imbalances in class distribution, whereas κ adjusts for chance agreement, making it particularly valuable for assessing classifier performance in datasets with imbalanced classes or when considering the potential for chance agreement. For regression, we use R^2 and Root Mean Square Error (RMSE), which are common metrics for regression problems. R^2 gauges the proportion of variance explained by the regression model, indicating how well it fits the data, while RMSE measures the average size of prediction errors, directly reflecting model accuracy. For both classification and regression, we also measure the time required to train each band selection algorithm in seconds.

2.3.2. Objectives

The specific objectives of our experiments are to compare and analyze the performance of BSDR with the benchmark algorithms as per the following aspects:

1. The number of model parameters used by each DL-based algorithm.
2. Regarding classification tasks for the algorithms MCUVE, SPA, BS-Net-FC, BS-Net-Classifier, and BSDR:
    - 2.1 Time efficiency.
    - 2.2 Predictive accuracy in terms of OA and κ.
    - 2.3 Changes in target band indices throughout the training process.
3. Regarding the regression task for the algorithms MCUVE, SPA, BS-Net-FC, and BSDR:
    - 3.1 Time efficiency.
    - 3.2 Estimation accuracy in terms of R^2 and RMSE.
    - 3.3 Changes in target band indices throughout the training process.
4. Bands selected by BSDR for each of the 10 training folds across all three datasets. For presentation purposes, only the case with a target size of five has been chosen.
5. Bands selected by BSDR for each of the 10 training folds where two or more target band indices have converged to a single value for at least one of the training folds.

3. Results

3.1. Model Parameters

To accomplish Objective 1, which is listed in Section 2.3.2, the number of model parameters used by each DL-based algorithm across various datasets is presented in Table 2. The values presented for BSDR correspond to a target size of 5. For larger target sizes, the number of required parameters increases slightly. For the other algorithms, the number of parameters does not depend on the target size. As evident from the table, BSDR requires a significantly lower number of parameters compared to the other algorithms.

Table 2. Number of model parameters used by each DL-based band selection algorithm for the three datasets; BSDR requires substantially fewer model parameters than the other algorithms.

Algorithm	GHISACONUS	Indian Pines	LUCAS
BS-Net-FC [16]	118,674	154,416	2,226,416
BS-Net-Classifier [15]	135,724	207,740	N/A
Proposed BSDR	9354	10,069	9094

3.2. Classification

3.2.1. Time Complexity

Figure 5 corresponds to Objective 2.1 of Section 2.3.2, which shows the training time required for each algorithm for the GHISACONUS and Indian Pines datasets. Since the scale of the training time varies greatly across the algorithms, for the convenience of illustration, the results are shown on a logarithmic scale.

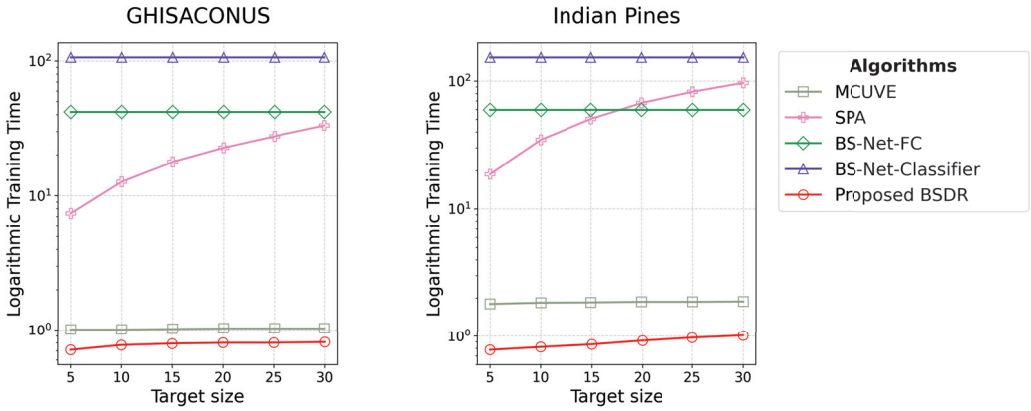


Figure 5. Training time (seconds) in logarithmic scale required by each band selection algorithm for the classification task for GHISACONUS and Indian Pines datasets, demonstrating the superior time efficiency of the proposed algorithm BSDR.

The comparative differences in training time for the algorithms are consistent across the two datasets. Training time is independent of the target size for the attention-based models, BS-Net-FC [16] and BS-Net-Classifier [15]. BSDR requires significantly less training time than SPA and the attention-based algorithms. MCUVE is close to BSDR in terms of training time, although it is slightly higher.

3.2.2. Predictive Accuracy

Objective 2.2 under Section 2.3.2 is addressed in Figure 6 and Tables 3 and 4, which represent the predictive accuracy of the classification tasks.

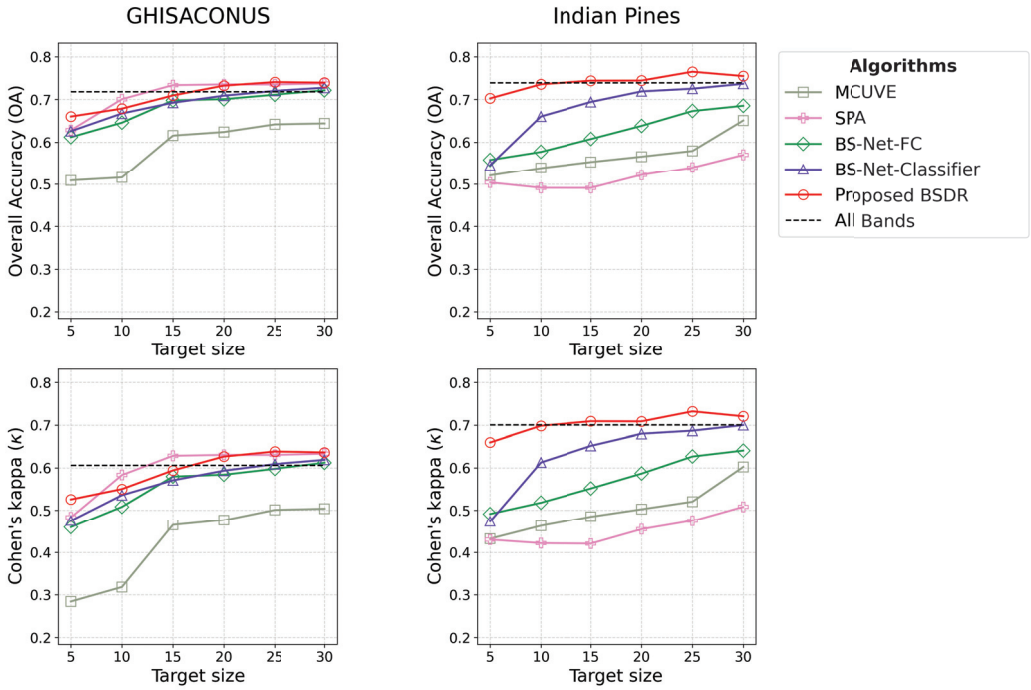


Figure 6. Predictive performance of the algorithms for classification tasks, shown in terms of Overall Accuracy (OA) and Cohen's kappa (κ) for the GHISACONUS and Indian Pines datasets, demonstrating the superior accuracy of the proposed BSDR algorithm.

Both OA and κ are consistent across the GHISACONUS and Indian Pines datasets as shown in Figure 6. Unlike the training time, however, the comparative predictive accuracy of the algorithms varies between the datasets. While SPA dominates in the GHISACONUS dataset, its performance is notably inferior in the other dataset. BSDR consistently outperforms all other algorithms in the Indian Pines dataset, with performance comparable to SPA in GHISACONUS. Notably, BSDR surpasses the performance with all bands while selecting 20 or more bands in GHISACONUS, and 15 or more bands in Indian Pines. The optimal performance is achieved with 25 bands for Indian Pines. It is also noteworthy that BS-Net-Classifier performs better than BS-Net-FC in both datasets.

Tables 3 and 4 summarize the OA results from the 10-fold cross-validation of the algorithms for different target sizes in the GHISACONUS and Indian Pines datasets, respectively. The standard deviation for BSDR across different executions is minimal, ranging from 0.02 to 0.04. BSDR demonstrates its most outstanding comparative performance for a target size of 5 on the Indian Pines dataset, achieving an OA of 0.7. This represents a 25% improvement over the closest competitor between the two attention-based algorithms, BS-Net-FC [16], which achieved an OA of 0.56, and a 34.62% improvement over the closest competitor between the two traditional algorithms, MCUVE, which achieved an OA of 0.52.

Table 3. Overall Accuracy (OA) based on 10-fold cross-validation for different target sizes on the GHISACONUS dataset, with the best results in bold and the second best in blue and italicized, showing the proposed BSDR algorithm achieves the best or comparable OA.

	Target Size					
Algorithm	**5**	**10**	**15**	**20**	**25**	**30**
MCUVE [13]	0.51 ± 0.05	0.52 ± 0.03	0.62 ± 0.04	0.62 ± 0.03	0.64 ± 0.04	0.64 ± 0.04
SPA [14]	*0.63 ± 0.03*	**0.70 ± 0.03**	**0.73 ± 0.03**	**0.74 ± 0.03**	**0.74 ± 0.03**	**0.74 ± 0.04**
BS-Net-FC [16]	0.61 ± 0.03	0.65 ± 0.03	0.70 ± 0.03	0.70 ± 0.02	0.71 ± 0.03	0.72 ± 0.03
BS-Net-Classifier [15]	*0.63 ± 0.03*	0.67 ± 0.04	0.69 ± 0.04	0.71 ± 0.03	*0.72 ± 0.04*	*0.73 ± 0.04*
Proposed BSDR	**0.66 ± 0.03**	*0.68 ± 0.03*	*0.71 ± 0.03*	*0.73 ± 0.03*	**0.74 ± 0.04**	**0.74 ± 0.03**

Table 4. Overall Accuracy (OA) based on 10-fold cross-validation for different target sizes on the Indian Pines dataset, with the best results in bold and the second best in blue and italicized, demonstrating that the proposed BSDR algorithm outperforms other algorithms by a noticeable margin.

	Target Size					
Algorithm	**5**	**10**	**15**	**20**	**25**	**30**
MCUVE [13]	0.52 ± 0.04	0.54 ± 0.05	0.55 ± 0.03	0.57 ± 0.03	0.58 ± 0.02	0.65 ± 0.04
SPA [14]	0.50 ± 0.03	0.49 ± 0.03	0.49 ± 0.02	0.52 ± 0.02	0.54 ± 0.02	0.57 ± 0.03
BS-Net-FC [16]	*0.56 ± 0.02*	0.58 ± 0.02	0.61 ± 0.03	0.64 ± 0.03	0.67 ± 0.02	0.69 ± 0.02
BS-Net-Classifier [15]	0.55 ± 0.06	*0.66 ± 0.05*	*0.69 ± 0.03*	*0.72 ± 0.03*	*0.73 ± 0.03*	*0.74 ± 0.03*
Proposed BSDR	**0.70 ± 0.04**	**0.74 ± 0.03**	**0.74 ± 0.02**	**0.74 ± 0.03**	**0.77 ± 0.02**	**0.75 ± 0.02**

3.2.3. Band Selection Across Training Iterations

Figure 7 addresses Objective 2.3 of Section 2.3.2, illustrating the selection of important bands by BSDR throughout the training epochs for the GHISACONUS and Indian Pines datasets (Figure 7a), along with corresponding updates in the OA and κ (Figure 7b). Although the band indices undergo minor updates throughout the epochs, these cumulative changes enhance OA and κ as training progresses. These metrics have been measured within BSDR using the validation dataset described in Section 2.3 to monitor the training progress.

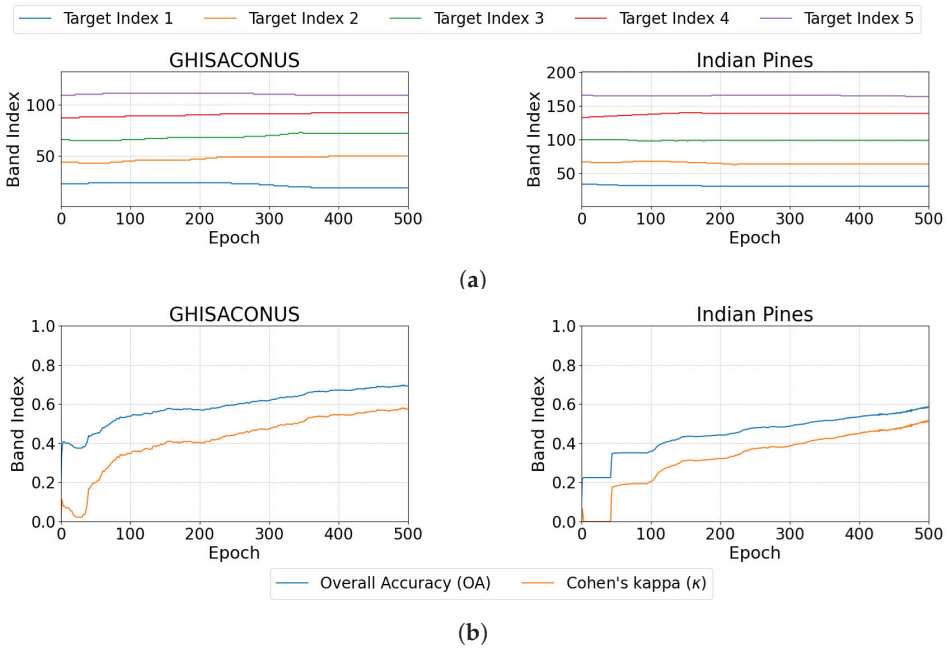


Figure 7. Changes in (**a**) band selection and corresponding (**b**) validation performance for two datasets across the training epochs; the validation performance improves as training progresses.

3.3. Regression

3.3.1. Time Complexity

Figure 8 and Table 5 correspond to Objective 3.1 under Section 2.3.2. Figure 8 shows the training time required by the algorithms for the regression task in the LUCAS dataset. Similar to the classification tasks, BSDR requires a significantly smaller amount of time than the other algorithms, while the time required to train SPA is excessively high. Table 5 summarizes the training time required by the algorithms for the regression task in the LUCAS dataset, based on 10-fold cross-validation. As mentioned in Section 2.3, SPA was executed only once due to its excessive training time; therefore, its stability for training time in different sets of training data has not been explored. In contrast, BSDR not only requires significantly less time than the other algorithms but also demonstrates consistent training times across different training sets, with a small standard deviation. Furthermore, as the target size increases, the training time for BSDR shows only a slight increase. BSDR exhibits its most efficient performance at a target size of 5, requiring only 2.54 s. This represents a 96.8% reduction in execution time compared to the attention-based competitor BS-Net-FC [16], which took 79.33 s, and a 97.18% reduction compared to the closest competitor between the two traditional algorithms, SPA, which took 90.02 s.

Table 5. Training time in seconds based on 10-fold cross-validation results for different target sizes on the LUCAS dataset, with the best results in bold and the second best in blue and italicized; the small standard deviation across different executions for the proposed BSDR algorithm demonstrates the reliability of its superior time efficiency.

Algorithm	Target Size					
	5	**10**	**15**	**20**	**25**	**30**
MCUVE [13]	90.02 ± 1.67	89.90 ± 1.81	93.02 ± 2.89	95.50 ± 22.95	93.81 ± 10.00	86.85 ± 1.76
SPA [14] [1]	1.54×10^4	2.10×10^4	4.22×10^4	6.91×10^4	8.23×10^4	8.88×10^4
BS-Net-FC [16]	*79.33 ± 1.54*	*79.33 ± 1.54*	*79.33 ± 1.54*	*79.33 ± 1.54*	*79.33 ± 1.54*	*79.33 ± 1.54*
Proposed BSDR	**2.54 ± 0.02**	**2.94 ± 0.01**	**3.19 ± 0.01**	**3.40 ± 0.03**	**3.71 ± 0.02**	**3.83 ± 0.02**

[1] Since SPA was executed only once due to its excessive training time, the standard deviation of its execution time could not be determined.

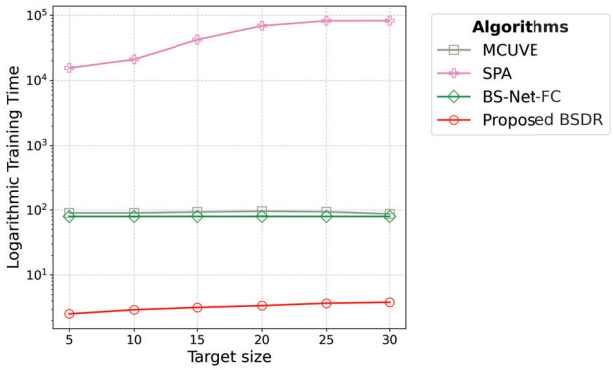


Figure 8. Training time (seconds) in logarithmic scale for each band selection algorithm for the regression task on the LUCAS dataset demonstrates that the proposed BSDR algorithm achieves superior time efficiency compared to the other algorithms by a significant margin.

3.3.2. Predictive Accuracy

Objective 3.2 under Section 2.3.2 is addressed by Figure 9 and Table 6. Figure 9 shows the estimation performance of the algorithms for the regression task in the LUCAS dataset. SPA outperforms the other algorithms in terms of accuracy. For BSDR, performance begins to deteriorate when selecting more than 15 bands and gradually converges with that of BS-Net-FC. Unlike in the GHISACONUS and Indian Pines datasets, the estimation performance with all bands is significantly lower in the LUCAS dataset. Based on the results measured from a single run of SPA, Table 6 shows that SPA outperforms the other algorithms for the regression task in the LUCAS dataset. BSDR maintains a low standard deviation ranging from 0.03 to 0.06 across the 10-fold validation, while remaining comparable to SPA.

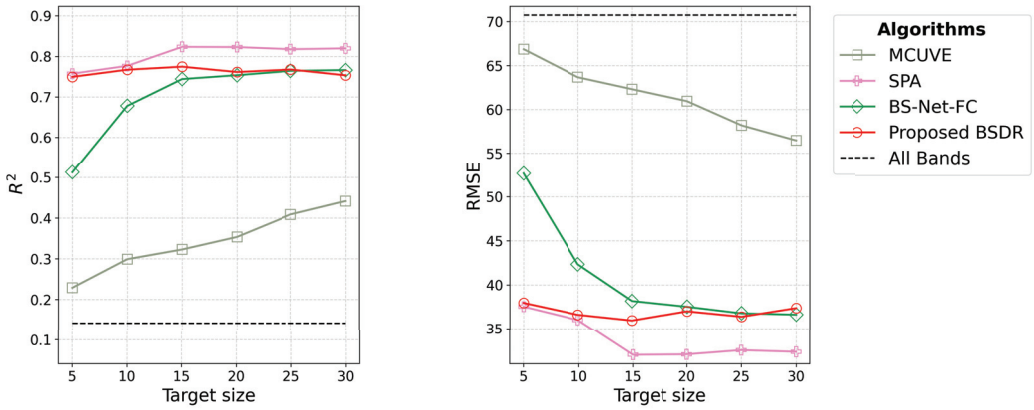


Figure 9. Band selection performance on the LUCAS dataset in terms of R^2 and RMSE for different algorithms; using all 4200 bands results in significantly inferior performance, as the evaluation algorithm (SVM) struggles to model the relationship between 4200 bands and the response variable (SOC) with only 1089 training samples.

Table 6. R^2 based on 10-fold cross-validation results for different target sizes on the LUCAS dataset; the best (bold) and the second best (blue and italicized) results are highlighted.

Algorithm	Target Size					
	5	**10**	**15**	**20**	**25**	**30**
MCUVE [13]	0.23 ± 0.17	0.30 ± 0.08	0.32 ± 0.11	0.35 ± 0.12	0.41 ± 0.09	0.44 ± 0.11
SPA [14] [1]	**0.76**	**0.78**	**0.82**	**0.82**	**0.82**	**0.82**
BS-Net-FC [16]	0.51 ± 0.11	0.68 ± 0.11	0.74 ± 0.05	0.75 ± 0.05	0.76 ± 0.04	*0.77 ± 0.03*
Proposed BSDR	*0.75 ± 0.03*	*0.77 ± 0.03*	*0.77 ± 0.04*	*0.76 ± 0.04*	*0.77 ± 0.05*	0.75 ± 0.06

[1] Since SPA was executed only once due to its excessive training time, the standard deviation of R^2 for this could not be determined.

3.3.3. Band Selection Across Training Iterations

In reference to Objective 3.3 in Section 2.3.2, Figure 10 illustrates the update of the selected bands by BSDR during training, along with the corresponding validation performance. Similar to the classification tasks, performance improves as the combination of bands is updated, despite the band indices undergoing minor updates throughout the training. The metrics R^2 and RMSE have been measured internally within BSDR with the validation data.

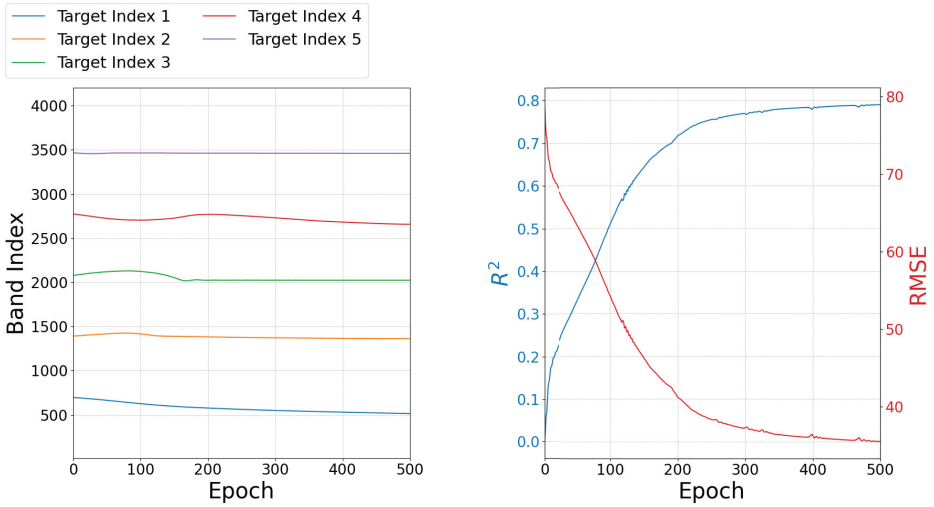


Figure 10. Changes in band selection and corresponding validation performance for regression across training epochs; the performance improves as the training progresses.

3.4. Selected Bands

Table 7 illustrates the bands selected in a sorted order by BSDR for each training fold in three datasets, specifically for a target size of five, as specified in Objective 4 listed in Section 2.3.2. The selected bands remain stable across the training folds with some minor differences due to the randomness introduced by the training folds. For example, in GHISACONUS, while the first band is always 19, and the most common indices for the other positions are 43, 63, 91, and 110, respectively, they slightly change in some training folds.

Table 7. Bands selected by BSDR across different training folds for three datasets at target size five; the selected bands remain stable with minor changes in some cases.

	Dataset		
Fold	**GHISACONUS**	**Indian Pines**	**LUCAS**
1	19, 50, 72, 92, 109	31, 64, 99, 139, 164	515, 1362, 2024, 2657, 3460
2	19, 43, 63, 91, 110	31, 64, 99, 141, 165	515, 1399, 2024, 2651, 3462
3	19, 44, 63, 91, 110	31, 64, 99, 141, 165	522, 1228, 2024, 2694, 3460
4	19, 44, 63, 91, 110	31, 64, 99, 139, 164	515, 1345, 2024, 2690, 3461
5	19, 44, 63, 91, 110	31, 64, 99, 140, 164	516, 1370, 2024, 2670, 3460
6	19, 43, 63, 91, 110	31, 64, 99, 141, 165	518, 1300, 2025, 2692, 3461
7	19, 43, 63, 91, 110	31, 64, 99, 140, 164	524, 1223, 2024, 2747, 3462
8	19, 43, 63, 91, 110	31, 64, 99, 141, 164	517, 1390, 2024, 2667, 3462
9	19, 43, 63, 91, 110	31, 64, 99, 141, 164	517, 1399, 2024, 2653, 3461
10	19, 44, 63, 91, 110	32, 64, 99, 140, 164	519, 1318, 2024, 2749, 3462

Table 8 addresses Objective 5 in Section 2.3.2, which lists the bands selected by BSDR for each of the 10 training folds where two or more target band indices have converged to a single value for at least one of the training folds. This figure corresponds to the case GHISACONUS dataset with a target size of 10. In this case, 10 bands have been selected in most of the cases as specified by the target size, and the selected bands remain stable

across the training sets with some minor differences, as shown in Table 7. However, in one instance, for the ninth fold, nine bands were selected instead of 10, meaning two of the target indices converged to a single value.

Table 8. Bands selected by BSDR across different training folds for the GHISACONUS dataset at a target size of 10; the ninth training fold demonstrates an example of achieving a lower number of bands than specified.

Fold	GHISACONUS
1	18, 24, 38, 48, 59, 78, 90, 92, 107, 119
2	17, 24, 38, 48, 59, 78, 91, 92, 109, 118
3	17, 24, 38, 48, 59, 78, 91, 92, 108, 118
4	17, 24, 38, 48, 59, 78, 91, 92, 109, 119
5	17, 24, 37, 48, 59, 78, 90, 92, 108, 119
6	17, 24, 38, 48, 59, 78, 90, 91, 108, 119
7	17, 24, 38, 48, 59, 78, 91, 92, 108, 119
8	17, 24, 38, 48, 59, 78, 90, 92, 107, 119
9	17, 24, 38, 48, 59, 78, 91, 107, 119
10	17, 24, 37, 48, 59, 78, 91, 92, 107, 119

4. Discussions

4.1. Model Parameters

For BSDR, the Band Selection Module's model parameters correspond to the target size. In the Inference Module, this count remains unchanged for regression tasks. For classification tasks, the weights between the last two linear layers slightly increase as the number of target classes increases. As a result, the number of model parameters increases slightly, from 9354 to 10,069, for GHISACONUS and Indian Pines, respectively, as shown in Table 2.

Both BS-Net-FC [16] and BS-Net-Classifier [15] utilize deep and wide layers, resulting in a significantly higher number of model parameters compared to BSDR. Furthermore, for both algorithms, the model parameter count increases linearly with the number of bands. The implication becomes evident when considering the significant number of model parameters, as large as 2,226,416, required by BS-Net-FC for the LUCAS dataset, consisting of 4200 bands. This is 14 and 18 times more than the number of parameters required by BS-Net-FC for Indian Pines and GHISACONUS, respectively, each comprising only a few hundred bands.

4.2. Classification

4.2.1. Time Complexity

The superior time efficiency depicted in Figure 5 stems from BSDR's use of full-batch training and its reliance on a minimal number of parameters compared to other deep learning-based algorithms. The modest increase in training time for BSDR with the target size is attributable to the corresponding slight increase in model parameters, as discussed in Section 3.1. BS-Net-Classifier requires 500 training epochs, compared to BS-Net-FC's 100 epochs. Consequently, it takes more time to train than BS-Net-FC. Training time for BS-Net-FC and BS-Net-Classifier remain constant across different target sizes, as these models attempt to derive weights for each band regardless of the target size. Since SPA selects new bands iteratively based on their information content, projecting them into orthogonal spaces through complex calculations, the training time escalates with the target size. This increase in training time is attributed to the growing number of iterations required as more bands are considered. Since MCUVE iteratively eliminates uninformative bands through a randomized approach and does not consider extensive combinations of features like SPA does, it requires less time to train than SPA.

4.2.2. Predictive Accuracy

In terms of predictive accuracy on the classification tasks, the closest competitor to BSDR is SPA, as shown in Figure 6. While BSDR consistently surpasses SPA on the Indian

Pines dataset, it trails SPA slightly on the GHISACONUS dataset for a few target sizes. This can be attributed to the consecutive missing bands across different regions. BSDR mistakenly treats two consecutive bands as adjacent, even when a large number of bands are missing in between, which leads to its degraded performance on the GHISACONUS dataset. However, the limited applicability of SPA is evident in the Indian Pines dataset, where it exhibits the most inferior performance. The disparity in SPA's performance between the two datasets arises because SPA selects informative bands in an unsupervised fashion without considering the target task. Given that the Indian Pines dataset contains more classes than GHISACONUS, 16 compared to five, SPA struggles to maintain consistent performance as the increased complexity from a greater number of classes hampers its ability to effectively discriminate among them. This rationale also explains why BS-Net-Classifier outperforms BS-Net-FC, especially on the more complex task on the Indian Pines dataset, which has a higher number of target classes. BS-Net-Classifier is tailored specifically to the classification task, fully utilizing training data to optimize performance. In contrast, BS-Net-FC learns band importance based on their capability to reconstruct the data in an unsupervised fashion without any information about the target task. Additionally, the superiority of BS-Net-Classifier is further enhanced by a sparse layer they utilized, which eliminates bands falling below a certain weight threshold, eliminating the uninformative bands concerning the target task.

The observed divergence between κ and OA for BS-Net-Classifier model, noted for the smaller target size of five in the Indian Pines dataset, suggests that the high OA may be more attributable to chance agreement rather than the true predictive efficacy of the algorithm for that instance. This divergence may stem from the large number of parameters in the model, which may induce overfitting. Additionally, a comparatively higher standard deviation in OA for target size five, reported as 0.06 in Table 4, further supports this observation.

For GHISACONUS, BS-Net-Classifier and BS-Net-FC require 25 and 30 bands, respectively, to achieve the OA attained using all bands. In contrast, BSDR surpasses the OA attained with all bands by utilizing only 20 bands. Additionally, the consistent κ for supports the true predictive accuracy of BSDR, indicating its effectiveness is not due to chance agreement.

In the Indian Pines dataset, BS-Net-Classifier and BSDR require 30 and 10 bands, respectively, to achieve the OA attained using all bands. Similarly, the consistent κ for BSDR at the target size of 10 supports the interpretation that its results reflect a true predictive capability rather than chance agreement.

In both datasets, BSDR achieves the best results with a target size of 25; however, beyond this point, its performance begins to deteriorate. The small standard deviation of the OA results for BSDR shown in Tables 3 and 4, based on the 10-fold cross-validation, supports the reliability of the presented results across various training data splits.

4.2.3. Band Selection Across Training Iterations

Throughout the training epochs, the selected band indices are updated, as illustrated in Figure 7. BSDR shows sensitivity to the initial values of the target indices, often converging toward a local optimum near the initial position. However, since many hyperspectral bands are linearly correlated, meaning that several bands can be reliably constructed as combinations of one or more bands, an effective combination can be achieved with small adjustments to each target band index, as demonstrated by BSDR.

As Figure 7 presents, the validation performance maintains a positive direction till the end of the training. It may seem prudent to continue training for more epochs. However, we observed that extended training contributes more significantly to the Inference Module. Beyond a certain point, gradient updates do not effectively propagate to the earlier layers, where the learnable band indices are located. Additionally, overtraining the Inference Module may result in diminished efficacy of the selected bands when using algorithms that do not employ DL.

4.3. Regression

4.3.1. Time Complexity

As mentioned in the Introduction, BS-Net-Classifier has been excluded from the regression task as it is exclusively designed for classification. Among the other algorithms, BSDR demonstrates the highest time efficiency, as shown in Figure 8 and Table 5, similar to the classification tasks. Even with a high spectral dimension of 4200 bands, the training time remains low because BSDR processes only a limited number of bands matching the target size, regardless of the spectral dimension size. The low standard deviation in execution time across multiple runs shown in Table 5 highlights the reliability of BSDR's superior time efficiency. The limitations of SPA become more pronounced with the LUCAS dataset. While the training time for SPA remained comparable to other algorithms for datasets containing a few hundred bands, as shown in Figure 5, it increased disproportionately for the LUCAS dataset. This raises questions about its applicability in high-dimensional datasets. For the same rationale presented in Section 4.2 regarding the random selection approach of MCUVE, the training time required for it remains within a comparable range with the other algorithms despite the high spectral dimension. Training time remains constant for BS-Net-FC as it evaluates weights for all the bands regardless of the target size.

4.3.2. Predictive Accuracy

As shown by the dashed lines in Figure 9, the performance on the LUCAS dataset using all bands is significantly lower than in the classification tasks with the GHISACONUS and Indian Pines datasets. This is primarily due to the much higher spectral dimension of the LUCAS dataset compared to the other two. Notably, an R^2 value of up to 0.91 has been reported in the literature [41] for SOC estimation from the LUCAS dataset using SVM. This result was achieved through rigorous preprocessing, outlier removal, and the exclusion of certain bands based on domain knowledge. Additionally, 14,277 training samples were used, whereas in our study, only 5% of the samples (1089) were used for training. Most notably, their approach involved downsampling the data by retaining one band every 10 nm, resulting in a reduced set of 195 bands.

Without such dimensionality reduction techniques, either through downsampling and band removal based on domain knowledge or automatic band selection using a suitable algorithm, many machine learning models may struggle to perform on high-dimensional datasets like LUCAS. Therefore, better estimation accuracy can be achieved even with five bands selected by the lowest-performing algorithm, MCUVE. These findings underscore the critical importance of dimensionality reduction, particularly for datasets with a high degree of collinearity or redundant features.

SPA performs consistently better than the other algorithms across different target sizes. However, this performance comes at the cost of a disproportionately long training time, ranging from 1.54×10^4 to 8.88×10^4 s based on the target size, as presented in Table 5. The performance of BSDR consistently improves across target sizes up to 15; beyond this point, it either deteriorates or remains unchanged at specific intervals. This deterioration could be attributed to the high degree of collinearity in the high-resolution visible and near-infrared spectra [42] provided by the LUCAS dataset. With as many as 4200 bands in the dataset, evaluating a large number of band combinations is crucial, as SPA demonstrates, to achieve effective band selection for target sizes exceeding 15.

The low standard deviation of the R^2 values for BSDR compared to BS-Net-FC up to target size 20, as presented in Table 6 based on 10-fold cross-validation, illustrates the consistent performance of the BSDR model across different training sets. Since SPA was executed only once due to its excessive training time, the standard deviation of its performance measures could not be determined.

4.3.3. Band Selection Across Training Iterations

Similar to classification tasks, significant changes in validation R^2 and RMSE are observed with minor updates in the learnable band indices as training progresses for the

regression task, as illustrated in Figure 10. The consistent increase in validation R^2 and decrease in validation RMSE suggest that the model is not only fitting effectively to the data but also generalizing effectively to unseen data. This trend implies a successful capture of the underlying patterns rather than memorization of the training set, as supported by the evaluated performance presented in Table 6.

4.4. Analysis of Band Selection Stability

The introduction of randomness by the training sets subtly influences the learning process of BSDR. These effects are manifested in the slight variations in band selection for certain indices, as demonstrated in Tables 7 and 8. The differences are comparatively bigger in LUCAS as the gradient-based search occurs through a higher spectral dimension, consisting of 4200 bands, in contrast to a few hundred bands in the other datasets.

As mentioned in Section 2.2, duplicate bands are removed after training concludes. Consequently, the number of selected bands might be lower than the intended target size in some instances. This is exemplified by the ninth training fold for the GHISACONUS dataset, where nine bands were selected, although the specified target size was 10, as shown in Table 8. The investigation through that particular instance of training reveals that after the 452nd epoch, both the seventh and eighth target indices assume the value of 91, which remains unchanged until the end of the training, as illustrated in Figure 11. The selected set was ultimately reduced to nine bands after removing the duplicate value of band 91.

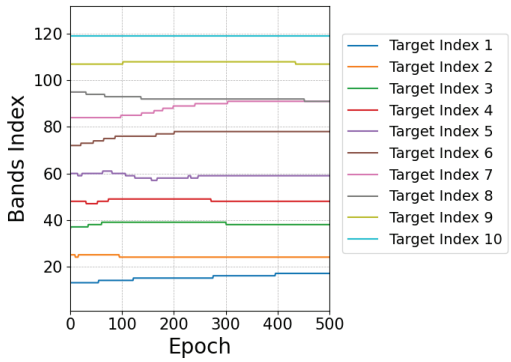


Figure 11. Example of multiple target indices converging to a single value, resulting in fewer selected bands than the specified target size. In this training case, toward the end of training, both the seventh and eighth target indices assume the value 91.

4.5. Overall Analysis

From Figures 6 and 9, it is evident that, except for some specific cases, BSDR consistently outperforms attention-based algorithms in predictive accuracy. This superior performance can be attributed not only to the data efficiency of BSDR, which stems from its reduced number of model parameters, but also to a fundamental limitation of ranking-based algorithms like BS-Net-FC [16] and BS-Net-Classifier [43]. While ranking-based methods provide a robust mechanism for ranking bands by importance, they fail to account for intraband correlations [44], often leading to a subset of bands with higher information redundancy. Moreover, some low-ranked bands that may exhibit strong correlations with the target variable when combined with other bands [45] are likely to be excluded.

In contrast, BSDR dynamically updates the selected bands based on their collective influence, thereby more effectively capturing intraband correlations. This consideration is particularly crucial for smaller target sizes, where redundancy reduces the proportion of informative bands, resulting in a more pronounced decline in predictive accuracy. The efficacy of BSDR in accounting for intraband correlations is demonstrated in Figures 6 and 9, as evidenced by the larger performance gap at lower target sizes. Across all three datasets, the performance gap between BSDR and the two attention-based algorithms is most pronounced at the lowest target size, 5.

5. Conclusions

This study demonstrates that the proposed model, BSDR, effectively selects hyperspectral bands, using a gradient-based search across the spectral dimension, as evidenced by a comprehensive evaluation of three benchmark datasets, which encompass diverse geographic regions, significant variability in soil types, and challenges such as class imbalance. By transforming discrete integer band indices into real numbers through discrete relaxation, BSDR reduces the NP-hard problem of band selection to polynomial time concerning the specified target size, providing a practical solution. Furthermore, this study shows that piecewise linear interpolation can effectively approximate the reflectance values for the fractional indices used in the gradient-based optimization.

Since BSDR processes only a limited number of bands matching the target size, regardless of the original spectral dimension, it requires significantly fewer model parameters. Consequently, it is not only highly data-efficient but also requires noticeably less training time compared to other recent DL-based band selection algorithms. This approach also allows BSDR to accommodate a high spectral dimension with as many as 4200 bands without requiring significantly more training time, which is unlike other comparably accurate algorithms such as SPA. Furthermore, BSDR demonstrated an effective approach to subset selection by learning through vector input and regulating the vector to closely emulate a set.

A limitation observed in BSDR is its high sensitivity to parameter initialization, which occasionally causes it to converge to a local optimum solution. In future studies, we intend to address this limitation by employing multiple initializations in parallel and deriving a combined decision based on them. Additionally, we will aim to develop a solution that addresses the issue of missing bands, thus ensuring that BSDR correctly distinguishes between adjacent and non-adjacent bands to improve performance on datasets like GHISACONUS.

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

Funding: This work has been supported by the Cooperative Research Centre for High Performance Soils (funding number 2.S.005a), whose activities are funded by the Australian Government's Cooperative Research Centre Program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the NASA EOSDIS Land Processes DAAC at https://doi.org/10.5067/Community/GHISA/GHISACONUS.001, reference number GHISACONUS.001, in the Purdue University Research Repository at https://doi.org/10.4231/R7RX991C, reference number R7RX991C, and in the European Soil Data Centre at https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data accessed on 27 November 2024, reference number LUCAS. These data were derived from the following resources available in the public domain: https://e4ftl01.cr.usgs.gov/COMMUNITY/GHISACONUS.001/2008.01.01/GHISACONUS_2008_001_speclib.csv accessed on 27 November 2024, https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html accessed on 27 November 2024, and https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data accessed on 27 November 2024.

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

References

1. Lu, B.; Dao, P.D.; Liu, J.; He, Y.; Shang, J. Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sens.* **2020**, *12*, 2659. [CrossRef]
2. Stuart, M.B.; McGonigle, A.J.; Willmott, J.R. Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems. *Sensors* **2019**, *19*, 3071. [CrossRef] [PubMed]

3. Tuşa, L.; Khodadadzadeh, M.; Contreras, C.; Rafiezadeh Shahi, K.; Fuchs, M.; Gloaguen, R.; Gutzmer, J. Drill-core mineral abundance estimation using hyperspectral and high-resolution mineralogical data. *Remote Sens.* **2020**, *12*, 1218. [CrossRef]
4. Wawerski, A.; Siemiątkowska, B.; Józwik, M.; Fajdek, B.; Partyka, M. Machine Learning Method and Hyperspectral Imaging for Precise Determination of Glucose and Silicon Levels. *Sensors* **2024**, *24*, 1306. [CrossRef]
5. Shimoni, M.; Haelterman, R.; Perneel, C. Hypersectral imaging for military and security applications: Combining myriad processing and sensing techniques. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 101–117. [CrossRef]
6. Sun, W.; Du, Q. Hyperspectral Band Selection: A Review. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 118–139. [CrossRef]
7. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [CrossRef]
8. Patro, R.N.; Subudhi, S.; Biswal, P.K.; Dell'acqua, F. A review of unsupervised band selection techniques: Land cover classification for hyperspectral earth observation data. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 72–111. [CrossRef]
9. Bai, Z.; Xie, M.; Hu, B.; Luo, D.; Wan, C.; Peng, J.; Shi, Z. Estimation of soil organic carbon using vis-nir spectral data and spectral feature bands selection in Southern Xinjiang, China. *Sensors* **2022**, *22*, 6124. [CrossRef]
10. Dave, K.; Vyas, T.; Trivedi, Y. Band selection technique for crop classification using hyperspectral data. *J. Indian Soc. Remote Sens.* **2022**, *50*, 1487–1498. [CrossRef]
11. Song, M.; Liu, S.; Xu, D.; Yu, H. Multiobjective optimization-based hyperspectral band selection for target detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5529022. [CrossRef]
12. Ni, Q.; Ji, J.; Feng, K.; Halkon, B. A novel correntropy-based band selection method for the fault diagnosis of bearings under fault-irrelevant impulsive and cyclostationary interferences. *Mech. Syst. Signal Process.* **2021**, *153*, 107498. [CrossRef]
13. Cai, W.; Li, Y.; Shao, X. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra. *Chemom. Intell. Lab. Syst.* **2008**, *90*, 188–194. [CrossRef]
14. Araújo, M.C.U.; Saldanha, T.C.B.; Galvao, R.K.H.; Yoneyama, T.; Chame, H.C.; Visani, V. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemom. Intell. Lab. Syst.* **2001**, *57*, 65–73. [CrossRef]
15. Zhang, L.; Wei, Y.; Liu, J.; Wu, J.; An, D. A hyperspectral band selection method based on sparse band attention network for maize seed variety identification. *Expert Syst. Appl.* **2024**, *238*, 122273. [CrossRef]
16. 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]
17. Wang, J.; Zhou, J.; Huang, W. Attend in bands: Hyperspectral band weighting and selection for image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 4712–4727. [CrossRef]
18. 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]
19. Li, T.; Cai, Y.; Cai, Z.; Liu, X.; Hu, Q. Nonlocal band attention network for hyperspectral image band selection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3462–3474. [CrossRef]
20. Feng, J.; Ye, Z.; Liu, S.; Zhang, X.; Chen, J.; Shang, R.; Jiao, L. Dual-graph convolutional network based on band attention and sparse constraint for hyperspectral band selection. *Knowl.-Based Syst.* **2021**, *231*, 107428. [CrossRef]
21. Zheng, Z.; Liu, Y.; He, M.; Chen, D.; Sun, L.; Zhu, F. Effective band selection of hyperspectral image by an attention mechanism-based convolutional network. *RSC Adv.* **2022**, *12*, 8750–8759. [CrossRef]
22. 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]
23. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [CrossRef]
24. Li, Z.; Xiong, F.; Zhou, J.; Wang, J.; Lu, J.; Qian, Y. BAE-Net: A band attention aware ensemble network for hyperspectral object tracking. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Virtual, 25–28 October 2020; pp. 2106–2110.
25. Zhao, L.; Yi, J.; Li, X.; Hu, W.; Wu, J.; Zhang, G. Compact band weighting module based on attention-driven for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 9540–9552. [CrossRef]
26. Li, Z.; Ye, X.; Xiong, F.; Lu, J.; Zhou, J.; Qian, Y. Spectral-spatial-temporal attention network for hyperspectral tracking. In Proceedings of the 2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 24–26 March 2021; pp. 1–5.
27. Li, W.; Hou, Z.; Zhou, J.; Tao, R. SiamBAG: Band attention grouping-based Siamese object tracking network for hyperspectral videos. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5514712. [CrossRef]
28. Li, W.; Chen, H.; Liu, Q.; Liu, H.; Wang, Y.; Gui, G. Attention mechanism and depthwise separable convolution aided 3DCNN for hyperspectral remote sensing image classification. *Remote Sens.* **2022**, *14*, 2215. [CrossRef]
29. Bins, J.; Draper, B.A. Feature selection from huge feature sets. In Proceedings of the Eighth IEEE International Conference on Computer Vision, ICCV 2001, Vancouver, BC, Canada, 7–14 July 2001; Volume 2, pp. 159–165. [CrossRef]
30. Alparone, L.; Arienzo, A.; Garzelli, A. Spatial Resolution Enhancement of Vegetation Indexes via Fusion of Hyperspectral and Multispectral Satellite Data. *Remote Sens.* **2024**, *16*, 875. [CrossRef]
31. Hancock, E.R.; Kittler, J. Discrete relaxation. *Pattern Recognit.* **1990**, *23*, 711–733. [CrossRef]
32. Zhang, W.; Li, X.; Zhao, L. Band priority index: A feature selection framework for hyperspectral imagery. *Remote Sens.* **2018**, *10*, 1095. [CrossRef]

33. Thenkabail, P.; Aneece, I. Global Hyperspectral Imaging Spectral-Library of Agricultural Crops for Conterminous United States V001 [Data Set]. *NASA EOSDIS Land Process. DAAC* **2019**. [CrossRef]
34. 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. [CrossRef]
35. Jones, A.; Fernandez-Ugalde, O.; Scarpa, S. LUCAS 2015 topsoil survey. *Present. Dataset Results EUR* **2020**, *30332*, 616084.
36. Sirita, J.; Phanichphant, S.; Meunier, F.C. Quantitative analysis of adsorbate concentrations by diffuse reflectance FT-IR. *Anal. Chem.* **2007**, *79*, 3912–3918. [CrossRef]
37. Zaheer, M.; Kottur, S.; Ravanbakhsh, S.; Poczos, B.; Salakhutdinov, R.R.; Smola, A.J. Deep sets. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [CrossRef]
38. Lee, J.; Lee, Y.; Kim, J.; Kosiorek, A.R.; Choi, S.; Teh, Y.W. Set transformer: A framework for attention-based permutation-invariant neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; Volume 97, pp. 3744–3753.
39. Xu, B.; Wang, N.; Chen, T.; Li, M. Empirical evaluation of rectified activations in convolutional network. *arXiv* **2015**, arXiv:1505.00853.
40. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the 30 th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; Volume 30, p. 3.
41. Zhong, L.; Guo, X.; Xu, Z.; Ding, M. Soil properties: Their prediction and feature extraction from the LUCAS spectral library using deep convolutional neural networks. *Geoderma* **2021**, *402*, 115366. [CrossRef]
42. Shen, Z.; Viscarra Rossel, R. Automated spectroscopic modelling with optimised convolutional neural networks. *Sci. Rep.* **2021**, *11*, 208. [CrossRef]
43. Zhang, H.; Gao, H.; Sun, H.; Sun, X.; Zhang, B. A Spatial-Spectrum Fully-Attention Network for Band Selection of Hyperspectral Images. *IEEE Geosci. Remote. Sens. Lett.* **2024**, *21*, 5507205. [CrossRef]
44. Sun, W.; Tian, L.; Xu, Y.; Zhang, D.; Du, Q. Fast and robust self-representation method for hyperspectral band selection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 5087–5098. [CrossRef]
45. Sosa-Cabrera, G.; Gómez-Guerrero, S.; García-Torres, M.; Schaerer, C.E. Feature selection: A perspective on inter-attribute cooperation. *Int. J. Data Sci. Anal.* **2024**, *17*, 139–151. [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.

sensors

MDPI

Article

Fluorescence and Hyperspectral Sensors for Nondestructive Analysis and Prediction of Biophysical Compounds in the Green and Purple Leaves of *Tradescantia* Plants

Renan Falcioni [1,*], Roney Berti de Oliveira [1], Marcelo Luiz Chicati [1], Werner Camargos Antunes [1], José Alexandre M. Demattê [2] and Marcos Rafael Nanni [1]

[1] Department of Agronomy, State University of Maringá, Av. Colombo, 5790, Maringá 87020-900, Paraná, Brazil; rboliveira@uem.br (R.B.d.O.); mlchicati@uem.br (M.L.C.); wcantunes@uem.br (W.C.A.); mrnanni@uem.br (M.R.N.)

[2] Department of Soil Science, Luiz de Queiroz College of Agriculture, University of São Paulo, Av. Pádua Dias, 11, Piracicaba 13418-260, São Paulo, Brazil; jamdemat@usp.br

* Correspondence: renanfalcioni@gmail.com or rfalcioni2@uem.br; Tel.: +55-4430118940

Abstract: The application of non-imaging hyperspectral sensors has significantly enhanced the study of leaf optical properties across different plant species. In this study, chlorophyll fluorescence (ChlF) and hyperspectral non-imaging sensors using ultraviolet-visible-near-infrared shortwave infrared (UV-VIS-NIR-SWIR) bands were used to evaluate leaf biophysical parameters. For analyses, principal component analysis (PCA) and partial least squares regression (PLSR) were used to predict eight structural and ultrastructural (biophysical) traits in green and purple *Tradescantia* leaves. The main results demonstrate that specific hyperspectral vegetation indices (HVIs) markedly improve the precision of partial least squares regression (PLSR) models, enabling reliable and nondestructive evaluations of plant biophysical attributes. PCA revealed unique spectral signatures, with the first principal component accounting for more than 90% of the variation in sensor data. High predictive accuracy was achieved for variables such as the thickness of the adaxial and abaxial hypodermis layers (R^2 = 0.94) and total leaf thickness, although challenges remain in predicting parameters such as the thickness of the parenchyma and granum layers within the thylakoid membrane. The effectiveness of integrating ChlF and hyperspectral technologies, along with spectroradiometers and fluorescence sensors, in advancing plant physiological research and improving optical spectroscopy for environmental monitoring and assessment. These methods offer a good strategy for promoting sustainability in future agricultural practices across a broad range of plant species, supporting cell biology and material analyses.

Keywords: chemometrics; chlorophyll a fluorescence; hyperspectral vegetation indices; partial least squares regression; principal component analysis; optical spectroscopy; UV-VIS-NIR-SWIR

1. Introduction

The use of non-imaging hyperspectral or proximal sensors along with chlorophyll fluorescence techniques has made substantial progress in evaluating the biophysical characteristics of leaves and plants. These state-of-the-art methods offer insights into the physiological conditions of plants, facilitating an extensive understanding of their responses to diverse environmental factors [1,2]. In addition to traditional assessment methods, these sensor technologies enable new ways to monitor plant health and detect changes in the biophysical attributes of leaves, aligning structure and ultrastructure components while optimizing sustainable agricultural practices and plant management [3]. The integration of biophysical analyses is important for improving predictive models, enhancing our understanding of plant-environment interactions, and guiding precision agriculture, ultimately contributing to more resilient and sustainable ecosystems.

Chlorophyll fluorescence (ChlF) curves are invaluable resources for gauging plant health by accurately mapping pigment levels and other biochemical properties, such as protein and carbohydrate content [4,5]. These curves measure the fluorescence emitted by chlorophyll a and provide critical information on photosynthetic efficiency and plant stress responses [6,7]. In this sense, the ChlF curve data reveal variations in biophysical properties at the cellular level, which is important for understanding and modeling plant behavior. The ability of this technique to detect early stress indicators before visible symptoms occur makes it particularly effective for proactive agricultural and environmental management and contributes valuable insights to the fields of plant physiology and ecology [8].

Ultraviolet-visible-near-infrared shortwave infrared (UV-VIS-NIR-SWIR) hyperspectral and chlorophyll fluorescence imaging and non-imaging sensors generate comprehensive spectral profiles, capturing detailed plant information across a broad spectrum of wavelengths [9–11]. In particular, non-imaging sensors allow swift data collection without a spatial biophysical context, which is highly efficient for large-scale or remote-sensing evaluations [12–14]. The precision and accuracy of these sensors for detecting minor physiological changes significantly increase their ability to monitor biophysical responses, as reviewed in detail [6,7].

The cellular and foliar architecture of plants is correlated, with various chloroplastic cells and leaf optical properties, such as thylakoid granum, adaxial hypodermis, chloroplast number, total leaf thickness, and parenchyma thickness [15,16]. Structural components such as the epidermis, hypodermis, and leaf parenchyma are crucial for maintaining the biomechanical integrity of plant tissues. Modifications of these cells and tissue arrangements can differ greatly and are influenced by factors such as leaf area expansion and cellular structural changes due to biophysical compounds [17–20]. These variations offer significant insights into the adaptive mechanisms of plants under environmental conditions [21–23]. Recent research has shown that fluorescence techniques and hyperspectral sensors use different methods to analyze and predict biochemical and biophysical compounds to monitor the biophysical properties of the leaves, steams, and roots of lettuces, tomatoes, and ornamental plants [11,24,25]. These tools assess changes in cellular size and volume correlated with water content and respond to pigment content, indirectly reflecting variations in concentrated pigments such as chlorophylls and carotenoids correlated with chloroplast numbers or thylakoid stacking in cells [26–28].

In leaves with biochemical and biophysical variations, optical properties such as reflectance, transmittance, and absorbance are primarily shaped by structural features, including thickness, cell organization, and mesophyll composition [11,29–31]. These properties are further influenced by biochemical elements, particularly pigment concentration and type, which affect light absorption and distribution. In *Tradescantia* plants, the use of these three spectral modes along with fluorescence data provides a precise evaluation of the biophysical mechanisms governing light interactions. This approach enhances the prediction of structural traits, such as leaf thickness, while efficiently estimating key biochemical components such as pigments. Additionally, the use of these sensors, along with others, has proven effective in analyzing foliar structural and ultrastructural components with interaction light with matter, or in this example for tissue plants [10,32–35].

Hyperspectral Vegetation Indices (HVIs) utilize reflectance values across different wavelengths to assess the plant growth status, health, and photosynthetic efficiency. For example, recent advancements have highlighted the value of HVIs in analyzing extensive datasets from both chlorophyll fluorescence and hyperspectral measurements [36–39]. For example, HVIs simplify complex spectral data into manageable indices, facilitating large-scale monitoring of biophysical and physiological conditions. These indices can identify regions and bands in which key biophysical components show significant changes. HVIs can detect variations in chlorophyll content, water stress, and nutrient levels [7, 40,41]. When combined with chlorophyll fluorescence data, HVIs enhance the ability to differentiate areas with varying photosynthetic rates or stress responses, enabling precise mapping of biophysical markers across landscapes [9,42,43].

Principal Component Analysis (PCA) is a powerful statistical method used to handle the complex, high-dimensional nature of hyperspectral and fluorescence data [11,43]. This technique effectively reduces data dimensionality while retaining most of the variance and uncovering the underlying patterns within the samples. PCA is particularly useful for determining the optimal number of principal components needed to capture essential data features and balancing information retention with noise reduction in datasets. The β-loadings from PCA clearly highlight spectral peaks, revealing the "fingerprints" of physical properties in the dataset, which are correlated with many biophysical parameters [11,24,44,45]. By combining PCA with regression coefficients from various spectral curves, such as OJIP and hyperspectral reflectance, transmittance, and absorbance curves, a more comprehensive analysis integrating biochemical and biophysical insights is possible using prediction models [17,28,46].

The second main method of development is Partial Least Squares Regression (PLSR), a robust method for predictive modeling with hyperspectral data that can manage highly collinear and multivariate datasets. PLSR projects predictor variables (spectral bands) onto a new subspace defined by latent variables that account for the predictor variance and their correlation with the response variable. This method is particularly beneficial when the number of predictors exceeds the number of observations [11,36,47]. PLSR effectively addresses multicollinearity among spectral bands, reducing predictors to a smaller set of uncorrelated components, while capturing key response variable patterns. Combining PLSR with cross-validation techniques enhances model stability and predictive performance, making it the preferred choice for handling complex, large hyperspectral datasets [11,36,47].

Machine Learning (ML), Deep Learning (DL), and Artificial Intelligence (AI) models have made significant contributions to biophysical analysis, particularly in predicting and selecting the most relevant wavelengths for hyperspectral data [2,36,48]. For example, random forests and Convolutional Neural Networks (CNNs) have been successfully applied to detect subtle variations in plant health and stress indicators, whereas Support Vector Machines (SVMs) [49] have been effective in classifying spectral data [50]. These advanced techniques allow automated feature selection and can identify critical spectral regions without human intervention. Despite these advancements, PLSR models have proven to be efficient, offering high predictive accuracy and robust performance in scenarios where interpretability and computational simplicity are key, making them a reliable choice alongside ML and DL models [37,43,51–53].

This study focused on improving the prediction of biophysical parameters using chlorophyll fluorescence kinetics and hyperspectral sensors in two *Tradescantia* species. The primary objective was to identify spectral bands in hyperspectral data that correlate strongly with chlorophyll a fluorescence, aiding in the estimation of eight biophysical parameters in green and purple *Tradescantia* leaves. By integrating advanced hyperspectral sensors for reflectance, transmittance, and absorbance data with fluorescence kinetics and utilizing multivariate statistical methods, such as PCA and PLSR, we hypothesized that the prediction accuracy for these parameters could be significantly enhanced, even with variations in leaf coloration (Figure 1).

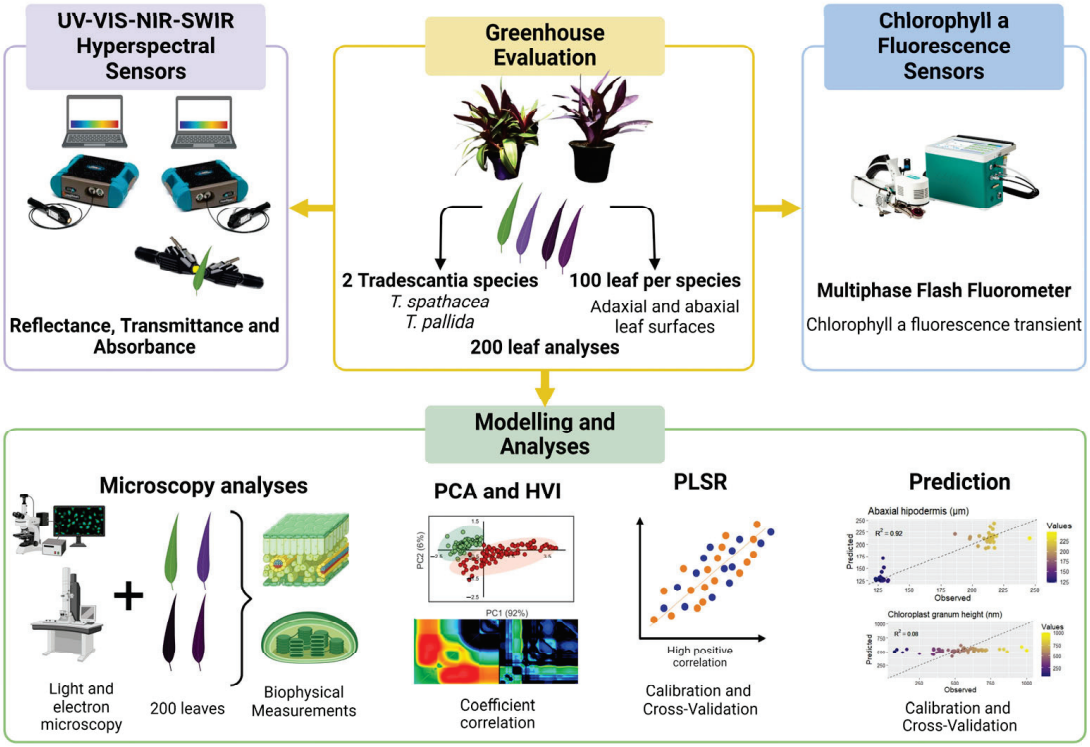


Figure 1. Flowchart for predicting biophysical compounds.

2. Material and Methods

2.1. Plant Materials and Experimental Analysis

The species of *Tradescantia spathacea* and *Tradescantia pallida*, which were chosen for their unique biophysical characteristics, were grown in 2-L pots at the Botanical Garden affiliated with the State University of Maringá, Brazil. The conditions within the greenhouse included exposure to natural light, with temperatures ranging from 22 to 26 °C and a prolonged light period of 16 h. The irrigation schedule was set to twice a day, once at 8 a.m. and then again at 6 p.m., to ensure regular soil hydration. For a thorough evaluation, leaf samples were harvested from different sections of the plant. This research aimed to conduct an in-depth hyperspectral reflectance analysis and detailed biophysical examination of 200 leaf samples, the results of which are presented in Figure 1.

2.2. Chlorophyll a Fluorescence Measurement

Chlorophyll a fluorescence was measured using an infrared gas exchange analyzer (IRGA) paired with a Multiphase Flash™ Fluorometer (LI-6800-01; LI-COR Inc., Lincoln, NE, USA). The sensor detects chlorophyll fluorescence at wavelengths greater than 720 nm. Leaves were dark-acclimated for 12 h before being subjected to a saturating light pulse of 15,000 µmol m^{-2} s^{-1} for 1 s to induce chlorophyll a fluorescence, ensuring the closure of all reaction centers [11,29].

2.3. Hyperspectral Optical Properties of Leaves

The reflectance and transmittance properties of the leaves were simultaneously measured using a FieldSpec® 3 spectroradiometer with an ASD Contact PlantProbe® (ASD Inc., Boulder, CO, USA). The spectroradiometer featured a VNIR detector with a 512-element silicon array for the 350–1000 nm range, a SWIR 1 detector with a graded index InGaAs photodiode for the 1001–1800 nm range, and a SWIR 2 detector for the 1801–2500 nm range. The standard white reference plates were used for calibration. During the measurements, a high-intensity light beam was directed at the adaxial leaf surface, and reflectance and transmittance were

recorded. The absorbance was calculated using the formula A = 1 − (R + T), ensuring precise measurements across a broad wavelength spectrum [11,29].

2.4. Sample Preparation and Microscopy

2.4.1. Sample Preparation for Microscopy

The leaf samples were prepared for optical microscopy (OM), transmission electron microscopy (TEM), and scanning electron microscopy (SEM). The samples were cut into small pieces, quickly submerged in a fixative solution of 2.5% glutaraldehyde and 2% paraformaldehyde in 0.05 M cacodylate buffer (pH 7.2), and left for six hours. The samples were fixed with a solution of 1% osmium tetroxide and 1.6% potassium ferrocyanide in the same buffer. After fixation, the samples were contrasted with 0.5% uranyl acetate for 12 h and dehydrated with a graded acetone series. Some samples were reserved for SEM, whereas others were embedded in Spurr low-viscosity epoxy resin and sectioned using an ultramicrotome [11,29].

2.4.2. Optical Microscopy

Leaf sections, 1 μm thick, were stained with 1% toluidine blue in borax buffer and briefly heated to enhance staining. A Leica ICC50 optical microscope (Leica Inc., Buffalo Grove, IL, USA) was used to observe the anatomical features, and quantitative analysis was performed using ImageJ (https://imagej.nih.gov/ij, accessed on 20 April 2024) and Image-Pro Plus version 4.5 software (Media Cybernetics Inc., Rockville, MD, USA) [11,29].

2.4.3. Scanning Electron Microscopy (SEM)

The samples prepared for SEM were dried using the CPD-030 Bal-Tec dryer (Bal-Tec AG, Balzers, Liechtenstein, Germany) critical-point drying method, mounted on aluminum stubs with carbon tape, and coated with gold. Observations were made using a Quanta 250 scanning electron microscope (FEI Company, Hillsboro, OR, USA) operating at 15–20 kV, and quantitative analysis was conducted using Image-Pro Plus version 4.5 software (Media Cybernetics Inc., Rockville, MD, USA) [11,29].

2.4.4. Transmission Electron Microscopy (TEM)

Ultrathin sections (60–70 nm) using an ultramicrotome (MTX Powertome X, Boeckeler Instruments, RMC Products, Phoenix, AZ USA) were placed on copper grids, contrasted with 3% uranyl acetate and lead citrate, and observed under a JEOL JEM-1400 transmission electron microscope Leica Microsystems Inc., Buffalo Grove, IL, USA) at 80 kV. A detailed examination of the cellular ultrastructures was performed, and the data were analyzed using the Image-Pro Plus version 4.5 software (Media Cybernetics Inc., Rockville, MD, USA) [11,29].

2.5. JIP Test and Hyperspectral Parameters Based on Hyperspectral Vegetation Indices with Optimal Wavelengths

To assess whether selecting the two most responsive hyperspectral wavelengths could enhance the predictive accuracy of eight biophysical parameters, including adaxial epidermis thickness (μm), adaxial hypodermis thickness (μm), parenchyma thickness (μm), abaxial hypodermis thickness (μm), total leaf thickness (μm), chloroplast count, granum height (nm), and thylakoid layer thickness, we examined all potential combinations of two spectral bands using the normalized difference vegetation index formula, as proposed by Crusiol et al. (2023) [51]. Each combination, representing a unique hyperspectral vegetation index (HVI), was correlated with leaf structural and ultrastructural parameters through Pearson's correlation coefficient (r) and the coefficient of determination (R^2), using a custom-written IDL code. Hyperspectral data were collected using a ground-based sensor covering the full spectral range of 350–2500 nm [25,29]. The resulting matrices were visualized as contour maps. Additionally, the reflectance, transmittance, and absorbance data for the

selected wavelengths were analyzed, as were the fluorescence responses from 0 to 1 s, expressed through OJIP curves captured by a fluorescence sensor.

$$\text{HVI} = \frac{\text{Wavelength 1} - \text{Wavelength 2}}{\text{Wavelength 1} + \text{Wavelength 2}} \quad (1)$$

2.6. Statistical Analyses

2.6.1. Descriptive, Univariate, and Multivariate Statistical Analyses

Descriptive statistics included count, mean, median, minimum, maximum, and coefficient of variation (CV). Pearson's correlation was used to explore the relationships between the biochemical and biophysical attributes. Statistical analyses were performed using Statistica 10® (StatSoft Inc., Tulsa, OK, USA) and R statistical package (R Core Team, 2020), with graphical representations created using SigmaPlot 10.0 (Systat Software, Inc., San Jose, CA, USA), Excel 365 software (Microsoft Office Professional 2024, Sunnyvale, CA, USA), and CorelDraw 2020® editor (Corel Corp., Ottawa, ON, Canada).

2.6.2. Principal Component Analysis (PCA)

PCA reduces the dimensionality of hyperspectral data by identifying the principal components that capture the most variance within a dataset. This method effectively filters out noise and highlights the most relevant spectral features, enabling more accurate analysis of biophysical parameters without overfitting or underfitting the model [53–58].

2.6.3. Spectroscopy Data Analysis via Partial Least Squares Regression (PLSR)

The hyperspectral data were mean-centered and analyzed using PLSR to develop prediction models for JIP test parameters. The data were split into training (75%) and prediction (25%) datasets. Calibration and cross-validation were used to assess model performance, with metrics such as R^2, RMSE, and RPD used to evaluate the quality, precision, and accuracy [59].

3. Results

3.1. Statistical Description of the Biophysical Compounds

Table 1 shows the results of the descriptive statistical analysis of various biophysical compounds measured in the leaves of *Tradescantia spathacea* and *Tradescantia pallida*, with 200 samples per parameter (Table 1).

Table 1. Descriptive statistics of the biophysical parameters. The data included the mean, median, minimum, maximum, and coefficient of variation for each parameter. (n = 200).

Parameter	Count (n)	Mean	Median	Minimum	Maximum	CV (%)
Adaxial epidermis (μm)	200	102.7	88.0	61	162	23.94
Adaxial hypodermis (μm)	200	323.4	325.7	212	452	25.79
Parenchyma thickness (μm)	200	252.6	251.3	224	296	3.80
Abaxial hypodermis (μm)	200	170.5	164.7	117	246	25.53
Total leaf thickness (μm)	200	849.1	820.5	735	978	7.02
Number of chloroplast	200	21.2	20.5	14	32	22.23
Granum height (nm)	200	522.5	531.9	75	1000	37.29
Thylakoid layer-granum	200	25.2	26.0	4.0	53.0	33.80

The descriptive analyses revealed that the mean thickness of the adaxial epidermis was 102.7 μm, the median thickness was 88.0 μm, the minimum thickness was 61 μm, and the maximum thickness was 162 μm, with a coefficient of variation of 23.94%. The adaxial hypodermis thickness had a mean of 323.4 μm, a median of 325.7 μm, a minimum of 212 μm, and a maximum of 452 μm, with a coefficient of variation of 25.79%. The parenchyma thickness had a mean of 252.6 μm, a median of 251.3 μm, a minimum of 224 μm, and a maximum of 296 μm, with a coefficient of variation of 3.80%. The abaxial hypodermis thickness had a mean value of 170.5 μm, a median value of 164.7 μm, a minimum value of 117 μm, and a maximum value of 246 μm, with a coefficient of variation of 25.53%. The

total leaf thickness had a mean of 849.1 µm, a median of 820.5 µm, a minimum of 735 µm, and a maximum of 978 µm, with a coefficient of variation of 7.02% (Table 1).

The mean number of chloroplasts per cell was 21.2, with a median of 20.5, a minimum of 14, and a maximum of 32, with a coefficient of variation of 22.23%. The mean granum height was 522.5 nm, the median height was 531.9 nm, the minimum height was 75 nm, and the maximum height was 1000 nm, with a coefficient of variation of 37.29%. The thickness of the thylakoid layers within the granum had a mean of 25.2 nm, a median of 26.0 nm, a minimum of 4.0 nm, and a maximum of 53.0 nm, with a coefficient of variation of 33.80% (Table 1).

3.2. Cluster Heatmap of Biophysical Parameters in Tradescantia Species

The hierarchical clustering heatmap (Figure 2) shows the biophysical parameters of the leaves of two species: *Tradescantia spathacea* (indicated in red) and *Tradescantia pallida* (indicated in green). The heatmap visualizes the Z-scores for various measured parameters, facilitating the comparison and identification of patterns across species (Figure 2).

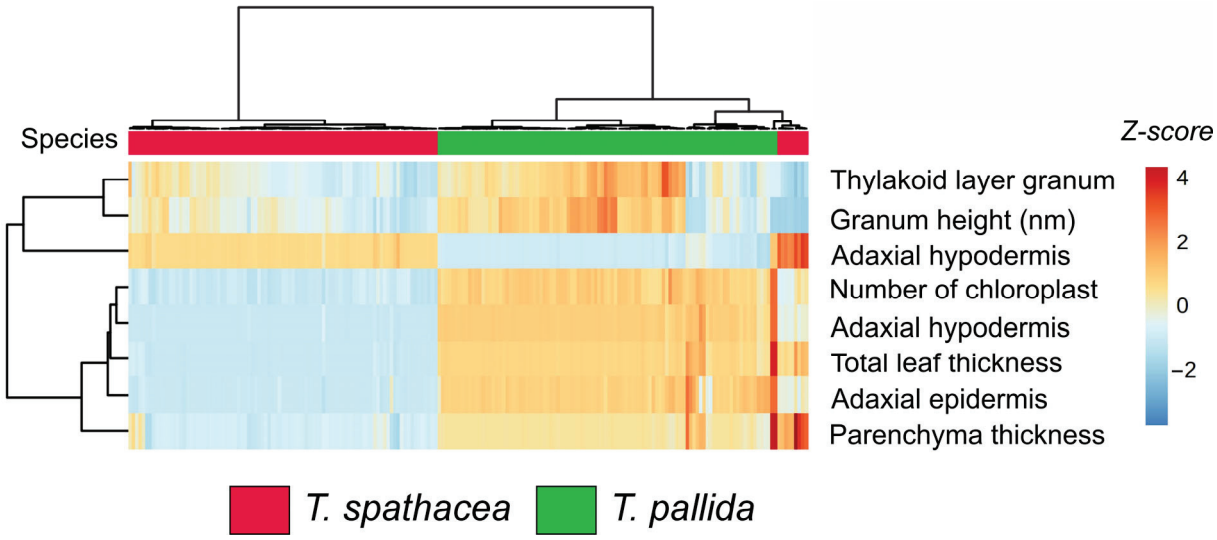


Figure 2. Cluster heatmap of biophysical parameters in leaves. Red represents *Tradescantia spathacea*, and green represents *Tradescantia pallida*. (n = 100).

The clustering dendrogram revealed distinct groupings. Specifically, *T. spathacea* presented greater Z-scores (indicated by more intense red hues) for parameters such as thylakoid layer granum, granum height, and adaxial hypodermis, suggesting that *T. spathacea* has more substantial structural attributes than *T. pallida* (Figure 2). Conversely, *T. pallida* displayed lower Z-scores (shades of blue) for these parameters, indicating relatively reduced biophysical characteristics.

Furthermore, the number of chloroplasts and total leaf thickness varied significantly among the species. *T. spathacea* tended to have a greater number of chloroplasts and greater leaf thickness, which could imply a greater photosynthetic capacity and structural robustness. The parenchyma thickness and adaxial epidermis also differed notably, with *T. pallida* exhibiting thicker parenchyma and adaxial epidermis, potentially contributing to its adaptive traits (Figure 2).

3.3. Analysis of Chlorophyll a Fluorescence and Spectral Properties

Compared with that of *T. pallida*, the chlorophyll a fluorescence of *T. spathacea* was distinctly changed (Figure S1A). The fluorescence measurements for *T. spathacea*, shown with green lines, revealed higher yields than the red–pink lines, which represent *T. pallida* at multiple observation points. We examined several parameters, including the initial fluorescence (O), intermediate phases (J, I), and peak (P). *T. spathacea* consistently presented

elevated values, indicating differences in photosynthetic competency and light absorption capacity (Figure S1). The difference in fluorescence yield was statistically significant, as evidenced by the values of F = 230.85, accuracy (Acc) = 0.98, and Kappa (K) = 0.96 (refer to Figure S1A).

The reflectance spectra showed considerable variations between the two species at specific wavelengths. In the UV–VIS spectrum (350–700 nm), a notable difference was observed at 544 nm, where *T. spathacea* had a higher reflectance. The most significant difference is observed at 700 nm in the NIR range (700–1300 nm). In the SWIR1 spectrum (1300–1800 nm), the key difference appeared at 1441 nm, and in the SWIR2 range (1800–2500 nm), it appeared at 2488 nm. These results were statistically significant, with F = 58.71, Acc = 0.90, and K = 0.80, underscoring the importance of these wavelengths for species differentiation based on reflectance properties (Figure S1B).

Transmittance measurements revealed significant differences at specific wavelengths. In the UV–VIS spectrum, the largest difference was observed at 400 nm. In the NIR spectrum, it was 1125 nm, in the SWIR1 spectrum at 1607 nm, and in the SWIR2 spectrum at 2175 nm. These results, with low p-values, suggest the potential of using transmittance at these wavelengths to distinguish between the two species. The transmittance data were characterized by F = 47.17, Acc = 0.85, and K = 0.70 (Figure S1C).

The absorbance measurements revealed notable differences at certain wavelengths across the various spectral ranges. Some wavelengths were 495, 498, 733, and 800 nm in the NIR region, 1575 nm in the SWIR1 region, and 2175 nm in the SWIR2 region. The low p-values confirmed the statistical importance of these wavelengths, indicating their utility in distinguishing between *T. spathacea* and *T. pallida*. The absorbance data revealed F = 87.94, Acc = 0.97, and K = 0.94 (Figure S1D). Overall, the differences observed in the reflectance, transmittance, and absorbance spectra highlight the distinct optical properties of the two *Tradescantia* species (Figure S1).

3.4. Principal Component Analysis of the Spectral Data

Figure 3 shows the use of principal component analysis (PCA) to determine the spectral properties of the leaves of *T. spathacea* and *T. pallida* via fluorescence, reflectance, transmittance, and absorbance readings (Figure 3A–D). For example, Figure 3A shows the PCA results for the fluorescence sensors, with *T. spathacea* indicated by red dots and *T. pallida* indicated by green dots. The first principal component (PC1) accounted for 92% of the variability, clearly distinguishing the two plants based on their fluorescence signatures, whereas the second principal component (PC2) explained 6% of the variability (Figure 3).

The PCA of the reflectance data in Figure 3B shows a clear separation along PC1, which accounts for 60% of the variation, and PC2 elucidates an additional 25%, demonstrating unique reflectance characteristics between the species. The PCA results are depicted in Figure 3C, revealing a marked contrast between *T. spathacea* and *T. pallida* along both PC1 and PC2, with PC1 accounting for 77% of the total variance and PC2 contributing 17%, indicating differential species-specific transmittance traits. Figure 3D shows the PCA related to the absorbance sensors, with PC1 representing 72% of the variation and PC2 representing another 18%. These data suggest significant differences in absorbance across species.

The shaded areas in each plot represent the clustering of individual observations showing intraspecific variability. Overall, the PCA results demonstrated that these spectral assessments were effective for differentiating *T. spathacea* from *T. pallida*, as shown in Figure 3.

Figure 4 shows an in-depth analysis of the variance, β-loading, and regression coefficients across the different spectral measurements for fluorescence, reflectance, transmittance, and absorbance in the two *Tradescantia* species. This figure shows the contribution of the principal components to the interpretation of spectral data.

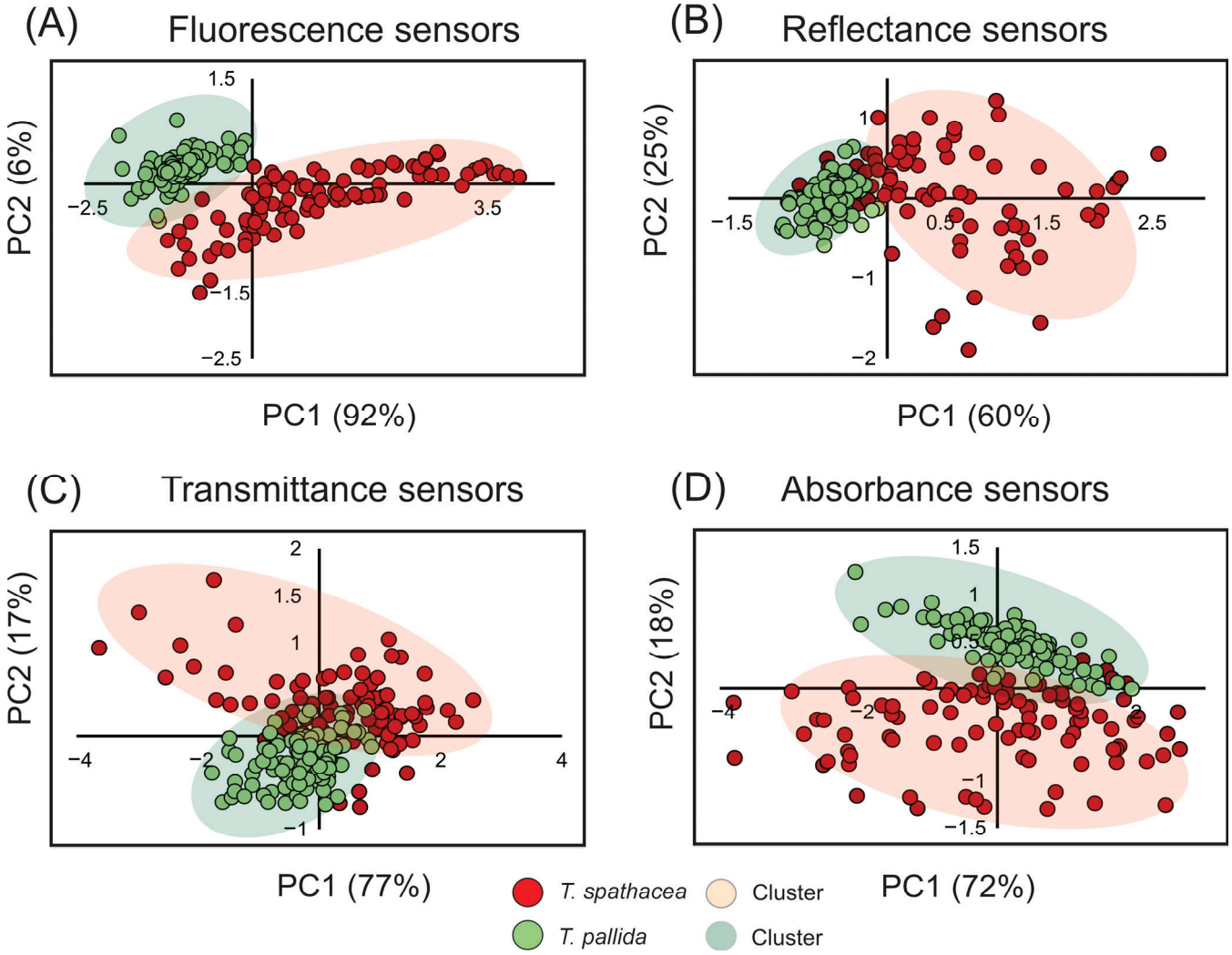


Figure 3. Principal component analysis (PCA) of spectral data from 350 to 2500 nm. (**A**) Fluorescence sensor data. (**B**) Hyperspectral reflectance sensor data. (**C**) Hyperspectral transmittance sensor data. (**D**) Hyperspectral absorbance sensor data. Clusters are indicated by red circles for *T. spathacea* and green circles for *T. pallida*. (n = 100).

Figure 4A–D depicts the variance explained by the principal components for each type of spectral measurement. The fluorescence data (Figure 4A) revealed that PC1 accounted for 90.27% of the variance, with the subsequent components contributing significantly less. For the reflectance data (Figure 4B), PC1 explained 60.05% of the variance, while PC2 explained 24.68%. According to the transmittance data (Figure 4C), PC1 captured 77.09% of the variance, while PC2 accounted for 15.50%. The absorbance data (Figure 4D) indicate that PC1 explained 72.45% of the variance, with smaller contributions from PC2 (18.08%).

The β-loading plots (Figure 4E–H) show the impact of each principal component across the time or wavelength range, highlighting the influence of specific spectral bands. For the fluorescence sensors (Figure 4E), the β-loadings of PC1, PC2, and PC3 showed significant contributions across the measured time points. Reflectance data (Figure 4F) identified key wavelengths for PC1 in the blue, green, red, NIR, SWIR1, and SWIR2 bands, with similar influential wavelengths for PC2 and PC3. The transmittance data (Figure 4G) highlight the crucial wavelengths for PC1 in the blue and SWIR2 bands, with consistent influential spectral bands for PC2 and PC3. The absorbance measurements (Figure 4H) for PC1 indicated key wavelengths spanning from the blue to SWIR2 range, which is consistent with the results for PC2 and PC3.

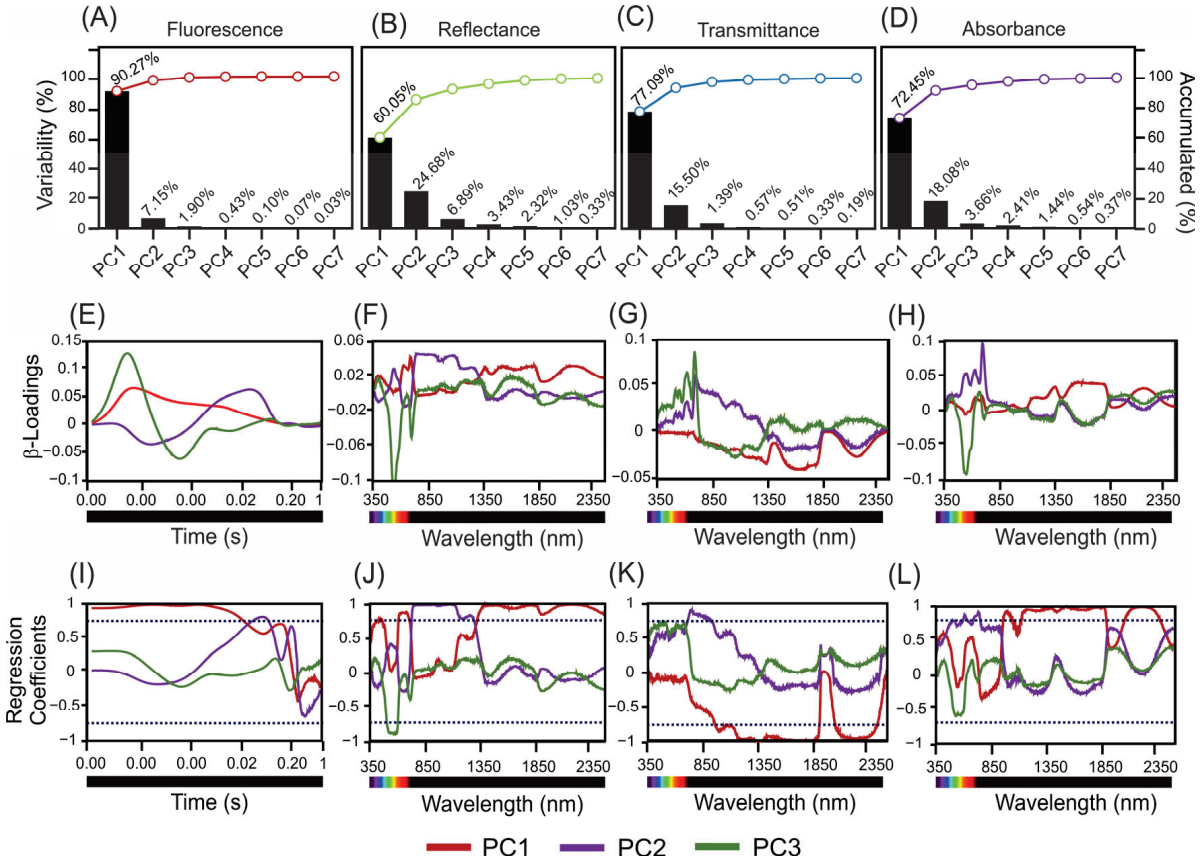


Figure 4. Present scores, β-loadings, and regression coefficients derived from sensors. (**A**–**D**) Variability and cumulative percentages are explained by three principal components. (**E**–**H**) β-Loading associated with hyperspectral and fluorescence sensors. (**I**–**L**) Regression coefficients for PC1–3 in the 350–2500 nm range. The dark blue dotted lines indicate −0.70 and 0.70 correlation of coefficients.

The regression coefficient plots (Figure 4I–L) further illustrate the influence of each wavelength on the respective spectral measurements, identifying the most impactful wavelengths. The regression coefficients for fluorescence (Figure 4I) confirmed a significant influence across the measured time-points. The reflectance data (Figure 4J) underscore the importance of the blue, green, red, NIR, SWIR1, and SWIR2 bands, whereas the transmittance data (Figure 4K) highlight the wavelengths in the blue and SWIR2 bands. The absorbance data (Figure 4L) validated the relevance of wavelengths ranging from blue to SWIR2.

Overall, the detailed analysis in Figure 4 highlights the significant contributions of specific wavelengths and spectral bands to the variance in the fluorescence, reflectance, transmittance, and absorbance measurements. This result showed the most influential spectral regions for distinguishing *T. spathacea* from *T. pallida*, providing valuable insights for further ecological and physiological research (Figure 4).

3.5. Evaluation of Hyperspectral Vegetation Indices via Fluorescence, Reflectance, Transmittance, and Absorbance Sensors

Figure 5 shows contour plots of the coefficient of determination (R^2) values from linear regression analyses that explored the relationships between various biophysical compounds and chlorophyll fluorescence (ChlF) over a time span ranging from 0.00001 to 1 s. These plots emphasize the strength of the correlation between fluorescence signals and structural leaf attributes, such as adaxial epidermis, adaxial hypodermis, parenchyma thickness, abaxial hypodermis, total leaf thickness, chloroplast count, granum height, and thylakoid layer granum.

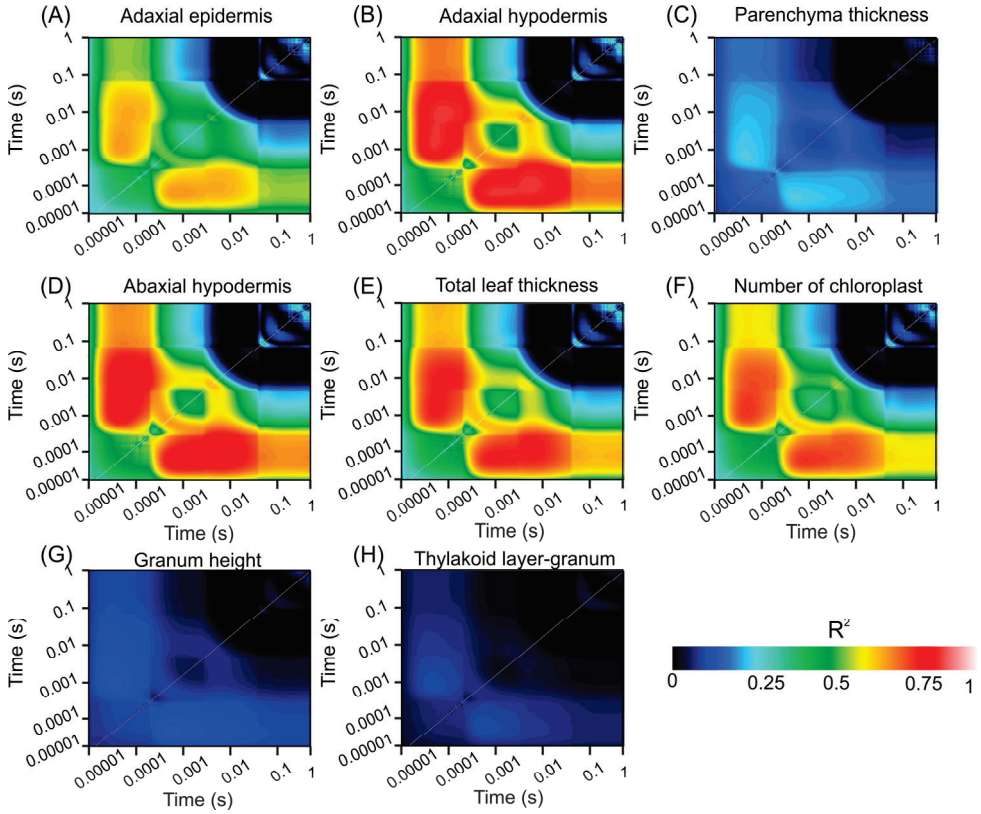


Figure 5. Count plot map of the coefficient of correlation (R^2). (**A**) Adaxial epidermis. (**B**) Adaxial hypodermis. (**C**) Parenchyma thickness. (**D**) Adaxial hypodermis. (**E**) Total leaf thickness. (**F**) Number of chloroplasts. (**G**) Granum height. (**H**) Thylakoid layer-granum. Dark blue to light red scale indicates increased associations. (n = 200).

Figure 5A displays the R^2 values for the adaxial epidermis, which were significantly correlated at the early time points. The transition from dark blue to light red signifies a strong relationship, indicating a rapid response of the adaxial epidermis to fluorescence changes. Similarly, Figure 5B (adaxial hypodermis) and Figure 5D (abaxial hypodermis) show significant early correlations, highlighting the crucial role of these layers in ChlF dynamics.

In contrast, Figure 5C shows lower R^2 values for parenchyma thickness across all time points, depicted predominantly by blue hues, suggesting a weaker association with the fluorescence parameters. Figure 5E (total leaf thickness) and Figure 5F (chloroplast count) show notable correlations at early and intermediate time points, with higher R^2 values (green to red transition), indicating their significant roles in ChlF characteristics.

Figure 5G,H show minimal correlations for the granum height and thylakoid layer–granum ratio, respectively, as indicated by the extensive dark blue regions, suggesting a lesser impact on fluorescence within the measured timeframe.

These contour plots visually represent the temporal and quantitative relationships between the chlorophyll fluorescence dynamics and various leaf structural components. The color gradation from dark blue to light red highlights the strongest relationships, providing insight into the most informative moments for ChlF concerning these structural attributes.

Figures S2–S4 present contour plots that illustrate the coefficient of determination (R^2) values from linear regression analyses, which were used to assess the relationships between the hyperspectral data (reflectance, transmittance, and absorbance) at two wavelengths and various leaf structural attributes. These attributes include adaxial epidermis, adaxial hypodermis, parenchyma thickness, abaxial hypodermis, total leaf thickness, chloroplast count, granum height, and thylakoid layer-to-granum ratio.

Figure S2 displays the R^2 values for the hyperspectral reflectance data. In Figure S2A, the adaxial epidermis shows strong correlations at specific wavelengths, particularly in the

blue and near-infrared (NIR) regions, as indicated by the gradient from dark blue to light red. Similarly, Figure S2B shows significant correlations for the adaxial hypodermis, with high R^2 values concentrated in the blue- and red-edge regions. Conversely, parenchyma thickness (Figure S2C) generally resulted in lower R^2 values across most wavelength pairings, implying a weaker relationship with the reflectance data.

Figure S2D highlights the R^2 values for the abaxial hypodermis, which are strongly correlated at various wavelengths, particularly in the green and NIR regions. Figure S2E,F show notable correlations for total leaf thickness and chloroplast count, respectively, at several wavelengths. Strong correlations for total leaf thickness were observed in the blue and NIR regions, whereas correlations for chloroplast count were more dispersed. Figure S2G,H depict the minimal correlations for the granum height and thylakoid layer-to-granum ratio, respectively, suggesting a lesser impact on the reflectance within the measured spectral range.

Figure S3 shows contour maps of the R^2 values obtained from the linear regression analysis of the hyperspectral transmittance data. Figure S3A shows strong correlations for the adaxial epidermis at specific wavelengths within the visible (VIS) and NIR spectra. Similarly, Figure S3B shows significant correlations for the adaxial hypodermis, with high R^2 values in the VIS and NIR spectra. For parenchyma thickness (Figure S3C), lower R^2 values were observed across various wavelength combinations, indicating a less robust correlation. Figure S3D highlights the significant correlations for the abaxial hypodermis, especially in the green and NIR regions. Figure S3E,F show discernible correlations for the total leaf thickness and chloroplast count, respectively, across multiple wavelengths. Finally, Figure S3G,H present the nominal correlations for the granum height and thylakoid layer-to-granum ratio, respectively.

Figure S4 shows the R^2 values obtained from the linear regression analysis of the hyperspectral absorbance data. Figure S4A shows significant correlations for the adaxial epidermis at specific wavelengths, particularly in the VIS and NIR regions at approximately 700 and 1600 nm. Figure S4B shows notable correlations for the adaxial hypodermis, with high R^2 values concentrated at approximately 700 and 1500–1700 nm. Figure S4C shows lower R^2 values for parenchyma thickness, suggesting a weaker relationship with absorbance data. Figure S4D highlights the strong correlations for the abaxial hypodermis, especially in the green and NIR regions. Figure S4E,F show notable correlations for total leaf thickness and chloroplast count, respectively, at several wavelengths. Finally, Figure S4G,H show minimal correlations for the granum height and thylakoid layer-to-granum ratio, respectively, suggesting a lesser impact on the absorbance within the measured spectral range.

3.6. Parameters Predicted by Biophysical Compounds

Calibration, Validation and Prediction Models

Table 2 and Table S1 present the statistical metrics from the partial least squares regression (PLSR) models used to predict the biophysical parameters of *Tradescantia* species via various spectrometric methods. These models were assessed based on their calibration and cross-validation performance using fluorescence, reflectance, transmittance, and absorbance measurements.

The calibration model for adaxial epidermal thickness using fluorescence sensors had an R^2 of 0.92, which remained consistent during cross-validation. The root mean square error (RMSE) values were 6.0 for calibration and 6.3 for cross-validation, with relative prediction deviations (RPDs) of 2.62 and 2.52, respectively, indicating model robustness. For the adaxial hypodermis, the model performed similarly well, with an R^2 of 0.93 in both calibration and cross-validation, RMSE values of 21.5 and 22.5, and RPD values of 2.79 and 2.68, respectively. The parenchyma thickness predictions exhibited moderate accuracy, with calibration and cross-validation R^2 values of 0.69 and 0.66, RMSE values of 3.0 and 3.2, and RPD values of 1.38 and 1.34, respectively.

Table 2. The performance of partial least squares regression (PLSR) models that use various spectroscopic techniques to estimate biophysical compound concentrations in *Tradescantia* species. The evaluation metrics included the correlation coefficient (r), R-squared (R^2), slope, offset, standard error of prediction (SEP), root mean squared error of prediction (RMSEP), ratio of performance to deviation (RPD), and bias, offering a detailed analysis for each sensor. (n = 50).

Sensors	Parameter	Maximum Factor PLS	Predicted							
			r	R^2	Slope	Offset	SEP	RMSEP	RPD	Bias
Fluorescence	Adaxial epidermis (μm)	5	0.66	0.44	0.54	45.80	22.80	22.70	1.33	2.14
	Adaxial hypodermis (μm)	5	0.86	0.74	0.95	17.80	47.50	47.10	1.95	2.06
	Parenchyma thickness (μm)	4	0.42	0.18	0.14	219.01	14.00	14.20	1.10	2.92
	Abaxial hypodermis (μm)	5	0.42	0.17	1.30	152.80	43.70	88.80	1.10	77.53
	Total leaf thickness (μm)	5	0.77	0.59	0.77	203.70	43.40	43.60	1.56	7.28
	Number of chloroplast	5	0.76	0.57	0.75	5.50	3.40	3.40	1.53	0.36
	Granum height (nm)	2	0.20	0.04	0.09	522.20	204.90	246.00	1.02	138.62
	Thylakoid layer-granum	2	0.10	0.01	0.04	25.30	11.00	11.40	1.01	3.30
Reflectance	Adaxial epidermis (μm)	7	0.77	0.59	0.48	51.90	19.60	19.50	1.57	2.40
	Adaxial hypodermis (μm)	6	0.94	0.88	0.81	62.00	29.10	28.90	2.91	1.80
	Parenchyma thickness (μm)	6	0.26	0.07	0.07	236.50	14.80	15.00	1.04	3.10
	Abaxial hypodermis (μm)	7	0.94	0.88	0.79	39.20	15.80	16.30	2.84	4.50
	Total leaf thickness (μm)	7	0.88	0.77	0.68	274.30	31.00	31.50	2.10	6.90
	Number of chloroplast	7	0.88	0.77	0.67	7.20	2.50	2.40	2.09	0.40
	Granum height (nm)	7	0.11	0.01	0.06	539.10	217.10	256.80	1.01	140.00
	Thylakoid layer-granum	7	0.03	0.00	0.01	25.60	11.60	11.90	1.00	3.10
Transmittance	Adaxial epidermis (μm)	7	0.76	0.57	0.55	41.20	19.50	20.10	1.53	5.30
	Adaxial hypodermis (μm)	7	0.92	0.84	0.91	18.60	33.70	34.70	2.51	9.30
	Parenchyma thickness (μm)	6	0.25	0.06	0.08	234.30	15.00	15.30	1.03	3.80
	Abaxial hypodermis (μm)	7	0.92	0.84	0.91	26.70	17.50	20.50	2.52	10.80
	Total leaf thickness (μm)	7	0.86	0.74	0.77	196.50	32.30	32.00	1.96	0.10
	Number of chloroplast	6	0.82	0.68	0.77	4.70	2.80	2.80	1.76	0.20
	Granum height (nm)	7	0.14	0.02	0.07	503.30	210.80	236.10	1.01	109.60
	Thylakoid layer-granum	6	0.13	0.02	0.04	24.20	10.90	11.10	1.01	2.40
Absorbance	Adaxial epidermis (μm)	7	0.78	0.60	0.56	42.20	18.90	19.00	1.59	2.90
	Adaxial hypodermis (μm)	7	0.94	0.89	0.95	14.90	27.90	27.70	3.03	0.50
	Parenchyma thickness (μm)	7	0.25	0.06	0.08	235.00	14.90	15.20	1.03	3.40
	Abaxial hypodermis (μm)	7	0.94	0.89	0.93	17.10	14.80	15.90	2.96	6.10
	Total leaf thickness (μm)	7	0.88	0.78	0.80	174.90	29.20	30.30	2.12	6.00
	Number of chloroplast	7	0.88	0.77	0.82	4.30	2.40	2.40	2.09	0.40
	Granum height (nm)	7	0.02	0.00	0.01	556.80	230.70	267.60	1.00	138.90
	Thylakoid layer-granum	6	0.02	0.00	0.01	26.10	12.10	12.40	1.00	3.10

Reflectance measurements for the adaxial epidermis yielded calibration and cross-validation R^2 values of 0.88 and 0.87, respectively. The RMSE values were 7.4 for calibration and 8.0 for cross-validation, with RPD values of 2.14 and 2.00, respectively. For the adaxial hypodermis, high accuracy was achieved, with R^2 values of 0.91 for calibration and 0.89 for cross-validation, RMSE values of 25.6 and 27.6, and RPD values of 2.36 and 2.20, respectively. Parenchyma thickness predictions were moderately accurate, with calibration and cross-validation R^2 values of 0.70 and 0.65, RMSE values of 3.0 and 3.2, and RPD values of 1.39 and 1.32, respectively.

Transmittance measurements for the adaxial epidermis indicated R^2 values of 0.88 for calibration and 0.87 for cross-validation, with RMSE values of 7.4 and 7.9, and RPD values of 2.14 and 2.01, respectively. For the adaxial hypodermis, high accuracy was demonstrated, with calibration and cross-validation R^2 values of 0.90 and 0.88, RMSE values of 26.3 and 28.3, and RPD values of 2.29 and 2.15, respectively. The R^2 values for predicting parenchyma thickness were 0.75 for calibration and 0.70 for cross-validation, with RMSE values of 2.7 and 3.0 and RPD values of 1.52 and 1.40, respectively.

The absorbance measurements for the adaxial epidermis achieved R^2 values of 0.89 for calibration and 0.88 for cross-validation, with RMSE values of 7.1 and 7.7 and RPD values of 2.23 and 2.07, respectively. The adaxial hypodermis model showed high accuracy, with calibration and cross-validation R^2 values of 0.92 and 0.90, RMSE values of 23.8 and 26.0, and RPD values of 2.53 and 2.33, respectively. The predictions of parenchyma thickness showed moderate accuracy, with calibration and cross-validation R^2 values of 0.73 and 0.67, RMSE values of 2.8 and 3.1, and RPD values of 1.47 and 1.35, respectively.

For the prediction parameters of fluorescence, the model for adaxial epidermis thickness demonstrated moderate predictability, with an R^2 of 0.44 and an RMSEP of 22.7. The adaxial hypodermis showed higher accuracy, with an R^2 of 0.74 and an RMSEP of 47.1, indicating model robustness. In contrast, parenchyma thickness and abaxial hypodermis had limited predictive accuracy, with R^2 values of 0.18 and 0.17, respectively. The model for the chloroplast count had a moderate R^2 of 0.57 and an RMSEP of 3.4, suggesting reasonable reliability, as shown in Table 2.

Reflectance-based models generally produce relatively high R^2 values. The adaxial epidermis model had an R^2 of 0.59, whereas the adaxial hypodermis model had an R^2 of 0.88, indicating better predictability than fluorescence. However, parenchyma thickness showed limited accuracy, with an R^2 of 0.07. The model for total leaf thickness exhibited a notable R^2 value of 0.77, demonstrating a strong predictive power.

Transmittance measurements yielded robust R^2 values for both the adaxial and abaxial hypodermis, at 0.84, with RMSEP values of 34.7 and 20.5, respectively. The model for total leaf thickness also had a high R^2 value of 0.74, indicating strong predictability.

The absorbance-based models improved the R^2 values for most parameters. The adaxial epidermis model had an R^2 of 0.60, whereas the adaxial hypodermis model achieved an R^2 of 0.89, both of which indicated moderate-to-strong predictive capabilities. The model for total leaf thickness exhibited a strong R^2 value of 0.78, indicating high reliability.

Overall, the PLSR models displayed varying degrees of predictability across different spectroscopic methods. Fluorescence and reflectance methods generally provided moderate predictability, whereas transmittance and absorbance methods showed greater accuracy for specific parameters. The RMSEP and RPD values indicated the precision and reliability of these models, with higher RPD values suggesting superior predictive capabilities, particularly for structural parameters, such as the adaxial and abaxial hypodermis. The bias values highlighted areas for potential model refinement to enhance the prediction accuracy.

A comparison of the predicted and observed values in the scatter plots revealed varying prediction accuracies across the different spectroscopic methods.

For the fluorescence sensor (Figure S5), strong correlations were observed for adaxial and abaxial hypodermis thickness ($R^2 = 0.94$) and total leaf thickness ($R^2 = 0.87$), indicating a high predictive power. However, parameters such as parenchyma thickness ($R^2 = -0.01$) and thylakoid layer–granum ratio ($R^2 = 0.05$) exhibited weak correlations, suggesting poor predictability.

The reflectance data (Figure S6) showed robust R^2 values for the adaxial hypodermis thickness (0.8) and abaxial hypodermis thickness (0.91), indicating strong predictive relationships. Similar to fluorescence, parenchyma thickness ($R^2 = 0.18$) and the thylakoid layer–granum ratio ($R^2 = 0.05$) were predicted less accurately.

The transmittance data (Figure S7) displayed moderate to strong correlations for most parameters. The thicknesses of the adaxial ($R^2 = 0.87$) and abaxial hypodermis ($R^2 = 0.92$) were highly predictable. However, parenchyma thickness ($R^2 = 0.16$) and thylakoid layer–granum ratio ($R^2 = 0.07$) showed weak predictability.

The absorbance data (Figure S8) demonstrated moderate predictive accuracy, with strong correlations between adaxial hypodermis thickness ($R^2 = 0.94$) and total leaf thickness ($R^2 = 0.8$). Parenchyma thickness ($R^2 = -0.04$) and the thylakoid layer–granum ratio ($R^2 = 0.06$) were poorly correlated, indicating limitations in the use of absorbance for these measurements.

In summary, the analysis revealed that parameters such as hypodermis and total leaf thickness, particularly fluorescence and reflectance, were consistently well predicted across all methods. Conversely, parameters such as parenchyma thickness and the thylakoid layer–granum ratio exhibited low predictability, highlighting areas where model refinement was needed. A higher green color intensity in the absorbance data plots indicates stronger correlations for specific parameters, suggesting that absorbance could be a more reliable predictor of biophysical compounds.

4. Discussion

Considerable progress has been made by integrating hyperspectral sensors to predict the biophysical parameters of plants. Chlorophyll fluorescence and hyperspectral sensors collect extensive data, facilitating detailed and precise predictions that are closely aligned with the physiological and biophysical functions. The incorporation of hyperspectral vegetation indices (HVIs), along with diverse analytical tools, has led to significant enhancements in monitoring and forecasting various structural and ultrastructural compounds [60,61].

Our fluorescence, reflectance, transmittance, and absorbance measurements demonstrated varying degrees of predictive accuracy. For example, chlorophyll a fluorescence was highly effective in predicting parameters such as adaxial and abaxial hypodermal thickness (R^2 = 0.94) and total leaf thickness (R^2 = 0.87). Similarly, reflectance data showed strong predictive power for adaxial hypodermis thickness (R^2 = 0.80) and abaxial hypodermis thickness (R^2 = 0.91). Transmittance readings also yielded robust predictions for adaxial (R^2 = 0.87) and abaxial hypodermal thickness (R^2 = 0.92). Absorbance measurements provided high predictive accuracy for adaxial hypodermis thickness (R^2 = 0.94) and total leaf thickness (R^2 = 0.80). Other studies that sought to relate different spectral curves using similar approaches reported comparable values [24,25,45]. The findings presented here align with the key advancements our research group has made from 2017 to 2024, emphasizing the importance of analyzing spectral curves using reflectance, transmittance, and absorbance [10,11,29–31,59,62–66]. These distinct modes of spectral acquisition capture the varied interactions between the electromagnetic spectrum and leaf optical properties, offering a more accurate representation of the leaf biophysical processes. Additionally, our recent integration of fluorescence sensors with infrared gas analysis (IRGA) has improved the analysis of not only biochemical aspects but also the structural and ultrastructural changes across different leaf types, enhancing our overall understanding of leaf biophysics and plant function [10,11,29–31,59,62–66].

However, some parameters, such as parenchyma thickness and the thylakoid layer–granum ratio consistently exhibited lower predictability across all sensor types. This suggests that the current models may not fully capture the complexity of these structures, potentially owing to limitations in sensor sensitivity or spectral resolution for these specific measurements. The variability in tissue composition, structural heterogeneity, and influence of overlapping biophysical factors could also contribute to reduced accuracy. These findings indicate that further model refinement, perhaps through the integration of additional sensor types or advanced analytical techniques, is necessary to improve the predictive accuracy of these parameters. These limitations highlight the need for a more tailored approach to measuring traits that exhibit subtle optical responses [5,11,29].

The predictive models employed in this study highlight the importance of integrating multiple sensors. The high R^2 values for the key parameters validate the reliability of these models in estimating various pigment concentrations, establishing a basis for future research and practical applications in precision agriculture. The strong performance of our models across different spectroscopic methods confirmed their potential for improving plant health assessments and environmental monitoring.

This is consistent with existing studies that highlight the importance of non-imaging hyperspectral sensors in remote sensing and improve sustainability on the basis of many factors, such as photochemical, biochemical, and metabolic processes, as well as the mechanisms of plant photosynthetic reactions and their interactions with biotic and abiotic stresses on the physiological states of plants and green chemistry. The interactions between light and plant tissues, such as absorption, reflection, and transmission, provide crucial insights into physiological processes. In other studies, the interactions for the precise prediction of plant compounds enhanced the field of remote sensing and facilitated the use of optical spectroscopy for further sample characterization [3,17,19].

Hyperspectral vegetation indices (HVIs) are particularly effective for assessing a broad spectrum of physiological, biochemical, and structural parameters. These indices

are vital for monitoring environmental stresses such as water and nitrogen deficiencies. For example, narrowband HVIs are highly sensitive to fluctuations in water and nitrogen levels, allowing for detailed evaluation of crop responses under different irrigation and fertilization conditions [51,67].

Combining HVIs with hyperspectral measurements increases the precision of agricultural interventions, optimizes resource utilization, and enhances crop management strategies. The clear spectral differences observed in the reflectance, transmittance, and absorbance curves of *T. spathacea* and *T. pallida* further underscore the utility of these methods for species differentiation and physiological and green chemistry evaluations [11].

4.1. Hyperspectral and Principal Component Analyses of Reflectance, Transmittance, and Absorbance

Hyperspectral sensors reveal detailed spectral data that facilitate comprehensive analysis of plant biochemical and biophysical properties. Reflectance measurements provide essential insights into both the surface and internal structures of leaves, revealing variations in the chlorophyll content and potential cellular changes. This finding is consistent with previous research, suggesting that changes in reflectance can indicate physiological alterations in plant tissues. Similarly, transmittance data are critical for understanding the light penetration and distribution within a leaf, offering insights into its structural and functional attributes.

Given the complexity of hyperspectral datasets, advanced data-reduction techniques, such as principal component analysis (PCA), are indispensable. In this study, PCA was successfully applied to identify the key spectral regions associated with variations in chlorophyll a fluorescence sensors. This is important for developing robust sensors capable of real-time monitoring of multiple biophysical parameters associated with complex environments, which significantly enhances the precision of sensor-fusion methods for measurements. The predictive performance of these hyperspectral measurements was rigorously evaluated with a focus on the coefficients of variation and value ranges to ensure data reliability across the electromagnetic spectrum. Our findings demonstrate that appropriately applied passive sensor techniques provide detailed insights into plant health and photosynthetic efficiency, which are crucial for accurate plant phenotyping and leaf tissue alterations [24,68,69].

Our study also investigated the integration of various sensor types, including hyperspectral and chlorophyll fluorescence sensors, to enhance the predictive modeling capabilities [12,70]. This combination broadens the applicability of the data, improves the accuracy of the predictions, and monitors diverse plant health parameters. Therefore, hyperspectral non-imaging techniques have the potential to revolutionize plant physiological research and agricultural practices. By providing a deeper molecular-level understanding of plant dynamics, these methods offer rapid and precise insights into plant health and phenotyping, which are essential for improving agricultural efficiency and accuracy [71].

Moreover, our results align with recent advancements in hyperspectral technology that emphasize the importance of integrating reflectance, transmittance, and absorbance data for a more comprehensive analysis of plant health, which aligns with tissue, cell, and organelle analyses. Advanced algorithms and machine learning techniques are now being used to further increase the predictive power and accuracy of hyperspectral data, making them vital tools in modern agricultural practice [17,28,46].

Additionally, recent research has demonstrated the potential of hyperspectral non-imaging to detect early signs of plant stress before they become visually apparent [11,17,28,29,46]. This early detection capability is crucial for implementing timely interventions to mitigate stress factors such as nutrient deficiencies, pest infestations, or water stress. By identifying spectral signatures associated with specific stress responses, hyperspectral imaging can facilitate the development of targeted treatment strategies, thereby increasing crop resilience and yield [3,24,44,72].

Furthermore, the integration of hyperspectral data with other remote-sensing technologies, such as LiDAR and thermal imaging, presents a promising avenue for a more

holistic understanding of plant health. LiDAR can provide detailed three-dimensional (3D) structural information, whereas thermal imaging offers insights into plant water status and temperature regulation. Combining these datasets with hyperspectral information allows for a multifaceted analysis of plant physiology, leading to more precise and comprehensive monitoring systems [1,40,73].

Finally, the scalability of hyperspectral non-imaging technology is crucial for its application in agriculture. Advances in sensor miniaturization and cost reduction have made these sensors more accessible for both small- and large-scale farming. In this study, combining two sensors to measure reflectance and transmittance and then calculating absorbance, along with fluorescence measurements, provided a more comprehensive analysis of leaf optical properties. This approach enhances the ability to detect structural and biochemical changes in plants and offers deeper insights than the use of a single sensor alone. In this sense, the development of portable hyperspectral devices and their integration with UAVs enables high-resolution regular monitoring over large areas. This supports precision agriculture by improving resource efficiency and reducing environmental impacts. Similar studies have highlighted that the use of multiple sensors not only improves measurement accuracy but also expands the range of observable traits, providing more detailed information for advancing agricultural research [11,17,28,29,46].

4.2. Predictive Modeling-Based Reflectance, Transmittance, and Absorbance

Merging hyperspectral data with partial least squares (PLS) modeling has emerged as a transformative approach in plant science, greatly improving our ability to predict and understand intricate physiological traits. Integrating measurements of reflectance, transmittance, and absorbance with advanced multivariate algorithms, such as PLS, highlights the essential role of these parameters in reflecting photosynthetic efficiency and electron transport chain functionality. This integration significantly enhances the precision of the predictive models [74,75].

Variable Importance in Projection (VIP), many machine and deep learning algorithm scores, and hyperspectral vegetation indices (HVIs) are critical for identifying the most informative wavelengths in hyperspectral datasets, thus improving the accuracy of predictive models and allowing for a detailed understanding of plant characteristics. These methods underscore the importance of identifying spectral regions that are rich in information and are closely related to the physiological state of plants [63,76,77].

Principal component analysis (PCA) is employed to handle the complexity of large-scale hyperspectral data by pinpointing key spectral regions associated with variations in chlorophyll a fluorescence parameters. These spectral signatures are crucial for creating remote sensing tools that can noninvasively evaluate plant properties such as chlorophyll and nitrogen contents [78–80].

The predictive models derived from this integrated approach enable real-time analysis, thereby reinforcing the efficacy of remote sensing for extensive monitoring. This aligns with the current views on the transformative potential of these technologies for global monitoring and productivity management. Real-time monitoring is crucial for informed decision-making in agriculture, forestry, and other sectors that rely on remote sensing, enhancing productivity, and sustainability [2,81].

Importantly, although significant progress has been made in applying hyperspectral and PLS models, further research is necessary to extend these techniques to diverse plant species and environmental conditions. Whole-spectrum models, such as PLSR, LDA, and SVR, have shown great efficacy in classifying and predicting leaf properties. These models provide valuable insights for interpreting chlorophyll a fluorescence data, which is crucial for JIP test evaluations. By utilizing the full spectrum of available data, these models ensure a thorough analysis of complex leaf attributes, including the interactions between leaf optical properties and the intricate chemistry of molecules in variegated leaves [45,82].

In conclusion, a variety of spectroscopic methods, including those with and without non-imaging (proximal sensor) functions, have been extensively employed to decipher de-

tails about photochemical, biochemical, biophysical, and metabolic activities, as well as to understand how plant photosynthesis reacts to and interacts with living organisms and environmental stressors, influencing the physiological conditions of plants by accurate responses.

5. Conclusions

The use of hyperspectral and chlorophyll detection technologies highlights their efficacy in assessing biochemical elements in *Tradescantia* species, thereby offering a thorough understanding of plant health and structural anomalies. These methodologies demonstrated high accuracy and reliability through rigorous statistical analyses, confirming their suitability for sensor technology and chemometric studies. Identifying critical spectral zones spanning from ultraviolet—blue to shortwave infrared is essential for noninvasive biochemical measurements and cellular structure detection. The use of hyperspectral vegetation indices (HVIs) significantly improves the performance of partial least squares (PLS) regression models. The precision of these models underscores the importance of stationary hyperspectral and fluorescence probes for analyzing diverse foliage. The insights from this research establish a crucial framework for integrating various standoff sensors and hyperspectral imaging methods in botanical studies. Future research should expand these techniques to encompass a wider variety of plant species and environmental conditions to enhance the robustness and applicability of these models. Advancing these approaches will confirm the pivotal role of hyperspectral sensing technologies in agricultural surveillance and management.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/s24196490/s1, Figure S1: Spectral leaf reflectance, transmittance, and absorbance profiles from 350 to 2500 nm. (A) Chlorophyll a fluorescence kinetics. (B) Reflectance spectra. (C) Transmittance spectra. (D) Absorbance spectra. In these plots, *T. spathacea* is shown using red–pink lines, while *T. pallida* is represented by green lines. Accuracy (Acc) and Kappa coefficient (K) are noted. (n = 100); Figure S2: Reflectance by count plot map of the coefficient of correlation (R^2). (A) Adaxial epidermis. (B) Adaxial hypodermis. (C) Parenchyma thickness. (D) Adaxial hypodermis. (E) Total leaf thickness. (F) Number of chloroplasts. (G) Granum height. (H) Thylakoid layer-granum. Dark blue to light red indicates increased associations. (n = 200); Figure S3: Transmittance by count plot map of the coefficient of correlation (R^2). (A) Adaxial epidermis. (B) Adaxial hypodermis. (C) Parenchyma thickness. (D) Adaxial hypodermis. (E) Total leaf thickness. (F) Number of chloroplasts. (G) Granum height. (H) Thylakoid layer-granum. Dark blue to light red indicates increased associations. (n = 200); Figure S4: Absorbance count plot map of the coefficient of correlation (R^2). (A) Adaxial epidermis. (B) Adaxial hypodermis. (C) Parenchyma thickness. (D) Adaxial hypodermis. (E) Total leaf thick-ness. (F) Number of chloroplasts. (G) Granum height. (H) Thylakoid layer-granum. Dark blue to light red indicates increased associations. (n = 200); Figure S5: Observed vs predicted data estimated by partial least squares regression (PLSR) with hyperspectral ChlF data. (n = 50); Figure S6: The observed vs. predicted data were estimated using partial least squares regression (PLSR) with hyperspectral reflectance data. (n = 50); Figure S7: The measured data were compared to the data calculated via partial least squares regression (PLSR) with hyperspectral transmittance data. (n = 50); Figure S8: Observed vs. predicted data estimated by partial least squares regression (PLSR) of hyperspectral absorbance data. (n = 50); Table S1: The performance of partial least squares regression (PLSR) models that use various spectroscopic techniques to estimate biophysical compound concentrations in *Tradescantia* species. The evaluation metrics included the correlation coefficient (r), R-squared (R^2), slope, offset, standard error of prediction (SEP), root mean squared error of prediction (RMSEP), ratio of performance to deviation (RPD), and bias, offering a detailed analysis for each sensor. Statistical performance of the PLSR model in predicting phases. Statistical metrics from the PLSR model in the calibration and cross-validation phases. (n = 150).

Author Contributions: Conceptualization, R.F., J.A.M.D. and M.R.N.; Data curation, R.F., W.C.A., J.A.M.D. and M.R.N.; Formal analysis, R.F., J.A.M.D. and M.R.N.; Funding acquisition, R.F., J.A.M.D. and M.R.N.; Investigation, R.F., J.A.M.D. and M.R.N.; Methodology, R.F., R.B.d.O., M.L.C., J.A.M.D. and M.R.N.; Project administration, R.F., J.A.M.D. and M.R.N.; Resources, R.F., W.C.A., J.A.M.D.

and M.R.N.; Software, R.F., M.L.C., W.C.A., J.A.M.D. and M.R.N.; Supervision, R.F., J.A.M.D. and M.R.N.; Validation, R.F., R.B.d.O., J.A.M.D. and M.R.N.; Visualization, R.F., J.A.M.D. and M.R.N.; Writing—original draft, R.F., R.B.d.O., M.L.C., W.C.A., J.A.M.D. and M.R.N.; Writing—review and editing, R.F., M.L.C., W.C.A., J.A.M.D. and M.R.N. All authors have read and agreed to the published version of the manuscript.

Funding: National Council for Scientific and Technological Development: Programa de Apoio à Fixação de Jovens Doutores no Brasil 168180/2022−7; Fundação Araucária: CP 19/2022—Jovens Doutores; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior: 001; Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP pn 2021/05129−8 for sensor financial support. CEAGRE—Centro de Excelência em Agricultura Exponencial for financial support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Material, and further inquiries can be directed to the corresponding author.

Acknowledgments: We thank Programa de Pós-Graduação em Agronomia (PGA-UEM) at the State University of Maringá, COMCAP-UEM and FAPESP for their support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the study design, collection, analyses, interpretation of data, writing of the manuscript, or decision to publish the results.

References

1. da Silva Junior, C.A.; Nanni, M.R.; Shakir, M.; Teodoro, P.E.; de Oliveira-Júnior, J.F.; Cezar, E.; de Gois, G.; Lima, M.; Wojciechowski, J.C.; Shiratsuchi, L.S. Soybean Varieties Discrimination Using Non-imaging Hyperspectral Sensor. *Infrared Phys. Technol.* **2018**, *89*, 338–350. [CrossRef]
2. Wang, D.; Cao, W.; Zhang, F.; Li, Z.; Xu, S.; Wu, X. A Review of Deep Learning in Multiscale Agricultural Sensing. *Remote Sens.* **2022**, *14*, 559. [CrossRef]
3. Féret, J.-B.; le Maire, G.; Jay, S.; Berveiller, D.; Bendoula, R.; Hmimina, G.; Cheraiet, A.; Oliveira, J.C.; Ponzoni, F.J.; Solanki, T.; et al. Estimating Leaf Mass per Area and Equivalent Water Thickness Based on Leaf Optical Properties: Potential and Limitations of Physical Modelling and Machine Learning. *Remote Sens. Environ.* **2019**, *231*, 110959. [CrossRef]
4. Gupta, R. Tissue Specific Disruption of Photosynthetic Electron Transport Rate in Pigeonpea (*Cajanus cajan* L.) under Elevated Temperature. *Plant Signal. Behav.* **2019**, *14*, 1601952. [CrossRef] [PubMed]
5. Colpo, A.; Molinari, A.; Boldrini, P.; Živčak, M.; Brestič, M.; Demaria, S.; Baldisserotto, C.; Pancaldi, S.; Ferroni, L. Thylakoid Membrane Appression in the Giant Chloroplast of *Selaginella martensii* Spring: A Lycophyte Challenges Grana Paradigms in Shade-Adapted Species. *Plant Sci.* **2023**, *336*, 111833. [CrossRef]
6. Stirbet, A. Govindjee On the Relation between the Kautsky Effect (Chlorophyll a Fluorescence Induction) and Photosystem II: Basics and Applications of the OJIP Fluorescence Transient. *J. Photochem. Photobiol. B Biol.* **2011**, *104*, 236–257. [CrossRef]
7. Kalaji, H.M.; Jajoo, A.; Oukarroum, A.; Brestic, M.; Zivcak, M.; Samborska, I.A.; Cetner, M.D.; Łukasik, I.; Goltsev, V.; Ladle, R.J.; et al. *The Use of Chlorophyll Fluorescence Kinetics Analysis to Study the Performance of Photosynthetic Machinery in Plants*; Academic Press: Cambridge, MA, USA, 2014; Volume 2.
8. Cheng, T.; Song, R.; Li, D.; Zhou, K.; Zheng, H.; Yao, X.; Tian, Y.; Cao, W.; Zhu, Y. Spectroscopic Estimation of Biomass in Canopy Components of Paddy Rice Using Dry Matter and Chlorophyll Indices. *Remote Sens.* **2017**, *9*, 319. [CrossRef]
9. Bégué, A.; Arvor, D.; Bellon, B.; Betbeder, J.; de Abelleyra, D.; Ferraz, R.P.D.; Lebourgeois, V.; Lelong, C.; Simões, M.; Verón, S.R. Remote Sensing and Cropping Practices: A Review. *Remote Sens.* **2018**, *10*, 99. [CrossRef]
10. Falcioni, R.; Moriwaki, T.; Pattaro, M.; Herrig Furlanetto, R.; Nanni, M.R.; Camargos Antunes, W. High Resolution Leaf Spectral Signature as a Tool for Foliar Pigment Estimation Displaying Potential for Species Differentiation. *J. Plant Physiol.* **2020**, *249*, 153161. [CrossRef]
11. Falcioni, R.; Oliveira, R.B.d.; Chicati, M.L.; Antunes, W.C.; Demattê, J.A.M.; Nanni, M.R. Estimation of Biochemical Compounds in Tradescantia Leaves Using VIS-NIR-SWIR Hyperspectral and Chlorophyll a Fluorescence Sensors. *Remote Sens.* **2024**, *16*, 1910. [CrossRef]
12. Shurygin, B.; Chivkunova, O.; Solovchenko, O.; Solovchenko, A.; Dorokhov, A.; Smirnov, I.; Astashev, M.E.; Khort, D. Comparison of the Noninvasive Monitoring of Fresh-Cut Lettuce Condition with Imaging Reflectance Hyperspectrometer and Imaging PAM-Fluorimeter. *Photonics* **2021**, *8*, 425. [CrossRef]
13. Shorten, P.R.; Leath, S.R.; Schmidt, J.; Ghamkhar, K. Predicting the Quality of Ryegrass Using Hyperspectral Imaging. *Plant Methods* **2019**, *15*, 63. [CrossRef] [PubMed]

14. Alimohammadi, F.; Rasekh, M.; Afkari Sayyah, A.H.; Abbaspour-Gilandeh, Y.; Karami, H.; Rasooli Sharabiani, V.; Fioravanti, A.; Gancarz, M.; Findura, P.; Kwaśniewski, D. Hyperspectral Imaging Coupled with Multivariate Analysis and Artificial Intelligence to the Classification of Maize Kernels. *Int. Agrophys.* **2022**, *36*, 83–91. [CrossRef] [PubMed]
15. Sakamoto, W.; Uno, Y.; Zhang, Q.; Miura, E.; Kato, Y. Sodmergen Arrested Differentiation of Proplastids into Chloroplasts in Variegated Leaves Characterized by Plastid Ultrastructure and Nucleoid Morphology. *Plant Cell Physiol.* **2009**, *50*, 2069–2083. [CrossRef] [PubMed]
16. Falcioni, R.; Gonçalves, J.V.F.; Oliveira, K.M.d.; Antunes, W.C.; Nanni, M.R. VIS-NIR-SWIR Hyperspectroscopy Combined with Data Mining and Machine Learning for Classification of Predicted Chemometrics of Green Lettuce. *Remote Sens.* **2022**, *14*, 6330. [CrossRef]
17. Gitelson, A.; Solovchenko, A. Noninvasive Quantification of Foliar Pigments: Possibilities and Limitations of Reflectance- and Absorbance-Based Approaches. *J. Photochem. Photobiol. B Biol.* **2018**, *178*, 537–544. [CrossRef]
18. Sobejano-Paz, V.; Mikkelsen, T.N.; Baum, A.; Mo, X.; Liu, S.; Köppl, C.J.; Johnson, M.S.; Gulyas, L.; García, M. Hyperspectral and Thermal Sensing of Stomatal Conductance, Transpiration, and Photosynthesis for Soybean and Maize under Drought. *Remote Sens.* **2020**, *12*, 3182. [CrossRef]
19. Jacquemoud, S.; Ustin, S. *Applications of Leaf Optics*, 1st ed.; Cambridge University Press: Davis, CA, USA, 2019.
20. Carter, G.A.; Knapp, A.K. Leaf Optical Properties in Higher Plants: Linking Spectral Characteristics to Stress and Chlorophyll Concentration. *Am. J. Bot.* **2001**, *88*, 677–684. [CrossRef]
21. Li, K.-Y.; de Lima, R.; Burnside, N.G.; Vahtmäe, E.; Kutser, T.; Sepp, K.; Cabral Pinheiro, V.H.; Yang, M.-D.; Vain, A.; Sepp, K. Toward Automated Machine Learning-Based Hyperspectral Image Analysis in Crop Yield and Biomass Estimation. *Remote Sens.* **2022**, *14*, 1114. [CrossRef]
22. Sexton, T.; Sankaran, S.; Cousins, A.B. Predicting Photosynthetic Capacity in Tobacco Using Shortwave Infrared Spectral Reflectance. *J. Exp. Bot.* **2021**, *72*, 4373–4383. [CrossRef]
23. Zhong, S.; Shi, H.; Xue, C.; Wang, L.; Xi, Y.; Li, J.; Quail, P.H.; Deng, X.W.; Guo, H. A Molecular Framework of Light-Controlled Phytohormone Action in Arabidopsis. *Curr. Biol.* **2012**, *22*, 1530–1535. [CrossRef] [PubMed]
24. Hassanzadeh, A.; Murphy, S.P.; Pethybridge, S.J.; van Aardt, J. Growth Stage Classification and Harvest Scheduling of Snap Bean Using Hyperspectral Sensing: A Greenhouse Study. *Remote Sens.* **2020**, *12*, 3809. [CrossRef]
25. Falcioni, R.; Santos, G.L.A.A.d.; Crusiol, L.G.T.; Antunes, W.C.; Chicati, M.L.; Oliveira, R.B.d.; Demattê, J.A.M.; Nanni, M.R. Non-invasive Assessment, Classification, and Prediction of Biophysical Parameters Using Reflectance Hyperspectroscopy. *Plants* **2023**, *12*, 2526. [CrossRef] [PubMed]
26. Ling, B.; Goodin, D.G.; Raynor, E.J.; Joern, A. Hyperspectral Analysis of Leaf Pigments and Nutritional Elements in Tallgrass Prairie Vegetation. *Front. Plant Sci.* **2019**, *10*, 142. [CrossRef] [PubMed]
27. Féret, J.B.; Gitelson, A.A.; Noble, S.D.; Jacquemoud, S. PROSPECT-D: Toward Modelling Leaf Optical Properties through a Complete Lifecycle. *Remote Sens. Environ.* **2017**, *193*, 204–215. [CrossRef]
28. Gitelson, A.; Solovchenko, A.; Viña, A. Foliar Absorption Coefficient Derived from Reflectance Spectra: A Gauge of the Efficiency of in Situ Light-Capture by Different Pigment Groups. *J. Plant Physiol.* **2020**, *254*, 153277. [CrossRef]
29. Falcioni, R.; Antunes, W.C.; Oliveira, R.B.d.; Chicati, M.L.; Demattê, J.A.M.; Nanni, M.R. Hyperspectral and Chlorophyll Fluorescence Analyses of Comparative Leaf Surfaces Reveal Cellular Influences on Leaf Optical Properties in Tradescantia Plants. *Cells* **2024**, *13*, 952. [CrossRef]
30. Falcioni, R.; Gonçalves, J.V.F.; Oliveira, K.M.d.; Oliveira, C.A.d.; Demattê, J.A.M.; Antunes, W.C.; Nanni, M.R. Enhancing Pigment Phenotyping and Classification in Lettuce through the Integration of Reflectance Spectroscopy and AI Algorithms. *Plants* **2023**, *12*, 1333. [CrossRef]
31. Huang, C.-Y.; Asner, G.P. Applications of Remote Sensing to Alien Invasive Plant Studies. *Sensors* **2009**, *9*, 4869–4889. [CrossRef]
32. Bloem, E.; Gerighausen, H.; Chen, X.; Schnug, E. The Potential of Spectral Measurements for Identifying Glyphosate Application to Agricultural Fields. *Agronomy* **2020**, *10*, 1409. [CrossRef]
33. Boshkovski, B.; Doupis, G.; Zapolska, A.; Kalaitzidis, C.; Koubouris, G. Hyperspectral Imagery Detects Water Deficit and Salinity Effects on Photosynthesis and Antioxidant Enzyme Activity of Three Greek Olive Varieties. *Sustainability* **2022**, *14*, 1432. [CrossRef]
34. Yendrek, C.R.; Tomaz, T.; Montes, C.M.; Cao, Y.; Morse, A.M.; Brown, P.J.; McIntyre, L.M.; Leakey, A.D.B.; Ainsworth, E.A. High-Throughput Phenotyping of Maize Leaf Physiological and Biochemical Traits Using Hyperspectral Reflectance. *Plant Physiol.* **2017**, *173*, 614–626. [CrossRef]
35. Bauriegel, E.; Giebel, A.; Herppich, W.B. Hyperspectral and Chlorophyll Fluorescence Imaging to Analyse the Impact of *Fusarium culmorum* on the Photosynthetic Integrity of Infected Wheat Ears. *Sensors* **2011**, *11*, 3765–3779. [CrossRef]
36. Ge, Y.; Atefi, A.; Zhang, H.; Miao, C.; Ramamurthy, R.K.; Sigmon, B.; Yang, J.; Schnable, J.C. High-Throughput Analysis of Leaf Physiological and Chemical Traits with VIS–NIR–SWIR Spectroscopy: A Case Study with a Maize Diversity Panel. *Plant Methods* **2019**, *15*, 66. [CrossRef] [PubMed]
37. Crusiol, L.G.T.; Sun, L.; Sibaldelli, R.N.R.; Junior, V.F.; Furlaneti, W.X.; Chen, R.; Sun, Z.; Wuyun, D.; Chen, Z.; Nanni, M.R.; et al. Strategies for Monitoring Within-Field Soybean Yield Using Sentinel-2 Vis-NIR-SWIR Spectral Bands and Machine Learning Regression Methods. *Precis. Agric.* **2022**, *23*, 1093–1123. [CrossRef]

38. Ferri, C.P.; Formaggio, A.R.; Schiavinato, M.A. Narrow Band Spectral Indexes for Chlorophyll Determination in Soybean Canopies [*Glycine max* (L.) Merril]. *Braz. J. Plant Physiol.* **2004**, *16*, 131–136. [CrossRef]
39. Zheng, W.; Lu, X.; Li, Y.; Li, S.; Zhang, Y. Hyperspectral Identification of Chlorophyll Fluorescence Parameters of *Suaeda salsa* in Coastal Wetlands. *Remote Sens.* **2021**, *13*, 2066. [CrossRef]
40. Fernandes, A.M.; Fortini, E.A.; Müller, L.A.d.C.; Batista, D.S.; Vieira, L.M.; Silva, P.O.; Amaral, C.H.d.; Poethig, R.S.; Otoni, W.C. Leaf Development Stages and Ontogenetic Changes in Passionfruit (*Passiflora edulis* Sims.) Are Detected by Narrowband Spectral Signal. *J. Photochem. Photobiol. B Biol.* **2020**, *209*, 111931. [CrossRef]
41. Gururani, M.A.; Venkatesh, J.; Ghosh, R.; Strasser, R.J.; Ponpandian, L.N.; Bae, H. Chlorophyll-a Fluorescence Evaluation of PEG-Induced Osmotic Stress on PSII Activity in *Arabidopsis* Plants Expressing SIP1. *Plant Biosyst.* **2018**, *152*, 945–952. [CrossRef]
42. Crusiol, L.G.T.; Sun, L.; Sun, Z.; Chen, R.; Wu, Y.; Ma, J.; Song, C. In-Season Monitoring of Maize Leaf Water Content Using Ground-Based and UAV-Based Hyperspectral Data. *Sustainability* **2022**, *14*, 9039. [CrossRef]
43. Crusiol, L.G.T.; Nanni, M.R.; Furlanetto, R.H.; Sibaldelli, R.N.R.; Sun, L.; Gonçalves, S.L.; Foloni, J.S.S.; Mertz-Henning, L.M.; Nepomuceno, A.L.; Neumaier, N.; et al. Assessing the Sensitive Spectral Bands for Soybean Water Status Monitoring and Soil Moisture Prediction Using Leaf-Based Hyperspectral Reflectance. *Agric. Water Manag.* **2023**, *277*, 108089. [CrossRef]
44. Koh, J.C.O.; Banerjee, B.P.; Spangenberg, G.; Kant, S. Automated Hyperspectral Vegetation Index Derivation Using a Hyperparameter Optimisation Framework for High-Throughput Plant Phenotyping. *New Phytol.* **2022**, *233*, 2659–2670. [CrossRef]
45. Falcioni, R.; Antunes, W.C.; Demattê, J.A.M.; Nanni, M.R. Biophysical, Biochemical, and Photochemical Analyses Using Reflectance Hyperspectroscopy and Chlorophyll a Fluorescence Kinetics in Variegated Leaves. *Biology* **2023**, *12*, 704. [CrossRef] [PubMed]
46. Gitelson, A.; Chivkunova, O.; Zhigalova, T.; Solovchenko, A. In Situ Optical Properties of Foliar Flavonoids: Implication for Nondestructive Estimation of Flavonoid Content. *J. Plant Physiol.* **2017**, *218*, 258–264. [CrossRef]
47. Chen, S.; Zhang, F.; Ning, J.; Liu, X.; Zhang, Z.; Yang, S. Predicting the Anthocyanin Content of Wine Grapes by NIR Hyperspectral Imaging. *Food Chem.* **2015**, *172*, 788–793. [CrossRef] [PubMed]
48. Kumar, A.; Kashyap, Y.; Kosmopoulos, P. Enhancing Solar Energy Forecast Using Multi-Column Convolutional Neural Network and Multipoint Time Series Approach. *Remote Sens.* **2023**, *15*, 107. [CrossRef]
49. Pierna, J.A.F.; Baeten, V.; Renier, A.M.; Cogdill, R.P.; Dardenne, P. Combination of Support Vector Machines (SVM) and near-Infrared (NIR) Imaging Spectroscopy for the Detection of Meat and Bone Meal (MBM) in Compound Feeds. *J. Chemom.* **2004**, *18*, 341–349. [CrossRef]
50. Nalepa, J. Recent Advances in Multi- and Hyperspectral Image Analysis. *Sensors* **2021**, *21*, 6002. [CrossRef]
51. Braga, P.; Crusiol, L.G.T.; Nanni, M.R.; Caranhato, A.L.H.; Fuhrmann, M.B.; Nepomuceno, A.L.; Neumaier, N.; Farias, J.R.B.; Koltun, A.; Gonçalves, L.S.A.; et al. Vegetation Indices and NIR-SWIR Spectral Bands as a Phenotyping Tool for Water Status Determination in Soybean. *Precis. Agric.* **2021**, *22*, 249–266. [CrossRef]
52. Cezar, E.; Nanni, M.R.; Crusiol, L.G.T.; Sun, L.; Chicati, M.S.; Furlanetto, R.H.; Rodrigues, M.; Sibaldelli, R.N.R.; Silva, G.F.C.; de Oliveira, K.M.; et al. Strategies for the Development of Spectral Models for Soil Organic Matter Estimation. *Remote Sens.* **2021**, *13*, 1376. [CrossRef]
53. de Oliveira, K.M.; Falcioni, R.; Gonçalves, J.V.F.; de Oliveira, C.A.; Mendonça, W.A.; Crusiol, L.G.T.; de Oliveira, R.B.; Furlanetto, R.H.; Reis, A.S.; Nanni, M.R. Rapid Determination of Soil Horizons and Suborders Based on VIS-NIR-SWIR Spectroscopy and Machine Learning Models. *Remote Sens.* **2023**, *15*, 4859. [CrossRef]
54. Franca, T.; Goncalves, D.; Cena, C. ATR-FTIR Spectroscopy Combined With Machine Learning For Classification of PVA/PVP Blends in Low Concentration. *Vib. Spectrosc.* **2022**, *120*, 103378. [CrossRef]
55. Falcioni, R.; Antunes, W.C.; Demattê, J.A.M.; Nanni, M.R. Reflectance Spectroscopy for the Classification and Prediction of Pigments in Agronomic Crops. *Plants* **2023**, *12*, 2347. [CrossRef] [PubMed]
56. de Oliveira, K.M.; Ferreira Gonçalves, J.V.; Falcioni, R.; Almeida de Oliveira, C.; de Fatima da Silva Haubert, D.; Mendonça, W.A.; Teixeira Crusiol, L.G.; Berti de Oliveira, R.; Reis, A.S.; Cezar, E.; et al. Classification of Soil Horizons Based on VisNIR and SWIR Hyperespectral Images and Machine Learning Models. *Remote Sens. Appl. Soc. Environ.* **2024**, *36*, 101362. [CrossRef]
57. Cui, X.; Zhao, Z.; Zhang, G.; Chen, S.; Zhao, Y.; Lu, J. Analysis and Classification of Kidney Stones Based on Raman Spectroscopy. *Biomed. Opt. Express* **2018**, *9*, 4175–4183. [CrossRef]
58. Minasny, B.; McBratney, A.B.; Malone, B.P.; Wheeler, I. Digital Mapping of Soil Carbon. In *Advances in Agronomy*; Elsevier: Amsterdam, The Netherlands, 2013; Volume 3, pp. 1–47.
59. Falcioni, R.; Gonçalves, J.V.F.; de Oliveira, K.M.; de Oliveira, C.A.; Reis, A.S.; Crusiol, L.G.T.; Furlanetto, R.H.; Antunes, W.C.; Cezar, E.; de Oliveira, R.B.; et al. Chemometric Analysis for the Prediction of Biochemical Compounds in Leaves Using UV-VIS-NIR-SWIR Hyperspectroscopy. *Plants* **2023**, *12*, 3424. [CrossRef]
60. Baio, F.H.R.; Santana, D.C.; Teodoro, L.P.R.; de Oliveira, I.C.; Gava, R.; de Oliveira, J.L.G.; Silva Junior, C.A.d.; Teodoro, P.E.; Shiratsuchi, L.S. Maize Yield Prediction with Machine Learning, Spectral Variables and Irrigation Management. *Remote Sens.* **2023**, *15*, 79. [CrossRef]
61. Jin, J.; Huang, N.; Huang, Y.; Yan, Y.; Zhao, X.; Wu, M. Proximal Remote Sensing-Based Vegetation Indices for Monitoring Mango Tree Stem Sap Flux Density. *Remote Sens.* **2022**, *14*, 1483. [CrossRef]
62. Falcioni, R.; Moriwaki, T.; Bonato, C.M.; de Souza, L.A.; Nanni, M.R.; Antunes, W.C. Distinct Growth Light and Gibberellin Regimes Alter Leaf Anatomy and Reveal Their Influence on Leaf Optical Properties. *Environ. Exp. Bot.* **2017**, *140*, 86–95. [CrossRef]

63. Fu, Y.; Yang, G.; Song, X.; Li, Z.; Xu, X.; Feng, H.; Zhao, C. Improved Estimation of Winter Wheat Aboveground Biomass Using Multiscale Textures Extracted from UAV-Based Digital Images and Hyperspectral Feature Analysis. *Remote Sens.* **2021**, *13*, 581. [CrossRef]
64. Ge, Y.; Bai, G.; Stoerger, V.; Schnable, J.C. Temporal Dynamics of Maize Plant Growth, Water Use, and Leaf Water Content Using Automated High Throughput RGB and Hyperspectral Imaging. *Comput. Electron. Agric.* **2016**, *127*, 625–632. [CrossRef]
65. Kycko, M.; Zagajewski, B.; Lavender, S.; Dabija, A. In Situ Hyperspectral Remote Sensing for Monitoring of Alpine Trampled and Recultivated Species. *Remote Sens.* **2019**, *11*, 1296. [CrossRef]
66. Hu, Y.; Wang, Z.; Li, X.; Li, L.; Wang, X.; Wei, Y. Nondestructive Classification of Maize Moldy Seeds by Hyperspectral Imaging and Optimal Machine Learning Algorithms. *Sensors* **2022**, *22*, 6064. [CrossRef] [PubMed]
67. Iqbal, I.M.; Balzter, H.; Firdaus-e-Bareen; Shabbir, A. Identifying the Spectral Signatures of Invasive and Native Plant Species in Two Protected Areas of Pakistan through Field Spectroscopy. *Remote Sens.* **2021**, *13*, 4009. [CrossRef]
68. Jin, J.; Arief Pratama, B.; Wang, Q. Tracing Leaf Photosynthetic Parameters Using Hyperspectral Indices in an Alpine Deciduous Forest. *Remote Sens.* **2020**, *12*, 1124. [CrossRef]
69. Fine, P.V.A.; Salazar, D.; Martin, R.E.; Metz, M.R.; Misiewicz, T.M.; Asner, G.P. Exploring the Links between Secondary Metabolites and Leaf Spectral Reflectance in a Diverse Genus of Amazonian Trees. *Ecosphere* **2021**, *12*, e03362. [CrossRef]
70. Nogales-Bueno, J.; Baca-Bocanegra, B.; Rooney, A.; Miguel Hernández-Hierro, J.; José Heredia, F.; Byrne, H.J. Linking ATR-FTIR and Raman Features to Phenolic Extractability and Other Attributes in Grape Skin. *Talanta* **2017**, *167*, 44–50. [CrossRef]
71. Davis, P.A.; Caylor, S.; Whippo, C.W.; Hangarter, R.P. Changes in Leaf Optical Properties Associated with Light-Dependent Chloroplast Movements. *Plant Cell Environ.* **2011**, *34*, 2047–2059. [CrossRef]
72. Tsimilli-Michael, M.; Strasser, R.J. In Vivo Assessment of Stress Impact on Plant's Vitality: Applications in Detecting and Evaluating the Beneficial Role of Mycorrhization on Host Plants. In *Mycorrhiza*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 679–703.
73. Coast, O.; Shah, S.; Ivakov, A.; Gaju, O.; Wilson, P.B.; Posch, B.C.; Bryant, C.J.; Negrini, A.C.A.; Evans, J.R.; Condon, A.G.; et al. Predicting Dark Respiration Rates of Wheat Leaves from Hyperspectral Reflectance. *Plant Cell Environ.* **2019**, *42*, 2133–2150. [CrossRef]
74. de Oliveira, M.R.R.; Ribeiro, S.G.; Mas, J.F.; dos Santos Teixeira, A. Advances in Hyperspectral Sensing in Agriculture: A Review. *Rev. Cienc. Agron.* **2020**, *51*, e20207739. [CrossRef]
75. Brezini, S.E.; Deville, Y. Hyperspectral and Multispectral Image Fusion with Automated Extraction of Image-Based Endmember Bundles and Sparsity-Based Unmixing to Deal with Spectral Variability. *Sensors* **2023**, *23*, 2341. [CrossRef] [PubMed]
76. Steidle Neto, A.J.; Moura, L.d.O.; Lopes, D.d.C.; Carlos, L.d.A.; Martins, L.M.; Ferraz, L.d.C.L. Nondestructive Prediction of Pigment Content in Lettuce Based on Visible-NIR Spectroscopy. *J. Sci. Food Agric.* **2017**, *97*, 2015–2022. [CrossRef] [PubMed]
77. Zhou, Q.; Yu, L.; Zhang, X.; Liu, Y.; Zhan, Z.; Ren, L.; Luo, Y. Fusion of UAV Hyperspectral Imaging and LiDAR for the Early Detection of EAB Stress in Ash and a New EAB Detection Index NDVI(776,678). *Remote Sens.* **2022**, *14*, 2428. [CrossRef]
78. Feng, W.; Yao, X.; Tian, Y.; Cao, W.; Zhu, Y. Monitoring Leaf Pigment Status with Hyperspectral Remote Sensing in Wheat. *Aust. J. Agric. Res.* **2008**, *59*, 748–760. [CrossRef]
79. Fan, K.; Li, F.; Chen, X.; Li, Z.; Mulla, D.J. Nitrogen Balance Index Prediction of Winter Wheat by Canopy Hyperspectral Transformation and Machine Learning. *Remote Sens.* **2022**, *14*, 3504. [CrossRef]
80. Chen, J.; de Hoogh, K.; Gulliver, J.; Hoffmann, B.; Hertel, O.; Ketzel, M.; Bauwelinck, M.; van Donkelaar, A.; Hvidtfeldt, U.A.; Katsouyanni, K.; et al. A Comparison of Linear Regression, Regularization, and Machine Learning Algorithms to Develop Europe-Wide Spatial Models of Fine Particles and Nitrogen Dioxide. *Environ. Int.* **2019**, *130*, 104934. [CrossRef]
81. Lebourgeois, V.; Bégué, A.; Labbé, S.; Mallavan, B. Can Commercial Digital Cameras Be Used as Multispectral. *Sensors* **2008**, *8*, 7300–7322. [CrossRef]
82. Cuba, N.I.; Torres, R.; San Román, E.; Lagorio, M.G. Influence of Surface Structure, Pigmentation and Particulate Matter on Plant Reflectance and Fluorescence. *Photochem. Photobiol.* **2021**, *97*, 110–121. [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.

sensors

MDPI

Article

Development of an Earth-Field Nuclear Magnetic Resonance Spectrometer: Paving the Way for AI-Enhanced Low-Field Nuclear Magnetic Resonance Technology

Eduardo Viciana [1], Juan Antonio Martínez-Lao [1], Emilio López-Lao [1], Ignacio Fernández [2,*] and Francisco Manuel Arrabal-Campos [1,2,*]

[1] Department of Engineering, Research Centre CIMEDES, Escuela Superior de Ingeniería, University of Almería, 04120 Almeria, Spain; evg010@ual.es (E.V.); ell658@inlumine.ual.es (E.L.-L.)
[2] Department of Chemistry and Physics, Research Centre CIAIMBITAL, University of Almería, 04120 Almeria, Spain
* Correspondence: ifernan@ual.es (I.F.); fmarrabal@ual.es (F.M.A.-C.)

Abstract: Today, it is difficult to have a high-field nuclear magnetic resonance (NMR) device due to the high cost of its acquisition and maintenance. These high-end machines require significant space and specialist personnel for operation and offer exceptional quality in the acquisition, processing, and other advanced functions associated with detected signals. However, alternative devices are low-field nuclear magnetic resonance devices. They benefit from the elimination of high-tech components that generate static magnetic fields and advanced instruments. Instead, they used magnetic fields induced by ordinary conductors. Another category of spectrometers uses the Earth's magnetic field, which is simple and economical but limited in use. These devices are called Earth-Field Nuclear Magnetic Resonance (EFNMR) devices. This device is ideal for educational purposes, especially for engineers and those who study nuclear magnetic resonance, such as chemistry or other experimental sciences. Students can observe their internal workings and conduct experiments that complement their education without worrying about damaging equipment. This article provides a detailed explanation of the design and construction of electrical technology devices for the excitation of atomic spin resonance using Earth's magnetic fields. It covers all necessary stages, from research to analysis, including simulation, assembly, construction of each component, and the development of comprehensive software for spectrometer control.

Keywords: Earth's field NMR; NMR spectrometer; EFNMR

1. Introduction

The history of nuclear magnetic resonance (NMR) is deeply intertwined with advances in multiple scientific disciplines, leading to its application in various fields such as chemistry, biochemistry, and medicine [1]. The initial breakthroughs in NMR can be traced back to the quantization of nuclear spin angular momentum observed by Stern and Gerlach in 1936 and the magnetic resonance method developed by Rabi and his team, which allowed for the precise measurements of magnetic moments [2,3]. Despite early challenges, such as Gorter's unsuccessful attempts to detect NMR in solids due to long relaxation times, these foundational experiments set the stage for further developments [1].

From the 1940s through the 1950s, significant strides in NMR technology were made. Bloch and Purcell, working independently, were the first to detect NMR signals in bulk matter, a discovery that earned them the Nobel Prize in Physics in 1952 [4,5]. Their work revealed much shorter relaxation times than previously anticipated, marking a pivotal moment in the understanding and application of NMR [6]. During this period, innovations such as the use of short radiofrequency pulses to excite nuclear spins and the discovery of spin echo by Hahn further refined NMR techniques, particularly in measuring the T_1

and T_2 relaxation times [7–9]. These advances were crucial in developing a new theoretical framework for NMR, including Bloch's introduction of the relaxation times and the BPP theory, which explained the effects of molecular motion on NMR signals [10,11].

The application of NMR has expanded rapidly with the discovery of chemical shifts and scalar coupling. Early observations, such as the two distinct resonance signals in ^{14}N with NH_4NO_3 and the identification of chemical shifts in various compounds, laid the foundation for using NMR in the determination of chemical structures [12–15]. The introduction of the term "chemical shift" by Herbert S. Gutowsky and the correlation of shifts with structural features significantly impacted the utility of NMR in chemistry [16]. High-resolution NMR techniques revealed the multiplicity of chemically shifted signals, further enhancing their application in structural analysis [17,18].

Dynamic nuclear polarization (DNP), discovered by Overhauser in 1953, and its subsequent verification and application in the nuclear Overhauser effect (NOE) by Solomon, provided new tools to enhance signal intensity in NMR, although these were not widely recognized until later [19–21]. As NMR technology matured, commercial spectrometers, such as those developed by Varian Associates, played a crucial role in making NMR more accessible to the scientific community [22,23]. Advances such as the Fourier transform NMR (FT-NMR) by Anderson and Ernst in the 1960s revolutionized data acquisition, allowing for faster and more efficient analysis, which was recognized with a Nobel Prize in Chemistry in 1991 [24].

NMR technology continued to advance through the 1970s and 1980s, with the development of two-dimensional (2D) NMR, proposed by Jean Jeener and advanced by Richard Ernst [25,26]. These techniques allowed for more detailed structural analyses, particularly in the study of biopolymers and complex molecules [27–29]. The integration of computational methods and improvements in hardware and software further enhanced the capabilities of NMR, making it a powerful tool for studying dynamic processes and the structure of biological substances [30].

In the field of medicine, the application of NMR took a significant leap with the introduction of magnetic resonance imaging (MRI) in the 1970s. The work of Lauterbur on spatial encoding and Mansfield on echoplanar imaging (EPI) led to the development of non-invasive diagnostic tools that revolutionized clinical medicine [31,32]. Solid-state NMR (ssNMR) also advanced, enabling the study of complex solid materials such as biological membranes and polymers [33,34].

The first Earth's field NMR (EFNMR) was performed in 1954 [35,36]; it was a thing of curiosity, and the main use was for geophysical magnetometry [37] or simply a low-cost demonstration of magnetic resonance principles. Despite its low field, it continues to attract the attention of researchers and engineers not only because of its rarity as a physical phenomenon but also because of certain low-field results that are difficult to obtain in a high field, such as the inflection obtained in the dependence of the inverse of T1 below 0.01 T of the signal detected in some tissues [38], as well as other applications encountered [39,40]. The strength of the Earth's magnetic field is very low. It is still possible to obtain narrow NMR lines and a high signal-to-noise ratio with an acceptable spectral resolution. In the 1960s, Béné and colleagues demonstrated that EFNMR could measure the *J*-coupling of nuclear elements [41]. Because high homogeneity and stability can be achieved successfully in low fields, experiments have proven that many important spin multipliers can be studied in low fields. Chemical shift differences have recently been demonstrated to be measured even in inhomogeneous low magnetic fields [42].

This device presents several potential applications. One significant application is the precise measurement of the chemical shift of hyperpolarized xenon-129 in the Earth's magnetic field, achieving a precision similar to high-field measurements [43]. Furthermore, ^{1} H and ^{19}F EFNMR spectra can be acquired with frequency resolutions nearly two orders of magnitude better than those obtainable with high-field high-resolution NMR [44]. Another application of the breakthrough was the observation of heteronuclear J-couplings using the classic COSY experiment [45], and more recently, the first observation of homonuclear

proton J-couplings in the Earth's magnetic field, previously considered impossible [46]. In the realm of condensed matter physics, low-field NMR allows for high-resolution spectra in solids by minimizing the line broadening caused by dipole–dipole interactions in high fields [47]. Moreover, low magnetic fields allow for the observation of "forbidden transitions", providing unique information that is not detectable at high magnetic fields Kohl et al. [48]. These applications highlight the versatility and potential of NMR in various scientific fields.

2. Motivation and Objectives

NMR has proven to be an exceptionally valuable tool in a variety of disciplines, from diagnostic medicine to chemical and biological research. However, most current NMR spectrometers require intense magnetic fields, which limits their use to specialized laboratory environments. The development of an NMR spectrometer that can effectively operate using the Earth's magnetic field represents a significant pedagogical shift. This advancement could revolutionize the educational application of NMR, enabling its integration into specialized NMR courses within the curriculum.

Current NMR equipment are prohibitively expensive for many institutions, especially those in developing countries and organizations with limited budgets. Using the Earth's magnetic field, a more accessible NMR spectrometer can significantly reduce costs and lower barriers to access, allowing for a much broader range of researchers to use this valuable tool. However, its use is currently limited to educational purposes.

The potential applications of an Earth-field NMR spectrometer extend beyond traditional laboratories. There are numerous situations where access to conventional NMR equipment is challenging or impossible, such as field research, low-resource environments, or remote geographic regions. An Earth-field NMR spectrometer could enable on-site analyses, provide real-time data, and eliminate the need for sample transportation. As it can be transported, this has allowed us to carry out extensive research of diffusion in explorations of Antartic ice and groundwater [49,50].

Research into an Earth-field NMR spectrometer also has the potential to deepen our fundamental understanding of nuclear magnetic resonance. By working with much lower magnetic field intensities than usual, we can gain new insights into the underlying principles of NMR and uncover previously unobserved behaviors and properties. This additional knowledge could open new avenues for research and application beyond the purely educational purposes mentioned above.

The main objective of this paper is to design and construct an NMR spectrometer that uses the Earth's magnetic field as a constant magnetic field to operate. In addition, the aim is to design systems that can generate acceptable and useful results under various operating conditions. To evaluate the efficiency and effectiveness of the Earth-field NMR spectrometer, its results are compared to those of conventional NMR spectrometers. One of the main advantages of Earth-based NMR spectrometers is the ability to use NMR spectrometers in situations where conventional NMR spectrometers cannot be used. Consequently, spectrometers are designed and developed so that they are as portable, durable, and user-friendly as possible to be used in a wider range of environments. Data collected by Earth's field NMR spectrometers must be processed and analyzed. Herein, we developed robust and user-friendly software systems to manage this process, simplify data interpretation, and allow users to extract the maximum value from the results obtained.

3. Principle of Operation and General Aspects

Nuclear magnetic resonance (NMR) is a physical phenomenon in which atomic nuclei in a magnetic field absorb and re-emit electromagnetic radiation. This process allows for an investigation of the physical and chemical properties of the nucleus itself and its surrounding environment.

Nuclear spin is a fundamental property of atomic nuclei related to the number of protons and neutrons they contain. Some nuclei with specific properties, such as ^{1}H

(protons), ^{13}C, ^{15}N, and ^{31}P, are NMR active, which means they have a nonzero spin quantum number. This spin imparts a magnetic moment to them, turning them into tiny magnetic bars.

In an NMR experiment, an external magnetic field, B_0, is applied. The magnetic moments of the nuclei align with the external field, but in quantum mechanics, this alignment is quantized; Figure 1. This means that nuclear spins can align in specific orientations dictated by their spin quantum number (m_l). For example, a nucleus with a spin quantum number of 1/2 (such as ^{1}H or ^{13}C) has two possible spin states in an external magnetic field: "spin-up" ($m_l = +1/2$) or "spin-down" ($m_l = -1/2$). These states have different energies, with the spin-down state being the higher energy state.

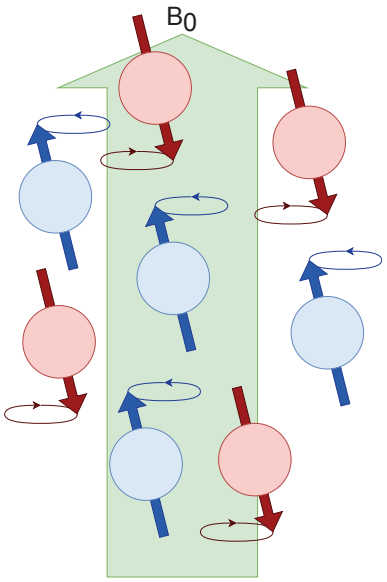


Figure 1. Energy levels considered before an external magnetic field B_0. Populations are generated that are quantized with two energy levels.

The constant that relates the Larmor frequency to the applied magnetic field is called the gyromagnetic constant, which, for ^1H, is 42.574 MHz/T. The energy levels produced can be quantified as follows:

$$E_I = h\nu_I = hm_I\gamma_I B_0 \tag{1}$$

where E_I is the energy level of the atom I, h is Planck's constant, m_I is the quantum spin number, γ_I is the gyromagnetic constant, and B_0 is the external magnetic field. These two energy levels are known as the alpha and beta states. For the hydrogen atom, the energy levels (E_I) for a 600 MHz (14.1 T) spectrometer are around 2×10^{-25} J. The thermal energy (E_T) that determines the population ratio of a partition distribution is around 4.1×10^{-21} J. Thus, the population associated with this energy state can be calculated as follows:

$$N_I = \frac{1}{2}e^{(E_I/E_T)} = \frac{1}{2}e^{\left(\frac{m_I h\gamma_I B_0}{k_b T}\right)} \tag{2}$$

The population ratio for the different levels can be estimated by subtracting the energy levels and dividing them by the sum of the two:

$$\frac{N_{+1/2} - N_{-1/2}}{N_{+1/2} + N_{-1/2}} = \frac{e^{\frac{h\gamma_I B_0}{k_b T}} - e^{\frac{-h\gamma_I B_0}{k_b T}}}{e^{\frac{h\gamma_I B_0}{k_b T}} + e^{\frac{-h\gamma_I B_0}{k_b T}}} = \tanh\left(\frac{h\gamma B_0}{2k_b T}\right) \approx \frac{h\gamma B_0}{2k_b T} \tag{3}$$

Equation (3) reveals that for a 600 MHz spectrometer at 288 K, the fraction of populations is around $\frac{\Delta N_{\pm 1/2}}{N} \approx 5 \times 10^{-5}$; therefore, only a small fraction will have a net magnetization M_0:

$$M_0 = N\frac{h\gamma}{2}\left(\frac{h\gamma B_0}{2k_bT}\right) = \frac{C_m V B_0}{T}\left(\frac{h^2\gamma^2}{4R}\right) \quad (4)$$

where N is the total number of populations, C_m is the molar concentration of the nuclei considered, V is the volume of the sample, B_0 is the external magnetic field, and R is the universal gas constant.

M_0 is the projection on the Z-axis parallel to the external magnetic field, referred to as the polarization in M_z of all spins, which defines the signal intensity in resonance. These energy levels are called Zeeman energy levels and determine the magnetization parallel to the external magnetic field (Figure 2). In NMR, a radiofrequency (RF) pulse is used to perturb the nuclear spins from their equilibrium state. If this RF pulse matches the Larmor frequency (a condition known as "on resonance"), the nuclear spins can absorb this energy and "flip" from the lower-energy state to the higher one. Once the RF pulse is turned off, the system will return to equilibrium, a process known as relaxation. During this process, the nuclei will emit the absorbed energy in the form of an electromagnetic signal, which can be detected and used to generate the NMR spectrum (Figure 2). The population difference between the alpha and beta states is very small but sufficient to produce an observable NMR signal, as explained previously. The system is in equilibrium when the ratio between the populations of the alpha and beta states remains constant. In this state, there is no net magnetization in the transverse plane (x-y). A short and intense burst of radiofrequency (RF) radiation, called an RF pulse, is applied perpendicular to B_0. If this RF pulse matches the Larmor frequency (the resonance frequency of the spins), the nuclear spins can absorb this energy, causing a 90° flip of their alignment with B_0 into the transverse plane (x-y). This generates a net magnetization in the transverse plane, creating a non-equilibrium state (Figure 2). It is a natural process based on the minimization of energy and maximization of entropy; see Table 1. In NMR, this natural process is observed by minimizing the spin energy and maximizing the nuclear spin phase.

Table 1. Model of equilibrium and non-equilibrium in the resonant process of nuclear spin.

Equilibrium $\mathbf{min(\Delta E - \Delta \phi)}$	**Non-Equilibrium** $\mathbf{max(\Delta E - \Delta \phi)}$
$M_z = M_0$	$M_z = M_0 \cos(\phi_p)$
$M_x = 0$	$M_x = 0$
$M_y = 0$	$M_y = M_0 \sin(\phi_p)$

The variables involved in the spin perturbation process are the intensity of the radiofrequency pulse (B_{RF}), the phase value, and the duration of the pulse (τ_p). The flip angle of the Z-axis over the X-Y plane is determined as $\phi_p = \tau_p \gamma B_{RF}$.

After turning off the RF pulse, the system does not remain in a non-equilibrium state. The spins gradually return to their equilibrium alignment along the Z-axis, a process known as relaxation. Relaxation occurs in two distinct processes: longitudinal relaxation (or T_1) and transverse relaxation (or T_2).

T_1 relaxation is an enthalpic process where a net electromagnetic emission occurs, producing a detectable signal as the system transitions from a non-equilibrium state to equilibrium. Also known as spin-lattice relaxation, T_1 relaxation is the process in which the net magnetization vector in the transverse plane returns to the Z-axis. This process is associated with the energy released by the spins as they return to equilibrium, causing the nuclei to revert to their original energy states. The relaxation time T_1 is a measure of how quickly this process occurs. During this relaxation process, the detectable signal mentioned earlier is observed as what is known as Free Induction Decay (FID). FID is

the time-dependent signal generated by the precessing net magnetization vector in the transverse plane immediately after the initial radiofrequency (RF) pulse is applied. As the magnetization vector relaxes back to its equilibrium state along the Z-axis, the FID signal decays exponentially, capturing the transition of the system back to equilibrium. The physical model of relaxation can be mathematically expressed as follows:

$$M_z = M_{z0}(1 - e^{-t/T_1}) \tag{5}$$

$$M_x = M_{x0} \sin(t\nu_H) e^{-t/T_1} \tag{6}$$

$$M_y = M_{y0} \cos(t\nu_H) e^{-t/T_1} \tag{7}$$

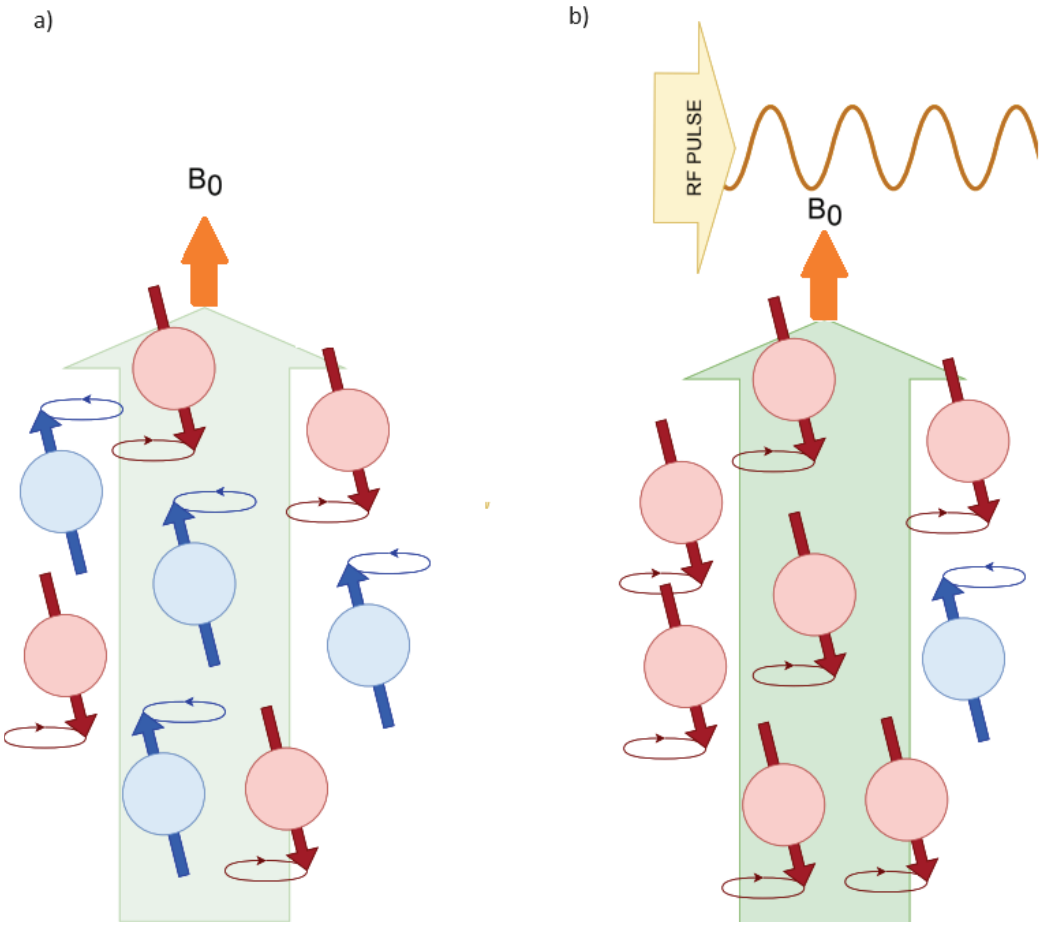


Figure 2. (**a**) Spin population distribution before the application of an RF pulse or polarizing coil in an external magnetic field Bo. Not all spins are aligned with Bo, leading to net magnetization. (**b**) Spin population distribution after the application of the RF pulse or polarizing coil. The pulse causes some spins to transition to a higher energy state, partially inverting the spin population, but not all spins are fully inverted. This diagram reflects the Zeeman energy levels and the alignment of nuclear spins with the external magnetic field. A radiofrequency (RF) pulse perturbs the nuclear spins from their equilibrium state, which are "on resonance".

Equation (5) shows what is called a coherent transition where photons are emitted, resulting in a detectable signal. The relaxation T_2 is an entropic process in which the detectable signal has a phase among the different nuclear spins, producing relaxation in the XY plane. Also known as spin–spin relaxation, T_2 relaxation is the process by which the spins in the XY plane dephase (i.e., lose their coherence). During this process, the transverse magnetization decays to zero as a result of interactions between the spins. The relaxation time T_2 is a measure of the rate at which the transverse magnetization decays. A

complete visualization of this relaxation process can be found in the following references for different experiments [7,8,51,52].

$$M_z = M_z(1 - e^{-t/T_1}) \tag{8}$$

$$M_x = M_x \sin(t\nu_H)e^{-t/T_2} \tag{9}$$

$$M_y = M_y \cos(t\nu_H)e^{-t/T_2} \tag{10}$$

This is the classical description of nuclear magnetization behavior developed by Bloch in his vector model in the presence of an external magnetic field and following excitation by an RF pulse. These equations, named after Felix Bloch, model nuclear spins as a bulk magnetization vector and describe their precession and relaxation behaviors. Although Bloch's model is incredibly useful and has greatly contributed to our understanding of NMR, it has several limitations.

- Lack of quantum effects: Bloch equations are classical and do not account for quantum mechanics effects. For example, they do not consider that the spin state of a nucleus is quantized.
- Assumption of a homogeneous magnetic field: The equations assume a perfectly homogeneous external magnetic field. In reality, magnetic fields often exhibit inhomogeneities, which can lead to various effects not covered by Bloch equations.
- Assumption of instantaneous excitation: The equations assume that the transition from equilibrium to non-equilibrium (during the RF pulse) is instantaneous. In real scenarios, the RF pulse has a finite duration, and its effect on the spin system might not be as abrupt as assumed by Bloch equations.
- Neglect of multiple spin interactions: Bloch's model does not consider the interaction of different spins with each other (spin–spin interactions), which can lead to more complex phenomena such as relaxation and spin–spin coupling.
- Assumption of large spin ensembles: Bloch equations consider the behavior of a spin ensemble, thus providing average values. They do not provide information on the behavior of individual spins within the ensemble.
- Simplification of relaxation processes: The equations simplify relaxation processes (T_1 and T_2) to single exponential decays, whereas in reality these processes can involve multiple components with different relaxation times.

Despite these limitations, the Bloch equations remain a fundamental tool in the understanding and interpretation of NMR, offering a simplified and efficient description of many phenomena encountered in this field. However, for a more detailed understanding of complex NMR phenomena, it is necessary to resort to quantum mechanical descriptions or more advanced models.

Nuclear magnetic resonance in the geomagnetic field is known as Earth's Field NMR (EFNMR), a special case of low-field NMR. When a sample is placed in a uniform magnetic field, nuclei with non-zero spin experience resonance at specific frequencies. This document exclusively covers proton nuclear magnetic resonance (NMR), primarily involving the ^{1}H isotope. Moreover, the experiments described in this study depend on the use of a polarizing coil to achieve the initial magnetization of the sample, which serves as an alternative to the traditional RF excitation pulse commonly used in high-field NMR. This choice is particularly important in the context of Earth-field NMR, where the magnetic field strength is extremely low, and consequently, the thermal polarization of the spins is also weak. The polarizing coil provides a stronger, more uniform magnetic field during the polarization phase, effectively increasing the initial magnetization of the sample before signal detection. This is very important in low-field conditions, where the natural magnetization would, otherwise, be insufficient to produce a detectable signal. This device uses separate coils for polarization and reception; each can be optimized independently for its specific function. The polarizing coil is designed to generate a strong and uniform magnetic field, while the receiving coil is optimized for maximum sensitivity and signal-to-noise ratio during

detection. This approach enhances the overall performance of the NMR system. Utilizing a dedicated polarizing coil helps to avoid potential artifacts or residual signals from the excitation phase that could interfere with the detection phase. This results in cleaner and more accurate signal acquisition, which is particularly important for reliable measurements in an Earth-field NMR setup.

4. Materials and Methods

This document exclusively covers proton NMR, mainly involving the ^{1}H isotope. In the Earth's magnetic field, which is around 45,000 nT depending on the location, the ^{1}H isotope has a Larmor frequency of approximately 2 kHz and generates a very weak signal of a few microvolts.

Figure 3 shows the block diagram and the complete architecture of this EFNMR spectrometer. The polarizing coil assembly includes two receiver coils (Rx) that are connected in an antiparallel configuration to cancel out environmental noise. The current flow through the polarizing coil is managed by a regulated power supply, controlled by a polarizing switch. The sample is placed inside one of the coils and the outputs of both Rx coils are fed into a differential amplifier through a relay-controlled switch.

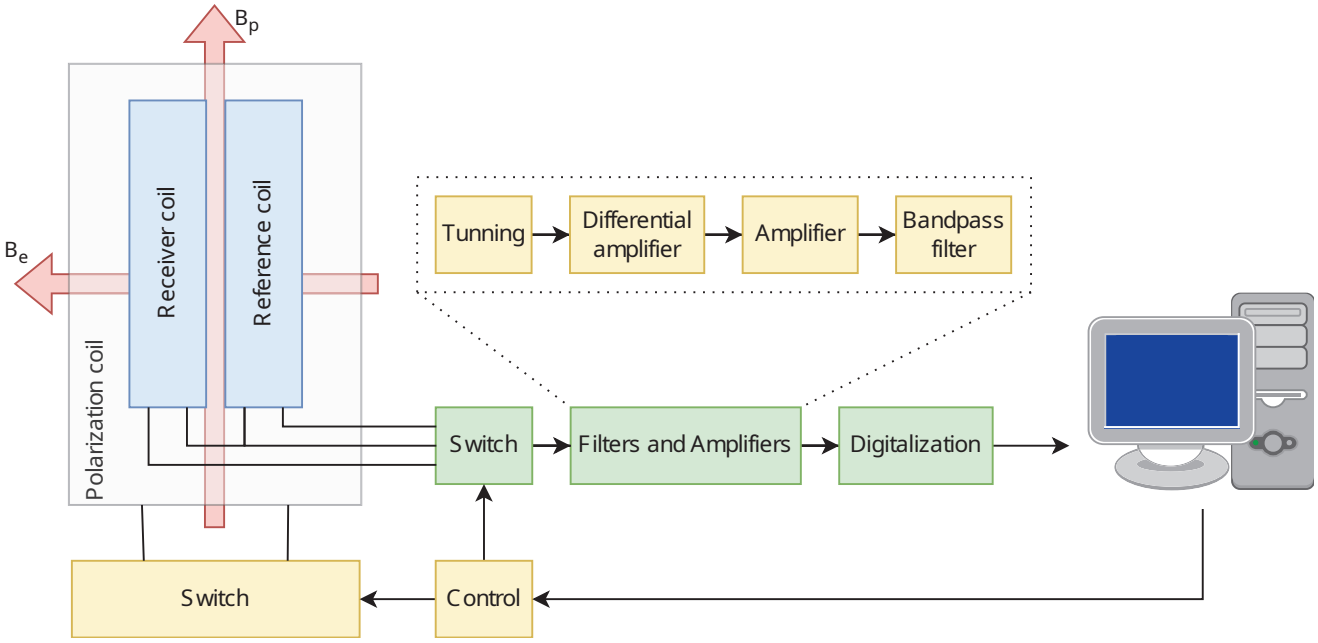


Figure 3. Block diagram of the electrotechnical device for nuclear spin excitation in the Earth's magnetic field. Bp and Be represent the polarization magnetic field and the Earth's magnetic field, respectively.

Given the small magnitude of the resonance signal, on the order of microvolts, signal amplification is essential before further processing. Therefore, the signal is channeled through a fixed-gain ultra-low noise differential amplifier and a low noise amplifier, and then to a bandpass filter, with a cutoff frequency around the Larmor frequency, thus eliminating unwanted noise and frequencies.

The filtered signal undergoes an additional amplification stage before being acquired by the microcontroller ΣΔ-ADC. Finally, signal processing is implemented on these data to extrapolate and visualize the desired magnetic resonance information.

4.1. CAD Design and 3D Printing

Three-dimensional printing is a very useful tool for prototyping. It allows for the fabrication of parts designed using CAD programs. In this project, 3D printing has been used to manufacture the cylinders on which the coils are wound and the base that holds them, as well as other complementary parts that will be discussed below. To design the parts used in this project, the CAD program was used. This software is a comprehensive 3D design tool that combines parametric modeling, rendering, simulation, and computer-aided manufacturing functions in a single integrated package.

The first part designed was the cylinder of the polarizing coil. This cylinder has a height of 145 mm, an outer diameter of 90 mm, and an inner diameter of 80 mm. Two holes with a diameter of 2 mm were made 20 mm from the base to allow the coil wire to pass inside the cylinder. The second part designed was the cylinder of the receiving coils. This cylinder has a height of 140 mm, an outer diameter of 32 mm, and an inner diameter of 28 mm. Similarly to the polarizing coil cylinder, this cylinder also has two holes made 20 mm from the base, but this time with a diameter of 1.5 mm to allow for the coil wires to pass inside the cylinders. Figure 4 shows the CAD design of the polarizing coil and the receiving coil cylinders.

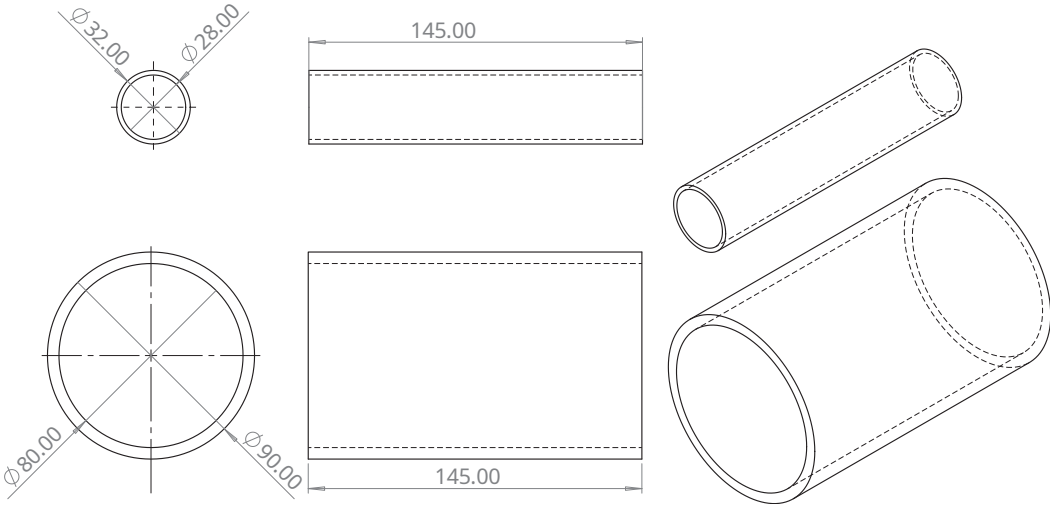


Figure 4. CAD design of the polarizing and receiving coil cylinders.

The three coils are aligned by means of a base that has grooves 15 mm deep and a width equivalent to the thickness of the walls of the cylinders where each of the coils is fitted, and this same base aligns the wires of connection of the coils to match the connections of the printed circuit.

In order to be able to modify the inclination of the coils, and to be able to align the terrestrial magnetic field in an optimal way, a support was designed, which holds the excitation coil and allows one to modify the angle with the base. The base of the equipment has a curved surface where the support can slide and two fixing screws, which allow us to secure the mounting in the desired position. The fixing screws used are plastic, so there is no need for any metal component that could alter the magnetic field near the coil. Figure 5 shows the CAD design of the various parts and the assembled model.

Once the design of all parts in CAD is completed, the files are transferred to a G-code translator. This kind of software is a sophisticated 3D model preparation software that enables the configuration and optimization of printing parameters. This process is essential before generating the G-code file required for 3D printing. The final result after the CAD design and printing is shown in Figure 6.

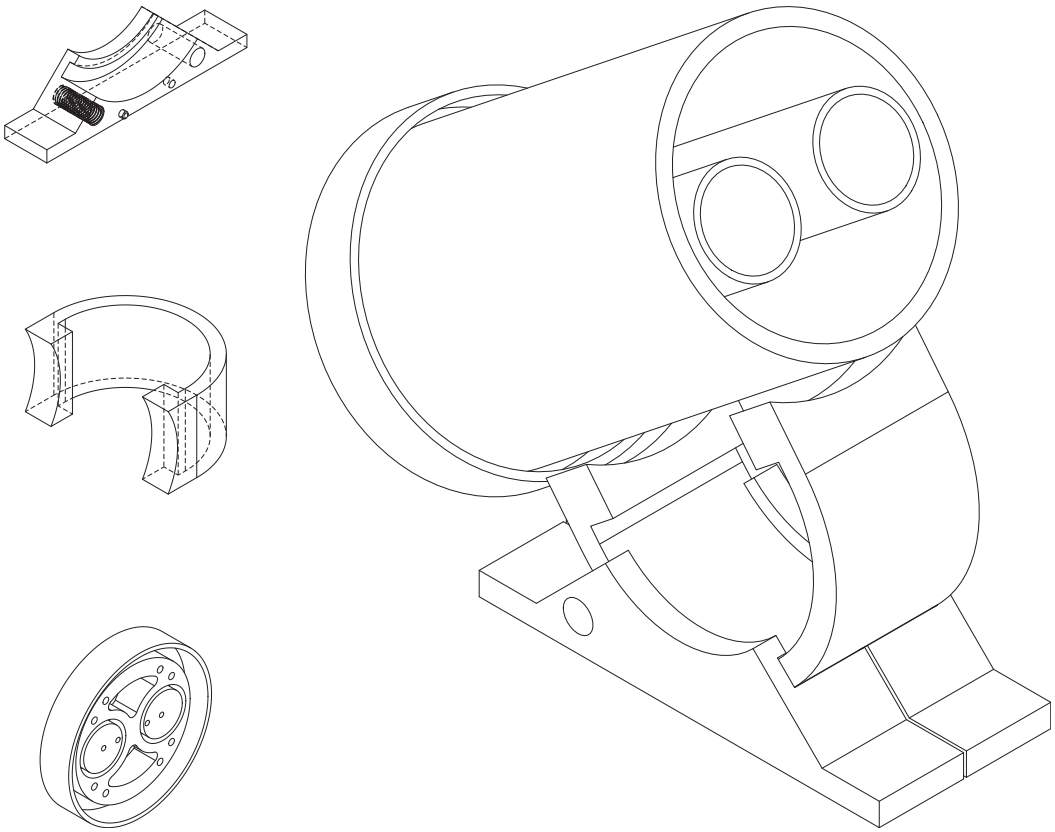


Figure 5. CAD design of the bases that support the coils and assembled model.

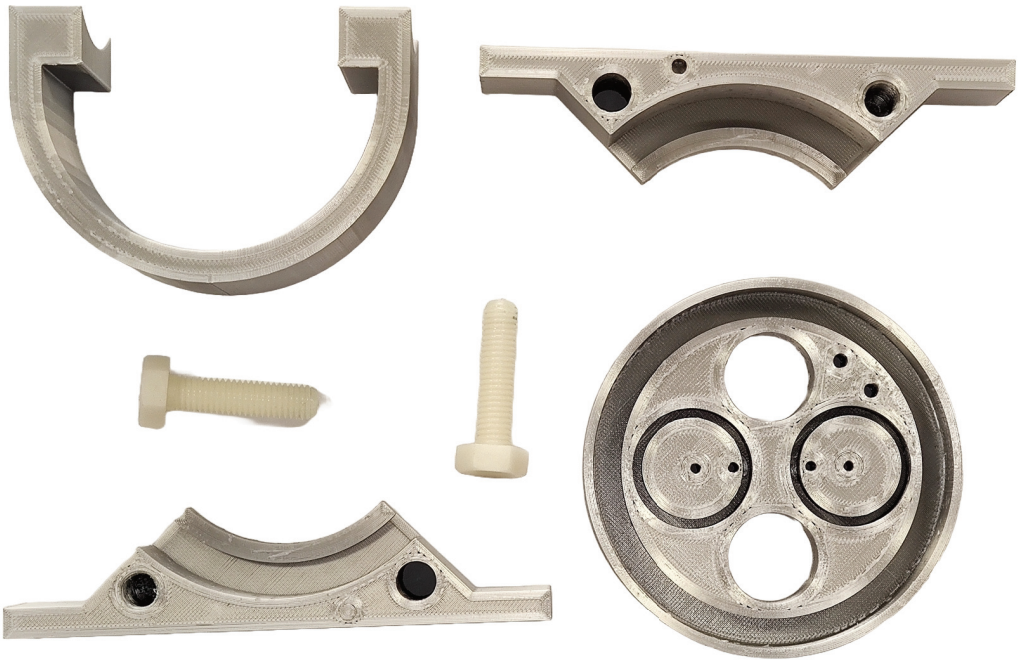


Figure 6. The 3D-printed parts with PETG and Nylon M10 threads

4.2. Polarizing Coil

The polarizing coil temporarily creates a polarizing magnetic field B_p. This field quickly turns off, leaving the magnetization instantly perpendicular to the Earth's magnetic field around which precession begins [53,54]. The polarizing coil and its supporting

structure must not contain any ferromagnetic material, see Figure 7 . Naturally, the Earth's field is very homogeneous, and if there is any ferromagnetic material near the coil, it will change the uniformity of the field, causing magnetic field inhomogeneities. To properly use the device during normal experiments, the entire system should be free of electromagnetic interference due to the weakness of the detected signal. However, since the goal of this project is to design a benchtop device, it has been designed to be robust under electromagnetic stress conditions. The purpose of the polarizing coil is to produce a magnetic field much greater than the Earth's magnetic field. To polarize the sample, a current of approximately 10 A is passed through the coil for 10 s. Therefore, the wire for winding the polarizing coil must have a large-enough cross-section to avoid heating while carrying this current. The specifications chosen for the polarizing coil are given in Table 2.

Table 2. Specifications of the polarizing coil.

Coil	Specification
Enameled copper wire	1.7 mm or 14 AWG
Inner diameter of the coil	80 mm
Outer diameter of the coil	90 mm
Height of the coil	145 mm
Number of wire layers	3
Number of turns per layer	60
Approximate inductance	3.4 mH
Approximate resistance	0.8 Ω

The polarizing coil cylinder is wound on a commercial PVC (vinyl polychloride) tube. The polarizing coil is placed in the center of the coil holder, which is provided with holes for passing the connections to the PCB, placed under this holder. The base, inclination part, and coil support are made of PETG, avoiding the use of metal. The tilt platform should form an angle close to 50° relative to the base. The angle ensures that when B_p is turned off, the magnetization is perpendicular to the Earth's field immediately prior to signal detection. This geometrical arrangement maximizes the strength of the signal. The Earth field in southern Spain (University of Almeria) has an inclination of 50.57° and a declination of 0.59°.

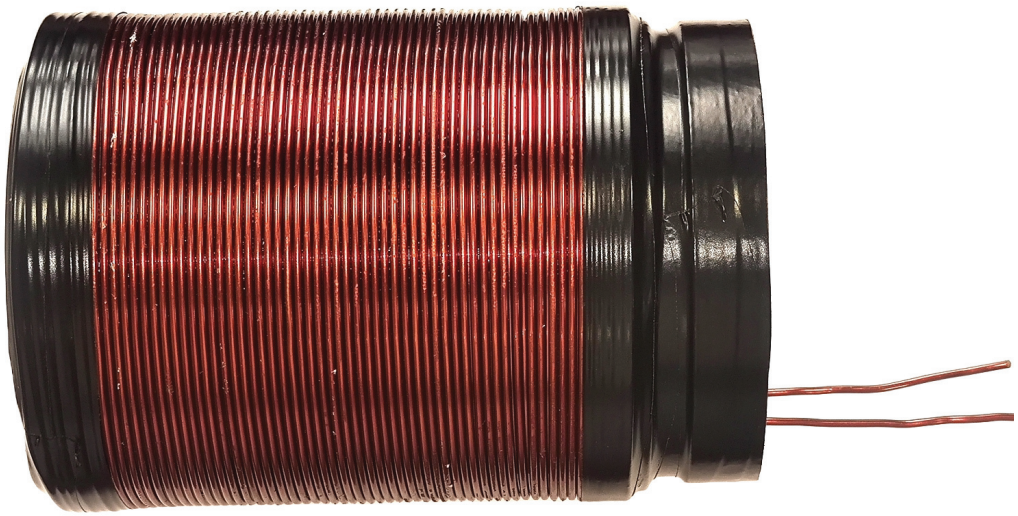


Figure 7. Wounded polarizing coil.

4.3. Receiver Coils (Antenna)

The receiving coils (R_x) are responsible for collecting the signal from the polarized sample in precession. The signal is extremely low, requiring many iterations of the experi-

ment to increase the signal-to-noise ratio (SNR). This process is known as signal averaging. The main problem affecting this device is noise, which is why the coils have been designed to minimize it. In addition to signal averaging, two nearly identical sensor coils are used to address this issue and maximize the SNR.

The two coils are wound in series but in opposite directions and placed on the same surface to cancel out noise (Figure 8). A real receiver coils built is shown in the Figure 9, this receiver coil is designed according its test tube. To receive the signal, the sample is placed in only one of the coils because if it is placed in both coils, the produced signals will cancel each other out, resulting in no signal being received at the output.

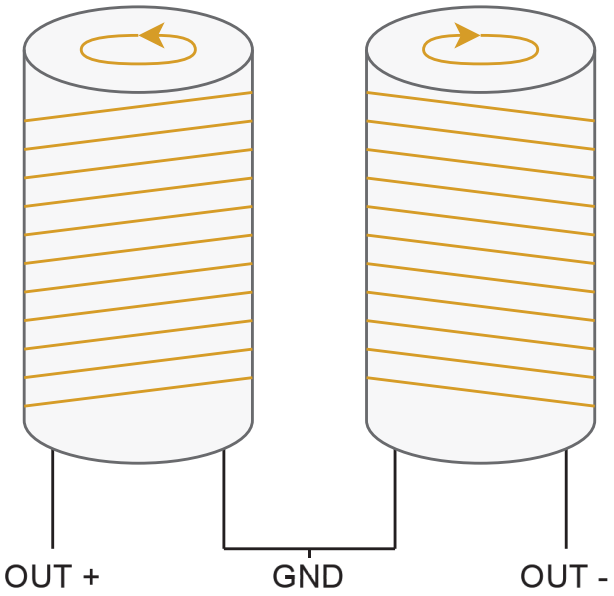


Figure 8. Winding diagram of the receiver coils.

The most important requirement for correct winding is that both coils must be as identical as possible to perfectly cancel out environmental noise. This is called common-mode rejection and can be achieved if the number of turns is equal for both coils and the wire is not twisted at any point during winding. To verify the similarity of the coils, one method is to weigh the coils before and after winding. As with the polarizing coil, the material used for the cylinders on which the receiving coils are wound is a commercial PVC pipe.

Figure 9. Wounded receiver coil and test tube.

As mentioned earlier, the device was designed to operate using the Earth's magnetic field, so any nearby ferromagnetic material must be avoided. Metallic objects such as copper, aluminum, and brass should also be kept away from the coil to avoid induced eddy currents when B_p is quickly turned off.

The wire's cross-section is also an important parameter. With a thin wire, we have more turns and a stronger signal, but it also has higher resistance and, therefore, more Johnson noise $\propto \sqrt{R}$, which we do not want. On the other hand, with a thick wire, we have fewer turns and a weaker signal compared to a thin wire, but the noise is lower and the quality factor can be higher. The thermal noise follows the resistance relationship in volts:

$$V_T = \sqrt{4K_B TR\Delta f} \tag{11}$$

where K_B is Boltzmann's constant (J/K), T is the temperature in K, R is the resistance, and Δf is the bandwidth. The specifications of the receiving coils are summarized in Table 3.

Table 3. Specifications of the receiver coil.

Coil	Specification
Enameled copper wire	0.812 mm or 21 AWG
Inner diameter of the coil	25 mm
Outer diameter of the coil	40 mm
Height of the coil	100 mm
Number of wire layers	4
Number of turns per layer	115
Approximate inductance per coil	2.56 mH
Approximate resistance per coil	2.5 Ω

To verify the proper functioning of the receiving coils, some preliminary tests were conducted. The coils were placed inside the polarizing coil, and a sinusoidal signal was applied to the polarizing coil using a function generator (Figure 10). This induces a sinusoidal wave in both receiving coils of the same amplitude and is approximately 180 degrees out of phase with the excitation source. Both waves cancel each other out, resulting in an almost null signal. All these signals were measured and analyzed in Matlab (Figure 11).

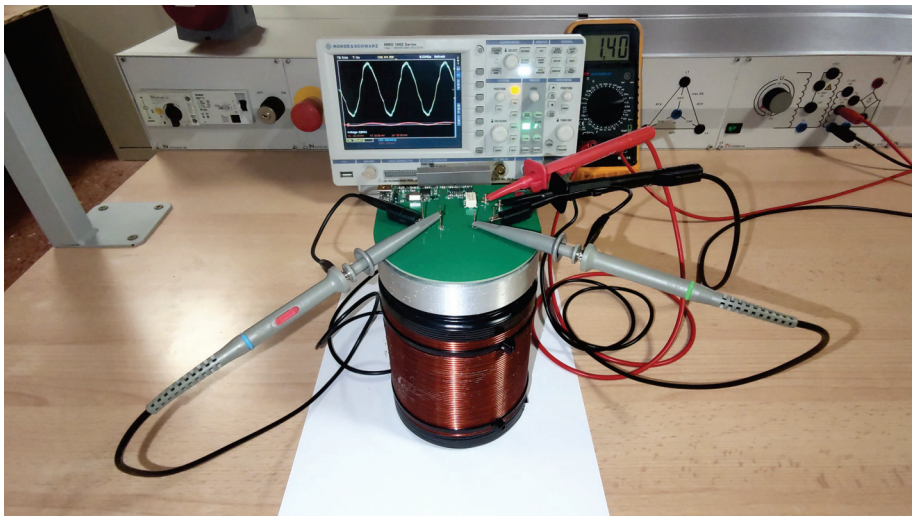


Figure 10. Testing of receiver coils in the electrotechnical laboratory .

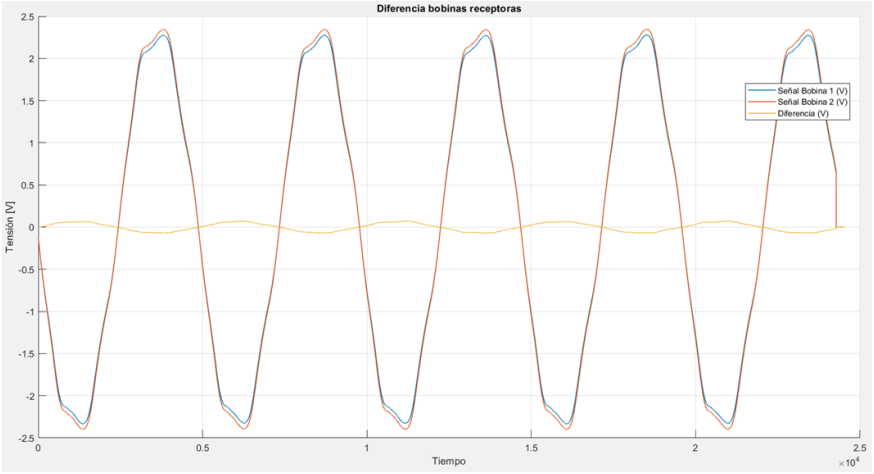


Figure 11. Test results give a good response of the avoiding signal between both coils. Only a little 180° dephased signal was measured, which ensures the cancelation of the signals.

4.4. Power Circuit

This circuit is responsible for turning the power supply of the polarizing coil on and off with a signal from the microcontroller. This circuit had to be carefully designed because, for the experiment to work correctly, the coil needs to discharge as quickly as possible, and consequently, so does the magnetic field it generates. Therefore, no switch can be used to turn off the polarizing coil. Mechanical relays could be used to turn the polarizing coil on and off, but they are not the best option due to several disadvantages. The relay operates like a basic switch and can produce an electric arc when it is turned off, as the coil has high induced voltages. Relays also have a short lifespan and take a few milliseconds to switch. In our case, the polarizing coil needs to be turned on and off within microseconds. Therefore, the simplest solution to this problem is to use a power MOSFET with a shorter switching time.

The MOSFET used is the IPD090N03L. The maximum allowed collector-emitter voltage is 30 V, and the base-emitter voltage is 20 V. The maximum current it can drain is 40 A. The IPD090N03L has an extremely low internal on-resistance of 0.009 Ω. The positive power supply of +12 V is connected to one terminal of the polarizing coil, while the other terminal is connected to the MOSFET's collector. The MOSFET emitter is connected to the ground. The MOSFET is operated by the microcontroller through a specific driver that boosts the voltage and current provided by the microcontroller to levels that ensure fast switching with the least amount of losses possible. Because high voltages are induced during the switching of the MOSFET, it is important to include a diode that absorbs these surges without damaging the MOSFET. Given the low resistance of the MOSFET, the amount of heat generated during switching can be dissipated by the PCB itself. The complete circuit schematic can be seen in Figure 12.

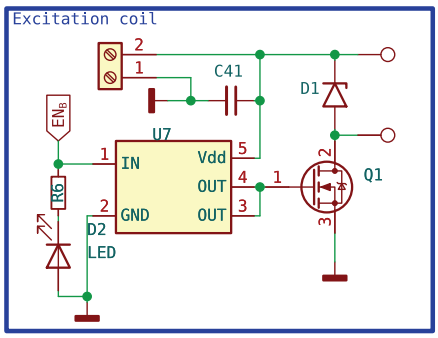


Figure 12. Circuit schematic of excitation coil control.

4.5. Receiver Circuit

The signal detected by the receiver coils is extremely small (on the order of microvolts). This signal needs to be converted by the microcontroller ADC to be processed by a computer, but it needs to be adapted; for this purpose, six different stages are designed as shown in Figure 13.

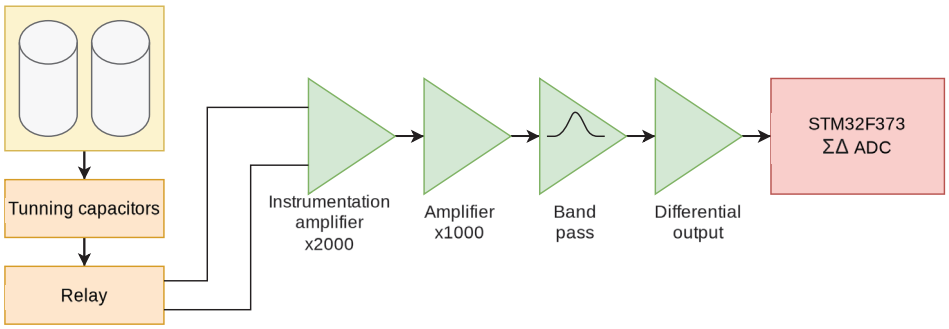


Figure 13. Block diagram of the amplification and filtering circuit.

The first stage is a capacitor block. These capacitors act as tuners and are responsible for enhancing the signal before it reaches the instrumentation amplifier. This is achieved by placing these capacitors in parallel with the receiving coil. The value of the capacitor can be found using the resonant frequency formula, Formula (12).

$$f = \frac{1}{2\pi\sqrt{LC}} \tag{12}$$

The frequency f is adjusted (adjusting C) to be as close as possible to the Larmor frequency of the proton, which is approximately 2 KHz. Here, L is the combined inductance of both coils, and its value is 5.12 mH. The optimal value of the tuning capacitor is 1.13 µF. This value is approximated by 1 µF and two 56 nF.

The second stage is a double relay controlled through the microcontroller that connects the output of the coils. The normal position of the relay switches is open. Each end of the receiving coils is connected to one of the relay switches, while the common point of the two coils is connected to the ground. When the microcontroller sends a signal to the transistor, it connects the relays to the +5 V power supply, causing the relay switches to activate, allowing the signal from the receiving coils to pass to the instrumentation amplifier. The circuit is shown in Figure 14.

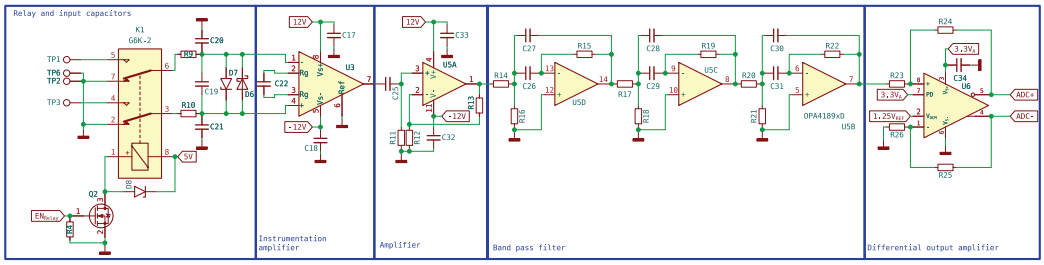


Figure 14. Switching amplification and filtering input stages circuit.

Third, there is an ultra-low-noise instrumentation preamplifier. An INA848 is used, which is a high-precision instrumentation amplifier that requires no external components for a fixed gain of 2000. Since the signal is in microvolts, only low-noise amplifiers can be used. Both ends of the coils are connected to the input of the instrumentation amplifier, while the sample is placed in only one of the coils, as mentioned earlier. When the signal arrives, the signals from the two inputs of the instrumentation amplifier are subtracted, and the signal emitted by the proton is obtained.

The fourth stage is an amplifier stage, with a gain of 1000, built with one of the four operational amplifiers available in the OPA4189. In the fifth stage, signal filtering is performed, using the remaining three amplifiers to perform a band-pass filter of order 6, designed to attenuate the frequencies outside the range of interest for this application. Specifically, 1.8 kHz center frequency and a width of 1.5 kHz have been chosen. The filter response is shown in Figure 15.

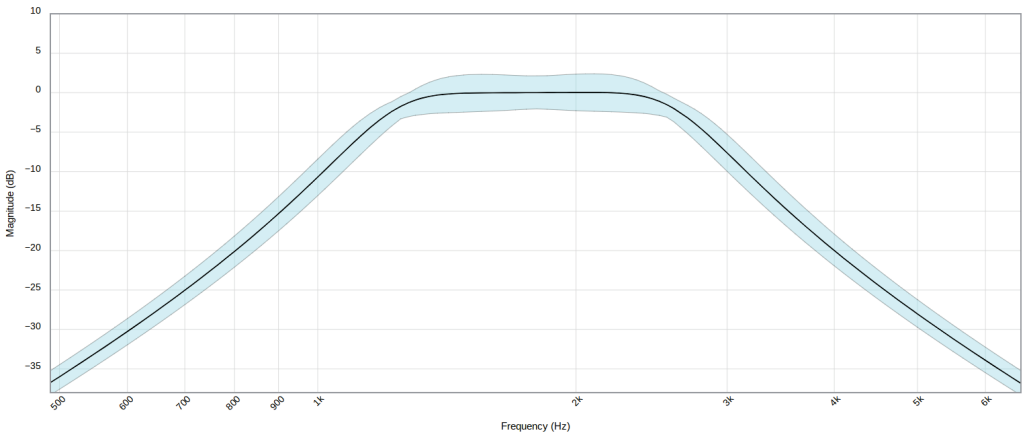


Figure 15. Filter stage frequency response. Blue band shadow includes components' tolerances.

The last stage transforms the signal to a differential mode and adapts the impedance values to those expected by the 16-bit ADC incorporated in the microcontroller. In this stage, the signal is shifted from referenced to GND to swing around $V_{ref}/2$, which is required to meet the input ranges of the ADC. Both the reference voltage and $V_{ref}/2$ are supplied by a precision voltage reference, in this case a REF2025.

The differential mode used significantly improves the conversion by increasing the noise immunity. No gain is introduced at this stage, so the total system gain is 2,000,000. The complete input schematic is shown in Figure 14.

4.6. Acquisition Sequence

The quality of the data obtained is strongly dependent on the precise timing of the different stages involved in the process. By design, this device alternates between the excitation stage, in which the polarizing coil is involved, and the relaxation stage, in which the receiver coils are involved. In this case, the acquisition process is fixed, and the timing parameters and the number of repetitions can be configured at the beginning of each experiment.

At the beginning of each scan of the experiment, the microcontroller activates the current conduction to the biasing coil. After the excitation time (τ_1), the coil is turned off. Once a preset time (τ_2) has elapsed, which is necessary for the field of the biasing coil to discharge, the relay closes and the receiver coils are connected to the amplifiers. Digitization begins after a second period (τ_2), which ensures complete closure of the relay, the duration of which varies according to the number of samples required (τ_3). At the end of the sampling, the relay opens and waits for the relaxation time (τ_4). This process is repeated for the indicated number of times, improving the system's ability to distinguish signal from noise.

Figure 16 details the temporal sequence that the device follows during each operation cycle. The time intervals associated with each phase of the sequence are detailed in Table 4.

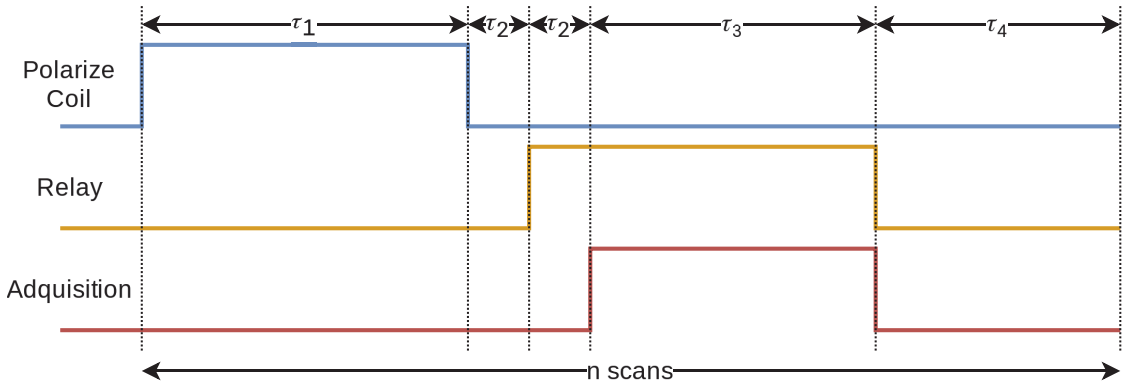


Figure 16. Sequence of pulses followed by the device as a function of time.

Table 4. Times for each stage of the sequence.

Stage	Symbol	Minimum	Maximum	Default
Polarizing coil activated	τ_1	1 s	32 s	10 s
Pre-acquisition	τ_2	-	-	100 ms
Acquisition	τ_3	1000 samples	5,000,000 samples	250,000 samples
Acquisition sample rate	-	-	-	50,000 SPS
Waiting between experiments	τ_4	1 s	32 s	5 s
Number of scans	N	1	5000	650

4.7. Control and Communications

The control tasks of the equipment, the digitalization of the amplified and adapted signals in the previous section, and the communication with a computer are handled by an ARM microcontroller (STM32F373). This microcontroller has the required features for the correct operation of the unit, including a 16-bit-resolution analog–digital converter and USB communication for sending the large number of samples that need to be acquired. Of course, it has diverse input and output pins, timers, and SPI and I2C communication ports, among others. To enable USB communication, it is required to use a 72 MHz clock frequency obtained from an 8 MHz crystal and the internal PLL.

The required design is simple, having only the outputs corresponding to the activation of the polarizing coil and the relay, the analog inputs for the signal, and a communication port with the computer. The rest of the elements are auxiliary to achieve the correct operation of the instrument. Figure 17 shows the microcontroller section as well as the rest of the auxiliary elements of the circuit.

Since the coils need to be properly aligned with the terrestrial magnetic field, a 3-axis magnetometer is also incorporated, which is accessed by the microcontroller via the I2C bus. As the PCB is precisely aligned with the coils, it is convenient and practical to be able to use this magnetometer to adjust the position before starting the measurements, ensuring better results. The read values are visible from the program interface.

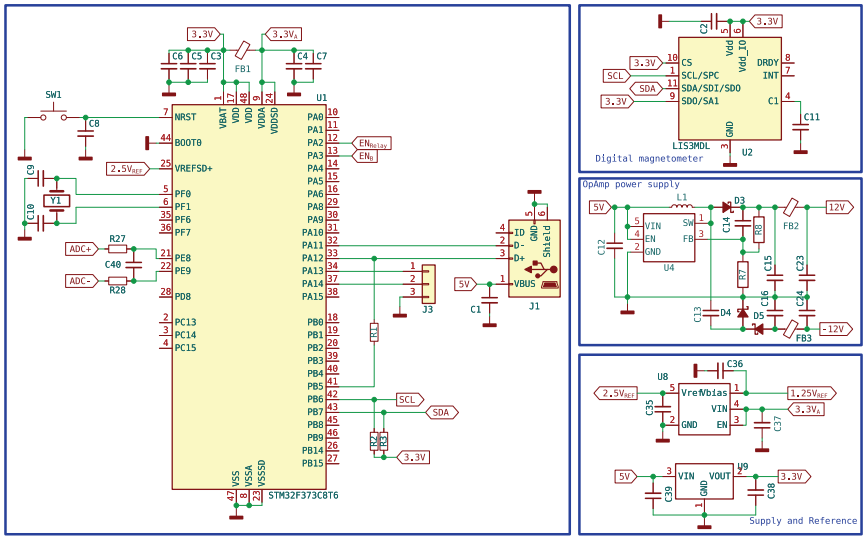


Figure 17. Microcontroller schematic and auxiliary components.

4.8. Computer Software

To ensure proper use of the unit, we developed basic control software in C#. This language for programming was chosen because it is driven by the extensive library support and the language's ease of use, facilitating the enhancement and swift implementation of new experiments and functionalities. The main window (Figure 18) presents configurable experimental parameters, enabling the user to adjust the settings as required.

- The number of samples to capture; according to this number and taking into account the fixed sampling frequency, the duration of this period is indicated, corresponding to τ_3.
- Number of scans or number of times the acquisition cycle will be repeated. Corresponds to the parameter N.
- Recovery time, or time required after acquisition until a new cycle starts. Corresponds to τ_4.
- Pulse time or polarizing coil activation time. Corresponding to τ_1.

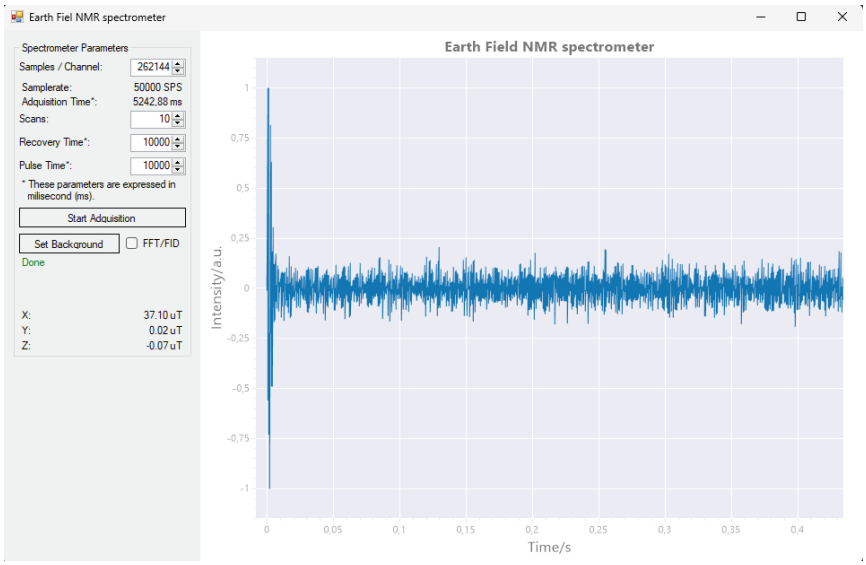


Figure 18. The main window of the computer software is divided into two parts. On the left side, there is a control panel with textboxes and buttons for managing the acquisition process. On the right side, there is a primary plot window that displays either the detected signal or its Fast Fourier Transform (FFT) analysis.

Although the software is intended to test the functionality of the machine, the ability to model the noise is incorporated, saving the results and allowing one to improve the behavior in subsequent experiments. This function is performed by running one or more scans without introducing any sample. The frequency response (FFT) obtained is stored in a file and used subsequently to suppress the background noise. The software also allows you to switch between the Fast Fourier Transform view (FFT) and the free-induction decay view.

4.9. Cost Analysis

The cost of the manufactured device is minimized due to its basic design and the use of readily available and inexpensive components. This approach ensures that the overall cost remains low while fulfilling the necessary functionality and performance requirements. The use of standard materials and simple manufacturing processes further enhances the cost effectiveness of the machine (see Table 5).

Table 5. Cost breakdown of EFRMN device components.

Item	Quantity	Cost
PETG Filament	1 kg	24 €
PCB Fabrication	-	10 €
Components	-	100 €
Copper Wire 1.7 MM	70 m	30 €
Copper Wire 0.8 MM	120 m	25 €
Wiring	-	10 €
Power supply	1	30 €
	Total	229 €

5. Results and Discussion

In this study, we successfully designed and built a complete assembly and electronic component EFNMR spectrometer from scratch. All parts, including the spectrometer framework, were manufactured with PETG material and 3D-printing technology with CAD design and 3D printing to ensure the device's robust and accurate structure.

The electronic system includes a custom coil to generate the required magnetic field and a sensitive receiver coil to detect NMR signals. The STM32F373 microcontroller was used to control pulse sequences and read Analog–Digital Converter (ADC) signals from the receiver. This controller provides the computational power and flexibility necessary for accurate timing and signal processing. In Figure 19 is shown the successfully assembled device, integrating all components into a fully functional and cohesive EFNMR spectrometer.

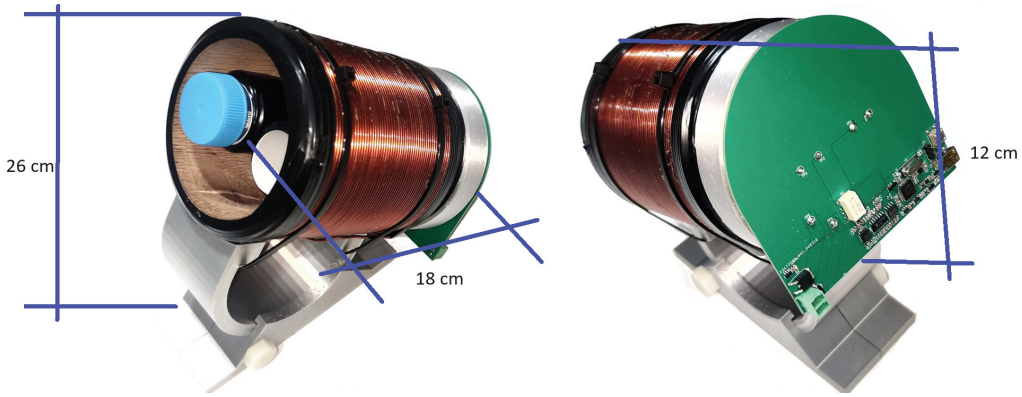


Figure 19. The final device was successfully assembled, integrating all components into a cohesive and fully functional EFNMR spectrometer.

The control software has a user-friendly interface and is written in C#. This C# application can be visualized in Figures 18 and 20. This interface facilitates easy interaction with

EFNMR spectrometers, allowing users to configure experimental parameters, begin measuring, and analyze the results. The software is designed to focus on simplicity and accessibility and enable users of different levels of expertise to use the spectrometer effectively.

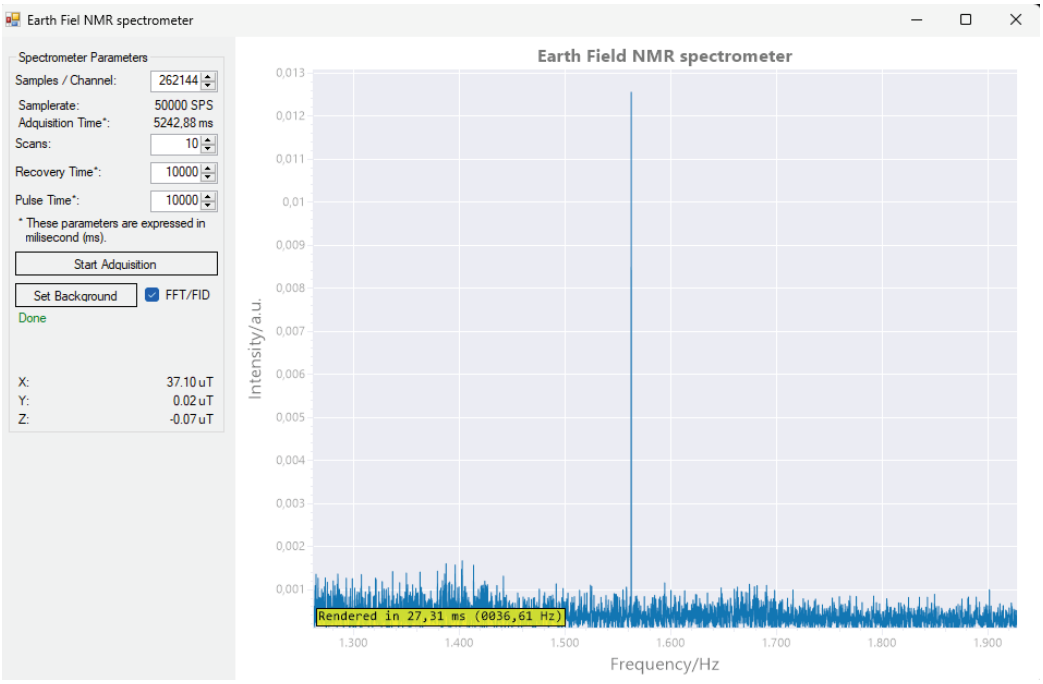


Figure 20. ^{1}H MR spectrum test was successfully conducted. The magnetometer detected a magnetic field strength of 37.10 µT, resulting in a resonance frequency of 1579 Hz for the ^{1}H nuclei.

All aspects of the project, including electronic designs, coil specifications, and control software, have been made available as open source resources on GitHub, https://github.com/fmarrabal/Earth-Field-NMR (accessed on 18 July 2024). This ensures that other researchers and enthusiasts can replicate, modify, and improve our work, fostering further advancements in the field of EFNMR spectroscopy.

The system was tested using standard samples, and the results demonstrated the capability of our EFNMR spectrometer to produce a ^{1}H NMR spectrum. The results indicate that our EFNMR spectrometer can serve as a powerful tool for a wide range of applications, including chemical analysis, geophysical exploration, and educational purposes. Figure 20 shows a clear signal peak at approximately 1579 Hz, which is an expected result for Earth-field NMR. However, the baseline noise is quite prominent, suggesting that the overall SNR is low. The noise appears to be relatively high compared to the signal peak. In an ideal NMR experiment, the signal should stand out distinctly above the noise floor. Here, the noise level is only marginally lower than the intensity of the peak, which implies that the SNR could be improved. In Earth-field NMR, the sensitivity is inherently lower than in high-field NMR because of the significantly smaller magnetic field strength. This leads to a reduced net magnetization, which directly impacts the strength of the detected signal. To improve the sensitivity, it is possible to perform the following:

- Increase the number of scans: averaging more scans can improve the SNR by reducing random noise.
- Optimize pulse parameters: adjusting pulse times and recovery times to ensure the optimal alignment of the spins prior to detection can enhance the signal.
- Signal averaging and background subtraction: techniques such as averaging multiple acquisitions and subtracting background noise can also be helpful in enhancing the detectable signal.

- Magnetization enhancement techniques: implementing techniques such as hyperpolarization or using stronger polarizing coils might also improve initial magnetization, thus increasing the signal intensity.

This device and software are designed to optimize sensitivity through a variety of advanced techniques. They are capable of setting the number of scans to improve the signal-to-noise ratio (SNR) by averaging multiple acquisitions, thus reducing random noise. The system can also fine-tune pulse times and recovery periods to ensure optimal spin alignment before detection, which enhances the detected signal's strength. In addition, the software facilitates signal processing methods, such as background subtraction and real-time signal averaging, to further enhance the clarity of the detected peaks. Although these capabilities significantly improve the sensitivity of the NMR measurements, it is important to note that the system does not incorporate magnetization enhancement techniques, which would require external interventions beyond the microcontroller and software scope.

In summary, the development of this EFNMR spectrometer demonstrates the feasibility of constructing high-quality, low-cost NMR systems using readily available materials and open-source technology. Our work lays the foundation for future innovations in portable and accessible NMR spectroscopy oriented to AI-enhanced low-field NMR technology.

Author Contributions: F.M.A.-C. and I.F.; methodology, F.M.A.-C., E.V. and J.A.M.-L.; software, F.M.A.-C. and E.V.; validation, I.F.; formal analysis, I.F.; investigation, E.L.-L.; resources, E.L.-L.; data curation, E.L.-L.; writing—original draft preparation, F.M.A.-C. and I.F.; writing—review and editing, F.M.A.-C. and I.F.; visualization, E.V. and E.L.-L.; supervision, J.A.M.-L.; project administration, F.M.A.-C. and I.F.; funding acquisition, F.M.A.-C. and I.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the State Research Agency of the Spanish Ministry of Science and Innovation (PID2021-126445OB-I00), and by the Gobierno de España MCIN/AEI/10.13039/501100011033 and Unión Europea "Next Generation EU"/PRTR (PDC2021-121248-I00, PLEC2021-007774 and CPP2022-009967).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The repository https://github.com/fmarrabal/Earth-Field-NMR (accessed on 18 July 2024) provides complete device schematics, PCB design, microcontroller firmware, and computer software for the device.

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

References

1. Becker, E.D.; Fisk, C.L.; Khetrapal, C.L. The Development of NMR. In *Encyclopaedia of Nuclear Magnetic Resonance*; Grant, D.M., Harris, R.K., Eds.; Wiley: Chichester, NJ, USA, 1996; Volume 1, pp. 1–158. [CrossRef]
2. Gerlach, W.; Stern, O. Experimentelle Prüfung der Richtungsquantelung im Magnetfeld. *Z. Phys.* **1922**, *9*, 349–352. [CrossRef]
3. Rabi, I.I.; Zacharias, J.R.; Millman, S.; Kusch, P. A New Method of Measuring Nuclear Magnetic Moment. *Phys. Rev.* **1938**, *53*, 318. [CrossRef]
4. Purcell, E.M.; Torrey, H.C.; Pound, R.V. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev.* **1946**, *69*, 37. [CrossRef]
5. Bloch, F.; Hansen, W.W.; Packard, M.E. Nuclear Induction. *Phys. Rev.* **1946**, *69*, 127. [CrossRef]
6. Heitler, W.; Teller, E. Time effects in the magnetic cooling method. *Proc. R. Soc. London. Ser. A* **1936**, *155*, 62. [CrossRef]
7. Hahn, E.L. Spin Echoes. *Phys. Rev.* **1950**, *80*, 580–594. [CrossRef]
8. Carr, H.Y.; Purcell, E.M. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys. Rev.* **1954**, *94*, 630. [CrossRef]
9. Meiboom, S.; Gill, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instruments* **1958**, *29*, 688. [CrossRef]
10. Bloch, F. Nuclear Induction. *Phys. Rev.* **1946**, *70*, 460–474. [CrossRef]
11. Bloembergen, N.; Purcell, E.M.; Pound, R.V. Relaxation Effects in Nuclear Magnetic Resonance Absorption. *Phys. Rev.* **1948**, *73*, 679. [CrossRef]
12. Proctor, W.G.; Yu, F.C. The Dependence of a Nuclear Magnetic Resonance Frequency upon Chemical Compound. *Phys. Rev.* **1950**, *77*, 717. [CrossRef]
13. Dickinson, W.C. Dependence of the F^{19} Nuclear Resonance Position on Chemical Compound. *Phys. Rev.* **1950**, *77*, 736–737. [CrossRef]
14. Knight, W.D. Nuclear Magnetic Resonance Shift in Metals. *Phys. Rev.* **1949**, *76*, 1259–1260. [CrossRef]

15. Arnold, J.T.; Dharmatti, S.S.; Packard, M.E. Chemical Effects on Nuclear Induction Signals from Organic Compounds. *J. Chem. Phys.* **1951**, *19*, 507. [CrossRef]
16. Meyer, L.H.; Saika, A.; Gutowsky, H.S. Electron Distribution in Molecules. III. The Proton Magnetic Spectra of Simple Organic Groups1. *J. Am. Chem. Soc.* **1953**, *75*, 4567. [CrossRef]
17. Proctor, W.G.; Yu, F.C. On the Nuclear Magnetic Moments of Several Stable Isotopes. *Phys. Rev.* **1951**, *81*, 20–30. [CrossRef]
18. Gutowsky, H.S.; McCall, D.W. Nuclear Magnetic Resonance Fine Structure in Liquids. *Phys. Rev.* **1951**, *82*, 748–749. [CrossRef]
19. Overhauser, A.W. Polarization of Nuclei in Metals. *Phys. Rev.* **1953**, *92*, 411–415. [CrossRef]
20. Carver, T.R.; Slichter, C.P. Polarization of Nuclear Spins in Metals. *Phys. Rev.* **1953**, *92*, 212–213. [CrossRef]
21. Solomon, I. Relaxation Processes in a System of Two Spins. *Phys. Rev.* **1955**, *99*, 559–565. [CrossRef]
22. Freeman, R.; Morris, G.A. The Varian story. *J. Magn. Reson.* **2015**, *250*, 80. [CrossRef]
23. Shoolery, J.N. The development of experimental and analytical high resolution NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **1995**, *28*, 37. [CrossRef]
24. Corey, E.J.; Burke, H.J.; Remers, W.A. Distinction between Cycloheptatriene and Bicycloheptadiene Structures by Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* **1955**, *77*, 4941. [CrossRef]
25. Müller, L.; Kumar, A.; Ernst, R.R. Two-Dimensional Carbon-13 NMR Spectroscopy. *J. Chem. Phys.* **1975**, *63*, 5490–5491. [CrossRef]
26. Aue, W.P.; Bartholdi, E.; Ernst, R.R. Two-dimensional Spectroscopy. Application to Nuclear Magnetic Resonance. *J. Chem. Phys.* **1976**, *64*, 2229. [CrossRef]
27. Andrew, E.R.; Szczesniak, E. A Historical Account of NMR in the Solid State. *Prog. Nucl. Magn. Reson. Spectrosc.* **1995**, *28*, 11–36. [CrossRef]
28. Bodenhausen, G.; Ruben, D.J. Natural Abundance Nitrogen-15 NMR by Enhanced Polarization Transfer. *Chem. Phys. Lett.* **1980**, *69*, 185. [CrossRef]
29. Bax, A.; Griffey, R.H.; Hawkins, B.L. Correlation of Proton and Nitrogen-15 Chemical Shifts by Multiple Quantum NMR. *J. Magn. Reson.* **1983**, *55*, 301. [CrossRef]
30. Sugiki, T.; Kobayashi, N.; Fujiwara, T. Modern Technologies of Solution Nuclear Magnetic Resonance Spectroscopy for Three-dimensional Structure Determination of Proteins Open Avenues for Life Scientists. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 328. [CrossRef]
31. Lauterbur, P.C. Image Formation by Induced Local Interactions: Examples of Employing Nuclear Magnetic Resonance. *Nature* **1973**, *242*, 190–191. [CrossRef]
32. Mansfield, P. Multi-Planar Image Formation Using NMR Spin Echoes. *J. Phys. C Solid State Phys.* **1977**, *10*, L55. [CrossRef]
33. Alam, T.M.; Jenkins, J.E. *HR-MAS NMR Spectroscopy in Material Science*; IntechOpen, London, UK, 2012 ; pp. 279–306. [CrossRef]
34. Reimer, J.A. Development of NMR: Solid-State NMR and Materials Science, Post 1995. In *Encyclopaedia of Nuclear Magnetic Resonance*; Grant, D.M., Harris, R.K., Eds.; Wiley: Chichester, UK, 2012; Volume 1, pp. 457–472. [CrossRef]
35. Packard, M.; Varian, R. Free Induction in the Earth's Magnetic Field. In *American Physical Society Minutes of the Stanford Meeting December 28, 29, and 30, 1953*; Kaplan, J., Ed.; American Chemical Society (ACS): Washington, DC, USA, 1954; Volume 93, p. Abstract A7.
36. Waters, G.S. A Measurement of the Earth's Magnetic Field by Nuclear Induction. *Nature* **1955**, *176*, 691. [CrossRef]
37. Courtillot, V.; Le Mouël, J. The study of Earth's magnetism (1269–1950): A foundation by Peregrinus and subsequent development of geomagnetism and paleomagnetism. *Rev. Geophys.* **2007**, *45*, 1–31. [CrossRef]
38. Koenig, S.H.; Brown, R.D.; Adams, D.; Emerson, D.; Harrison, C.G. Magnetic Field Dependence of 1/T1 of Protons in Tissue. *Investig. Radiol.* **1984**, *19*, 76–81. [CrossRef]
39. Balcı, E.; Rameev, B.; Acar, H.; Mozzhukhin, G.V.; Aktaş, B.; Çolak, B.; Kupriyanov, P.A.; Ievlev, A.V.; Chernyshev, Y.S.; Chizhik, V.I. Development of Earth's Field Nuclear Magnetic Resonance (EFNMR) Technique for Applications in Security Scanning Devices. *Appl. Magn. Reson.* **2016**, *47*, 87–99. [CrossRef]
40. Sato-Akaba, H.; Itozaki, H. Development of the Earth's Field NMR Spectrometer for Liquid Screening. *Appl. Magn. Reson.* **2012**, *43*, 579–589. [CrossRef]
41. Duval, E.; Ranft, J.; Béné, G. Détermination des signes relatifs des constantes de couplage 31 P- 1 H dans le tripropyl phosphate par R.M.N. dans le champ magnétique terrestre. *Mol. Phys.* **1965**, *9*, 427–431. [CrossRef]
42. Meriles, C.A.; Sakellariou, D.; Heise, H.; Moule, A.J.; Pines, A. Approach to High-Resolution ex Situ NMR Spectroscopy. *Science* **2001**, *293*, 82–85. [CrossRef] [PubMed]
43. Appelt, S.; Häsing, F.W.; Kühn, H.; Perlo, J.; Blümich, B. Mobile High Resolution Xenon Nuclear Magnetic Resonance Spectroscopy in the Earth's Magnetic Field. *Phys. Rev. Lett.* **2005**, *94*, 197602. [CrossRef]
44. Appelt, S.; Kühn, H.; Häsing, F.W.; Blümich, B. Chemical analysis by ultrahigh-resolution nuclear magnetic resonance in the Earth's magnetic field. *Nat. Phys.* **2006**, *2*, 105–109. [CrossRef]
45. Robinson, J.N.; Coy, A.; Dykstra, R.; Eccles, C.D.; Hunter, M.W.; Callaghan, P.T. Two-dimensional NMR spectroscopy in Earth's magnetic field. *J. Magn. Reson.* **2006**, *182*, 343–347. [CrossRef] [PubMed]
46. Appelt, S.; Häsing, F.W.; Kühn, H.; Sieling, U.; Blümich, B. Analysis of molecular structures by homo- and hetero-nuclear J-coupled NMR in ultra-low field. *Chem. Phys. Lett.* **2007**, *440*, 308–312. [CrossRef]
47. Zax, D.B.; Bielecki, A.; Zilm, K.W.; Pines, A.; Weitekamp, D.P. Zero field NMR and NQR. *J. Chem. Phys.* **1985**, *83*, 4877–4905. [CrossRef]
48. Kohl, M.; Odehnal, M.; Petříček, V.; Tichý, R.; Šafrata, S. Observation of higher order NMR larmor lines by SQUID in solids at low magnetic field. *J. Low Temp. Phys.* **1988**, *72*, 319–343. [CrossRef]

49. Hunter, M.W.; Dykstra, R.; Lim, M.H.; Haskell, T.G.; Callaghan, P.T. Using Earth's Field NMR to Study Brine Content in Antarctic Sea Ice: Comparison with Salinity and Temperature Estimates. *Appl. Magn. Reson.* **2009**, *36*, 1–8. [CrossRef]
50. Callaghan, P.; Eccles, C. NMR studies on Antarctic sea ice. *Bull. Magn. Reson.* **1996**, *18*, 62–64.
51. Anderson, A.G.; Garwin, R.L.; Hahn, E.L.; Horton, J.W.; Tucker, G.L.; Walker, R.M. Spin echo serial storage memory. *J. Appl. Phys.* **1955**, *26*, 1324–1338. [CrossRef]
52. Hahn, E.L. Nuclear induction due to free larmor precession. *Phys. Rev.* **1950**, *77*, 297–298. [CrossRef]
53. Melton, B.F.; Pollak, V.L.; Mayes, T.W.; Willis, B. Condition for Sudden Passage in the Earth's-Field NMR Technique. *J. Magn. Reson. Ser. A* **1995**, *122*, 164–170. [CrossRef]
54. NOAA National Centers for Environmental Information. Geomagnetic Calculators. 2023. Available online: https://www.ngdc.noaa.gov/geomag/calculators/magcalc.shtml?useFullSite=true (accessed on 18 July 2024).

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.

sensors

MDPI

Article

Handheld Near-Infrared Spectroscopy for Undried Forage Quality Estimation

William Yamada [1,*], Jerry Cherney [2], Debbie Cherney [2], Troy Runge [1] and Matthew Digman [1]

[1] Department of Biological Systems Engineering, University of Wisconsin—Madison, Madison, WI 53706, USA; trunge@wisc.edu (T.R.); digman@wisc.edu (M.D.)
[2] Section of Soil and Crop Sciences, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA; jhc5@cornell.edu (J.C.); djc6@cornell.edu (D.C.)
* Correspondence: wyamada@wisc.edu

Abstract: This study investigates the efficacy of handheld Near-Infrared Spectroscopy (NIRS) devices for in-field estimation of forage quality using undried samples. The objective is to assess the precision and accuracy of multiple handheld NIRS instruments—NeoSpectra, TrinamiX, and AgroCares—when evaluating key forage quality metrics such as Crude Protein (CP), Neutral Detergent Fiber (aNDF), Acid Detergent Fiber (ADF), Acid Detergent Lignin (ADL), in vitro Total Digestibility (IVTD)and Neutral Detergent Fiber Digestibility (NDFD). Samples were collected from silage bunkers across 111 farms in New York State and scanned using different methods (static, moving, and turntable). The results demonstrate that dynamic scanning patterns (moving and turntable) enhance the predictive accuracy of the models compared to static scans. Fiber constituents (ADF, aNDF) and Crude Protein (CP) show higher robustness and minimal impact from water interference, maintaining similar R^2 values as dried samples. Conversely, IVTD, NDFD, and ADL are adversely affected by water content, resulting in lower R^2 values. This study underscores the importance of understanding the water effects on undried forage, as water's high absorption bands at 1400 and 1900 nm introduce significant spectral interference. Further investigation into the PLSR loading factors is necessary to mitigate these effects. The findings suggest that, while handheld NIRS devices hold promise for rapid, on-site forage quality assessment, careful consideration of scanning methodology is crucial for accurate prediction models. This research contributes valuable insights for optimizing the use of portable NIRS technology in forage analysis, enhancing feed utilization efficiency, and supporting sustainable dairy farming practices.

Keywords: portable NIRS; miniaturization; feed composition; scanning methods

1. Introduction

Understanding and managing the nutritional variability of forages is crucial for optimizing dairy farm management and enhancing animal health. The nutritive value of forages, including alfalfa-grass haylage and corn silage, exhibits considerable variability, which can influence milk production efficiency and environmental sustainability [1]. Recent advancements have leveraged handheld near-infrared (NIR) spectroscopy devices, such as the NeoSpectra (1350–2550 nm, Si-ware Systems Inc., Cairo, Egypt), TrinamiX (1450–2450 nm, TrinamiX Inc., Ludwigshafen, Germany), and Agrocares Scanners F-Series (1300–2550 nm, AgroCares, Wageningen, The Netherlands). These tools offer rapid on-site prediction of forage quality, enabling nutritional management by detecting variations in forage nutritive values, including dry matter (DM) [2–5], Crude Protein (CP) [2–10], and fiber contents and properties, such as actual Neutral Detergent Fiber (aNDF) [2,6–10], Neutral Detergent Fiber Digestibility (NDFD) [4,7,8], Acid Detergent Fiber (ADF) [2–4,6–10], Acid Detergent Lignin (ADL) [2,4,7,8], and in vitro Total Digestibility (IVTD) [4,6–8,10]. By facilitating daily adjustments to animal diets based on accurate, real-time forage analysis,

handheld NIR devices can significantly enhance the efficiency of feed utilization, reduce environmental impact, and improve the overall profitability and sustainability of dairy farming operations [3,11,12].

The interest in compact, handheld spectrometers for analyses conducted directly on farms is gaining momentum [2,13,14]. These NIR devices have proven to have good performance on pre-treated samples of dried and ground forage [6]. These portable spectrometers are becoming more common on machinery used for forage harvesting and have also been modified for use with other agricultural implements like liquid manure spreaders [15,16]. In a study [17], it was found that a portable device (HarvestLab^{TM} 3000, Deere & Company, Moline, IL, USA) could approximate the quality of a mix of undried forage species, albeit with systematic errors that could be adjusted for accuracy. An essential element for the effectiveness of NIR technology in these applications is the development of a reliable calibration.

The utilization of handheld Near-Infrared Spectroscopy (NIRS) devices in forage assessment aims to permit direct sample analysis without the need for prior sample processing. This is enabled by calibrations developed for wet, unprocessed forage. However, deploying these devices faces several obstacles, notably the influence of the moisture absorption band and sample heterogeneity [18,19]. Additional operational challenges of handheld NIRS technology include managing the signal-to-noise ratio, ensuring a consistent power supply in a portable format, and maintaining functionality amid diverse and potentially adverse environmental conditions [3,7].

These studies underscore the practical considerations and performance of handheld NIRS instruments in the field. For instance, [7] describes the development of predictive models for various forage types, highlighting the significant influence of the scanning methodology on the accuracy of the constituent prediction. Similarly, ref. [3] provides an assessment of multiple handheld NIR devices, examining their precision and accuracy in on-farm forage evaluation, with a focus on dry matter content compared to traditional moisture meters and the robustness of available calibrations for nutritive value determination.

The ongoing development of NIRS technology for forage quality prediction indicates that miniaturized instruments have similar predictive power as benchtop instruments. As this technology is increasing in popularity, it is essential to understand if different portable instruments and scanning patterns affect the quality of the prediction. Thus, the objectives of this research are as follows:

- To assess the precision and accuracy of multiple handheld Near-Infrared Spectroscopy (NIRS) devices when used for on-farm forage evaluation, particularly focusing on the robustness of calibrations for nutritive value determination;
- To examine if different portable instruments and scanning patterns influence the quality of prediction;
- To evaluate the effects of using dried unground samples for forage quality prediction.

2. Materials and Method

2.1. Samples and Reference Analysis

Predictive NIRS models were developed using NIRS spectra and laboratory reference values for 600 silage samples of mixed haylage. Silage samples were collected between 2021 and 2023 from silage bunkers on 111 farms around New York State. After collection, the samples were vacuum-packed in oxygen-limiting polyethylene bags using a commercial vacuum packing machine for scanning at a later date.

The acquisition of NIR spectroscopic measurement data was achieved using three scanners: NeoSpectra (1350–2550 nm, Si-ware Systems Inc., Cairo, Egypt), TrinamiX (1450–2450 nm, TrinamiX Inc., Ludwigshafen, Germany), and AgroCares (1300–2550 nm, AgroCares, Wageningen, The Netherlands). The data collected with TrinamiX and AgroCares reported spectra from 1454 to 2446 nm at a fixed step of 4 nm, while the NeoSpectra scanner reported spectra from 1350 to 2550 nm at a variable step between 2.5 and 8.8 nm and a wavelength resolution of 16 nm. Each scanner used different detector types as shown in Table 1.

Table 1. Instruments and their characteristics.

Property	AgroCares F-Series	TrinamiX	NEO Spectra
Detector Type	MEMS	Linear Variable Filter	MEMS-FT-NIR
Spectral Range (nm)	1450–2450	1450–2450	1350–2500
Sample Scanning	Contact	Contact	Contact

Before the scanning process began, all samples were thoroughly mixed in a large plastic container to ensure homogeneity. All samples were scanned in a controlled laboratory environment to avoid any interference of humidity or temperature on the scans. Two primary methods were employed to capture spectra, alongside a third specialized technique. The first method involved placing the scanner's lens in direct contact with the sample, where it remained stationary. The second method required the scanner to be moved across the sample surface during the scanning period, maintaining continuous contact. After each scan conducted with the second method, the samples were mixed again to ensure consistency; this method was tested with the AgroCares and NEOSpectra scanners. The third technique utilized a rotating dish accessory (Si-ware Systems Inc., Cairo, Egypt) for the NEOSpectra instrument, allowing the sample to be scanned continuously. These methods were sequentially applied to each forage sample, with five replicate scans being collected to ensure accuracy and repeatability.

Post-acquisition, the samples were desiccated using forced-air ovens until a consistent mass was achieved at 60 °C, subsequently ground with a Wiley mill (Thomas Scientific, Swedesboro, NJ, USA) through a 1 mm mesh screen, and then stored in plastic containers. The forage constituents appraised included Neutral Detergent Fiber (aNDF), in vitro Total Digestibility (IVTD), Neutral Detergent Fiber Digestibility (NDFD), Acid Detergent Lignin (ADL), Acid Detergent Fiber (ADF), ash, and Crude Protein (CP), which served as reference variables for the calibration of Near-Infrared Spectroscopy (NIRS) predictive models.

2.2. Wet Chemistry

For the chemical analysis, the methodologies aligned with those delineated in the literature [20]. Concisely, forage samples were apportioned into ANKOM F57 filter bags (ANKOM Technology, Macedon, New York, NY, USA) to quantify NDF, ADF, ADL, and 48-h IVTD. To alleviate gaseous pressure, the filter bags were intermittently removed from their respective containers on both the initial and subsequent days. The digestibility of the Neutral Detergent Fiber was quantified in terms of the percentage of fiber hydrolyzed, with the values expressed on an NDF basis.

The nitrogen (N) content was measured through a combustion technique using a LECO CN628 analyzer (DairyOne, Ithaca, NY, USA), with Crude Protein (CP) being inferred from nitrogen values using the conversion factor of 6.25 as per AOAC guidelines (1995). Duplicate analyses were performed for all constituents, with nitrogen content being quantified in duplicate on a select sample subset to establish the Standard Error for the CP measurement. The laboratory's Standard Error (SEL) pertinent to these determinations has been documented in prior studies that used the same samples but with different instruments and scanning analysis [3,7].

2.3. Model Calibration

To ensure methodological consistency and mitigate the risk of overfitting, uniform data preprocessing and training protocols were employed across all instrument models. Spectral data from the NEOSpectra device were interpolated to achieve a consistent interval of 4 nm. All spectral data were converted to absorbance by employing the logarithmic transformation of the reciprocal reflectance, denoted as log(1/R). Data preprocessing was standardized using a Savitzky–Golay filter with a window length of four, a polynomial order of three, and a derivation order of one. The algorithm of choice for the modeling was Partial Least Squares Regression (PLSR), using Python 3.10.12 and the packages scipy

(to preprocess the data) and scikit-learn (to calibrate the PLSR model). In order to have a baseline model for each predicted variable, we opted to work with PLS-1 to understand how the scans affected the individual performance of the models. The selection of the optimal number of latent variables within the range of 1 to 20 was systematically determined using a grid search.

The dataset was divided into 90%/10% for training with five-fold cross-validation (CV) and a separate validation dataset, respectively. We randomly selected bunkers to split the dataset, ensuring that the training and test sets were independent. This approach uses 540 samples for calibration (432 for training and 108 for CV) and 60 samples for testing the final model, providing a robust evaluation of the model's generalizability. By having an external dataset for validation, created as described, we can effectively verify the overfitting of our model [21]. Overfitting can be identified by comparing performance metrics between the training and validation datasets. If the model performs significantly better on the training data than on the validation data or the validation performance deteriorates while the calibration improves, it is likely overfitting. In our study, we adopt a robust outlier detection method utilizing Partial Least Squares (PLS) regression tailored for Near-Infrared (NIR) spectroscopy data analysis. This approach leverages Q-residuals and Hotelling's T-squared statistics to identify deviations, ensuring outliers that could skew the model's predictive accuracy are effectively recognized, using a 95% confidence level [22]. This technique provides a systematic way to refine datasets for better analytical outcomes.

2.4. Evaluation

The calibration models were evaluated using the standards set forth by Malley et al. [23] and Williams et al. [24] as shown in Table 2. This section details the predictive performance of the models using a suite of metrics, including root mean square error (RMSE), Bias, Standard Error (SE), Cross-Validation Standard Error (SECV), Coefficient of Determination (R^2), Cross-Validated R^2 (R^2_{CV}), Ratio of Performance to Deviation (RPD), Cross-Validated RPD (RPD_{CV}), and the number of latent variables (LVs) employed on the PLS calibration.

Table 2. Calibration classification and level of success according to [23,24].

Level of Success [23]	**R^2** [23]	**RPD Value** [24]	**Classification** [24]	**Application** [24]
Not useful	<0.80	<2.0	Very poor	Not recommended
Moderately Successful	0.80 to 0.90	2.0 to 2.5	Poor	Rough screening
		2.5 to 3.0	Fair	Screening
Successful	0.90 to 0.95	3.0 to 3.5	Good	Quality control
		3.5 to 4.0	Very good	Process control
Excelent	>0.95	>4.0	Excellent	Any application

RPD—Ratio of Perfomance to Deviation.

We will compare our results with those obtained from various handheld NIRS devices reported in the literature. Specifically, we will reference studies that utilized different instruments on both dried and undried materials. For undried and unground material, we will compare our findings with those from the Aurora device as reported by Cherney et al. (2021) [3] and the NEOSpectra device as reported by Feng et al. (2023) [7]. For dried and ground material, comparisons will be made with results obtained using the MicroP-HAZIR (1600–2400 nm, ThermoFisher Scientific, Waltham, MA, USA) and DLP NIRscan Nano EVM (900–1700 nm, Texas Instruments, Dallas, TX, USA) devices as described by Acosta et al. (2020) [6], the NEOSpectra device as reported by Digman et al. (2022) [4], the ASD QualitySpec (350–2500 nm, Malvern Panalytical, Cambridge, UK) and Tellspec (900–1700 nm, Tellspec Inc., Toronto, ON, Canada) devices as detailed by Rukundo et al. (2021) [2], and the Aurora (950–1650 nm, GraiNit S.r.l., Padua, Italy), NIR-S-G1 (950–1650 nm, InnoSpectra, Hsinchu, Taiwan), and SCiO (740–1070 nm, Consumer Physics, Hod Hasharon, Israel) devices as discussed by Berzaghi et al. (2021) [8]. This comprehensive comparison will provide a robust context for evaluating the performance and accuracy of our results and evaluating the impact of using undried materials. In addition, to evaluate

the effects of water absorption, we will analyze the major PLS loading components in relation to the water absorption bands as described by Williams [25]. The primary water absorption bands are detailed in Table 3.

Table 3. Positions of main absorption bands in water.

Wavelength (nm)	1460	1778	1904	2208	2384
Relative Intensity	Large	Very Small	Very Large	Very Small	Very Small

3. Results and Discussion

3.1. Spectral Data

The average spectra and range of the 600 samples scanned are shown in Figure 1, which presents the mean spectral signatures captured by each instrument. In accordance with the observations reported by Feng et al. [7], overtone bands attributable to O-H bonds are discernible at approximately 1400 and 1900 nm, which is consistent with the presence of moisture in the undried forage samples. The statistical information of the laboratory measurements of the constituents is provided in Table 4.

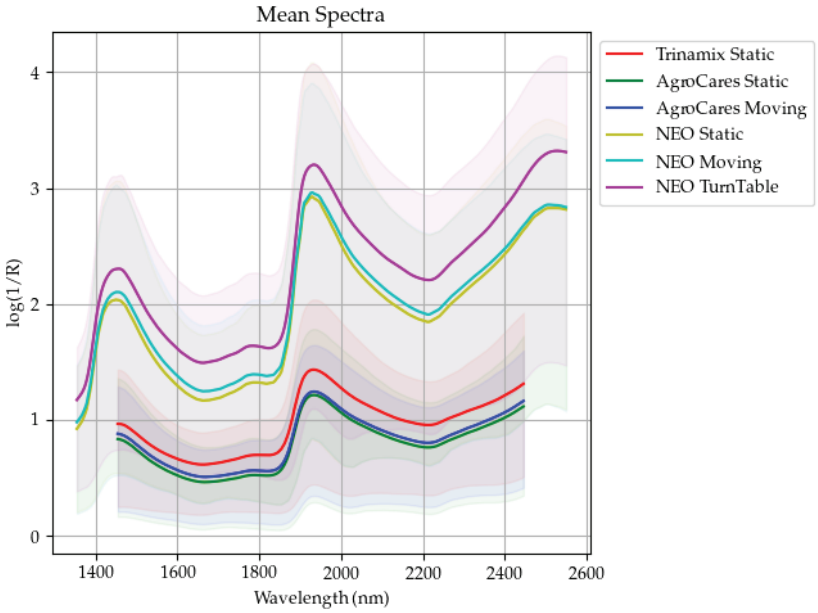


Figure 1. Comparative analysis of forage sample spectra: This graph illustrates the mean spectral signatures of forage samples (n = 600) as measured by three different scanners—TrinamiX (red line—static scan), AgroCares (green line—static scan; blue line—moving scan), and NEOSpectra (yellow—static scan; cyan—moving scan; magenta—turntable scan)—utilizing varying methods. Each line represents the average log(1/R) value across a range of wavelengths from 1400 to 2600 nm. The hue of each line represents the range between the maximum and minimum measured for each instrument.

Table 4. Laboratory reference values statistics.

	IVTD	aNDF	NDFD	ADF	ADL	CP
Unit			%DM			
Count			600			
Mean	79.22	50.13	58.84	37.18	7.35	17.62
SD	7.31	10.42	9.39	5.80	2.18	4.43
Min	38.13	28.81	11.40	24.22	3.12	6.12
Median	80.73	48.75	58.50	36.50	7.06	18.04
Max	92.92	81.60	80.87	59.06	20.60	27.72

IVTD—in vitro Total Digestibility, aNDF—actual Neutral Detergent Fiber, NDFD—Neutral Detergent Fiber Digestibility, ADF—Acid Detergent Fiber, ADL—Acid Detergent Lignin, CP—Crude Protein, DM—Dry Matter, SD—Standard Deviation.

3.2. Calibration Results

The calibration outcomes are summarized in Table 5. The NEOSpectra device in turntable mode yielded superior fit models for the calibration set variables ADF, ADL, CP, and aNDF. With respect to IVTD and NDFD, this device also demonstrated superior performance in certain metrics while remaining competitive in others. An observation is that both the moving and turntable scanning modes achieved the best calibration results when utilizing a greater number of latent variables, suggesting that dynamic scanning captures more relevant data for model calibration.

Table 5. Statistical performance metrics for calibration models using 540 samples undried and unground alfalfa samples across different instruments and modes. The table lists the RMSE, Bias, SE, SE_{CV}, R^2, R^2_{CV}, RPD, RPD_{CV}, and the number of latent variables (LVs) for the variables ADF, ADL, CP, IVTD, and NDFD.

Instrument	Mode	Variable	RMSE	Bias	SE	SE_{CV}	R^2	R^2_{CV}	RPD	RPD_{CV}	LVs
AgroCares	Static	ADF	2.754	0.000	2.756	3.379	0.771	0.655	2.090	1.703	10
	Moving	ADF	1.959	0.000	1.961	2.608	0.884	0.795	2.937	2.208	15
NEOSpectra	Static	ADF	2.463	0.000	2.465	2.885	0.817	0.749	2.336	1.996	12
	Moving	ADF	2.122	0.000	2.124	2.544	0.864	0.805	2.711	2.264	20
	Turntable	ADF	**1.861**	0.000	**1.862**	**2.198**	**0.895**	**0.854**	**3.093**	**2.620**	19
Trinamix	Static	ADF	2.261	0.000	2.263	2.662	0.846	0.786	2.545	2.163	13
AgroCares	Static	ADL	1.365	0.000	1.367	1.674	0.591	0.386	1.564	1.277	11
	Moving	ADL	1.293	0.000	1.294	1.559	0.633	0.468	1.651	1.371	10
NEOSpectra	Static	ADL	1.369	0.000	1.371	1.592	0.589	0.445	1.559	1.342	11
	Moving	ADL	1.242	0.000	1.244	1.460	0.661	0.533	1.718	1.464	18
	Turntable	ADL	**1.175**	0.000	**1.176**	**1.450**	**0.697**	**0.539**	**1.817**	**1.473**	20
Trinamix	Static	ADL	1.405	0.000	1.406	1.608	0.567	0.434	1.520	1.329	10
AgroCares	Static	CP	2.010	0.000	2.012	2.306	0.792	0.727	2.195	1.915	9
	Moving	CP	1.670	0.000	1.672	1.872	0.857	0.820	2.641	2.358	8
NEOSpectra	Static	CP	1.843	0.000	1.845	2.140	0.825	0.765	2.393	2.063	11
	Moving	CP	1.513	0.000	1.514	1.799	0.882	0.834	2.916	2.454	20
	Turntable	CP	**1.328**	0.000	**1.329**	**1.601**	**0.909**	**0.869**	**3.322**	**2.758**	20
Trinamix	Static	CP	1.643	0.000	1.645	1.900	0.861	0.815	2.684	2.324	12
AgroCares	Static	IVTD	4.279	0.000	4.283	5.016	0.660	0.533	1.714	1.463	9
	Moving	IVTD	3.465	0.000	3.469	**4.177**	0.777	**0.677**	2.117	**1.758**	10
NEOSpectra	Static	IVTD	4.114	0.000	4.118	4.632	0.686	0.602	1.783	1.585	10
	Moving	IVTD	3.665	0.000	3.668	4.455	0.750	0.632	2.002	1.648	19
	Turntable	IVTD	**3.409**	0.000	**3.412**	4.195	**0.784**	0.674	**2.152**	1.750	20
Trinamix	Static	IVTD	3.757	0.000	3.760	4.372	0.738	0.645	1.953	1.679	12
AgroCares	Static	NDFD	6.534	0.000	6.540	7.744	0.524	0.333	1.450	1.224	10
	Moving	NDFD	5.653	0.000	5.658	6.862	0.644	0.476	1.676	1.382	10
NEOSpectra	Static	NDFD	6.317	0.000	6.323	7.173	0.555	0.428	1.500	1.322	11
	Moving	NDFD	5.730	0.000	5.735	6.980	0.634	0.458	1.653	1.358	19
	Turntable	NDFD	5.485	0.000	5.490	**6.508**	0.665	**0.529**	1.727	**1.457**	19
Trinamix	Static	NDFD	**5.390**	0.000	**5.395**	7.081	**0.676**	0.442	**1.757**	1.339	20
AgroCares	Static	aNDF	3.811	0.000	3.814	4.370	0.863	0.820	2.700	2.356	10
	Moving	aNDF	3.256	0.000	3.259	3.829	0.900	0.862	3.159	2.689	8
NEOSpectra	Static	aNDF	3.243	0.000	3.246	3.752	0.901	0.867	3.172	2.744	15
	Moving	aNDF	3.007	0.000	3.010	3.573	0.915	0.880	3.421	2.881	20
	Turntable	aNDF	**2.605**	0.000	**2.608**	**3.031**	**0.936**	**0.913**	**3.949**	**3.397**	20
Trinamix	Static	aNDF	2.905	0.000	2.908	3.828	0.920	0.862	3.541	2.689	20

SE—Standard Error, RPD—Ratio of Performance to Deviation, CV—Cross-Validated, LVs—Latent Variables, ADF—Acid Detergent Fiber, ADL—Acid Detergent Lignin, CP—Crude Protein, IVTD—in vitro Total Digestibility, NDFD—Neutral Detergent Fiber Digestibility, aNDF—actual Neutral Detergent Fiber.

Figure 2 illustrates the relationship between the number of latent variables (LVs) and the root mean square error (RMSE) for various instruments and target variables. The analysis reveals that dynamic scans, represented by moving and turntable configurations, exhibit less sensitivity to larger number of latent variables. In contrast, static scans demonstrate signs of overfitting when more than 10 latent variables are used as indicated by the deterioration in CV performance, evidenced by increasing the RMSE, despite improvements in the calibration RMSE. This observation underscores the importance of scan dynamics in mitigating overfitting and enhancing the predictive accuracy of models across different latent variable configurations. The explained variability of the LVs can be found in Table A1 in Appendix A. Overall, the improvements using more latent variables are not that significant for more than 10 LVs, achieving less than one percent improvement in the RMSE.

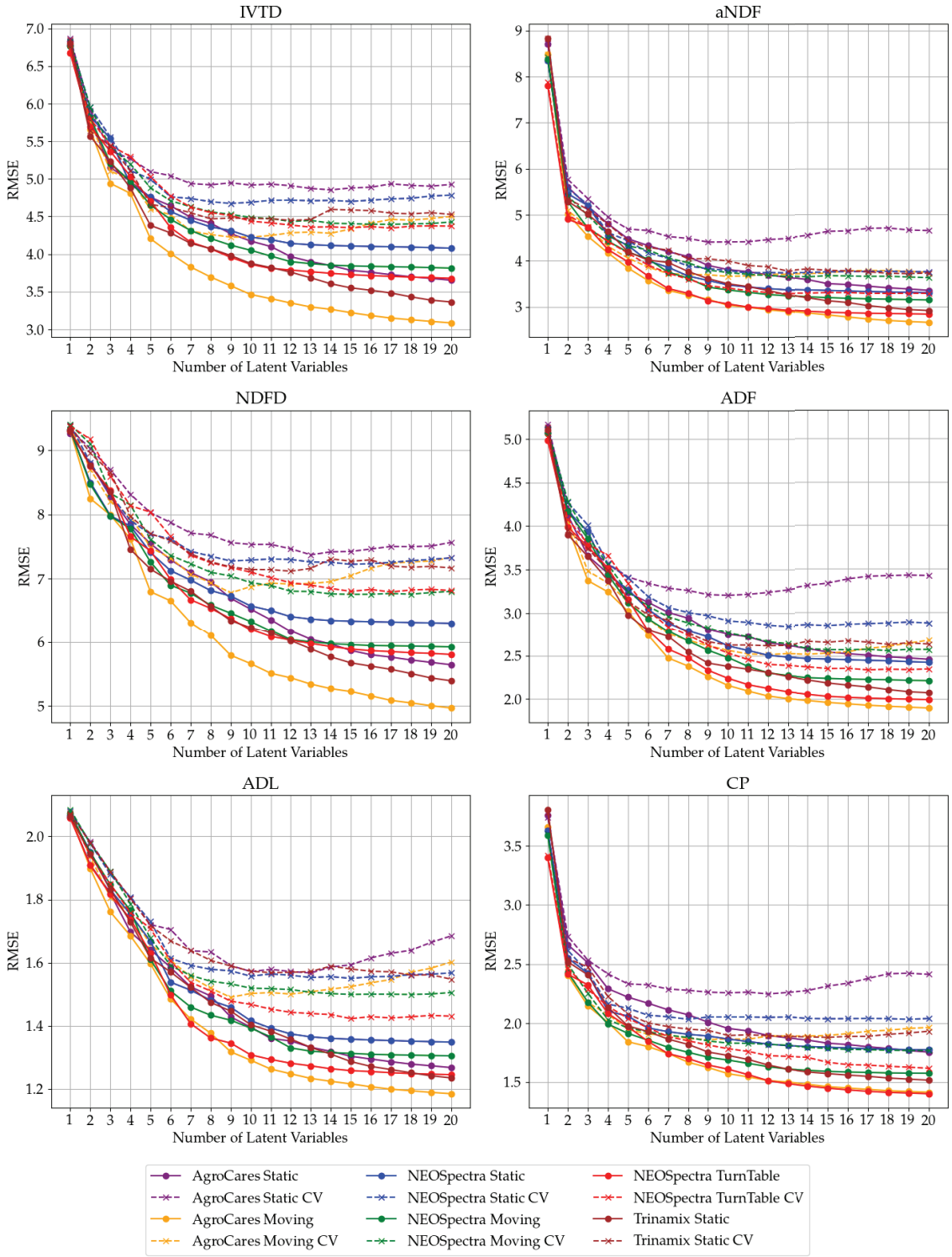


Figure 2. RMSE vs. latent variables for each variable. This figure shows the root mean squared error (RMSE) values for different numbers of latent variables across various instruments and target variables. The RMSE values for both calibration and cross-validation (CV) are plotted for each instrument, differentiated by color (purple—AgroCares Static, orange—Agrocares Moving, blue—NEOSpectra Static, green—NEOSpectra Moving, red—NEOSpectra Turntable, brown—Trinamix Static) and line style (continuous—calibration, dashed—CV).

3.3. Validation Results

The validation results are reported in Table 6 and summarized in Table 7, the performance of the NEOSpectra and Trinamix instruments across various modes and variables is quantified through metrics such as RMSE, SE, R^2, and RPD. For the NEOSpectra instrument, when operating in 'Moving' mode, CP predictions were moderately successful ($R^2 = 0.892$) with a corresponding RPD of 3.042, leading to a 'Good' classification. However, the same instrument's performance predicting IVTD in the same mode was not useful, with a lower R^2 of 0.743 and an RPD of 1.974, reflecting a 'Very poor' classification. When utilizing the 'Turntable' mode, predictions of ADF and ADL yielded 'Fair' and 'Very poor' classifications, respectively, indicating varied efficacy based on the forage constituent analyzed. In contrast, the Trinamix instrument in 'Static' mode demonstrated 'Successful' prediction for aNDF with an R^2 of 0.916 and an RPD of 3.452, garnering a 'Very good' classification. These results reflect the nuanced capabilities of each instrument and mode combination, emphasizing the importance of selecting the appropriate setup for specific analytical needs in forage assessment. Figure 3 makes it clear that the calibrated models did not perform well in predicting NDFD, ADL, and IVTD on the validation set for undried haylage samples.

Table 6. Validation performance metrics for different instruments operating in static, moving, and turntable modes. The metrics include RMSE, Bias, SE, R^2, Slope, Intercept, and RPD for the validation of variables ADF, ADL, CP, IVTD, NDFD, and aNDF using a set of 60 samples. This table facilitates the comparison of model precision and prediction accuracy across diverse instruments and scanning configurations for the validation dataset.

Instrument	Mode	Variable	RMSE	Bias	SE	R^2	Slope	Intercept	RPD
AgroCares	Static	ADF	2.949	−0.150	2.970	0.761	0.973	1.191	2.047
	Moving	ADF	3.015	−0.654	2.968	0.751	0.928	3.386	2.003
NEOSpectra	Static	ADF	2.490	−0.203	2.502	0.830	1.026	−0.805	2.425
	Moving	ADF	2.283	−0.446	2.258	0.857	0.923	3.374	2.645
	Turntable	ADF	**2.207**	−0.193	**2.217**	**0.866**	0.946	2.269	**2.736**
Trinamix	Static	ADF	2.536	−0.332	2.536	0.824	1.090	−3.124	2.381
AgroCares	Static	ADL	2.400	−0.040	2.420	0.109	0.644	2.705	1.059
	Moving	ADL	2.247	−0.251	2.252	0.219	0.835	1.450	1.132
NEOSpectra	Static	ADL	2.050	−0.079	2.066	0.350	0.963	0.356	1.240
	Moving	ADL	2.013	−0.321	2.004	0.373	1.018	0.194	1.263
	Turntable	ADL	**1.794**	−0.178	**1.800**	**0.502**	1.040	-0.115	**1.417**
Trinamix	Static	ADL	1.961	−0.307	1.953	0.405	1.257	−1.550	1.297
AgroCares	Static	CP	2.003	0.356	1.988	0.783	0.893	1.408	2.144
	Moving	CP	1.729	0.329	1.712	0.838	0.937	0.700	2.484
NEOSpectra	Static	CP	1.977	0.333	1.965	0.788	0.997	−0.288	2.172
	Moving	CP	**1.412**	0.074	**1.422**	**0.892**	0.986	0.150	**3.042**
	Turntable	CP	1.517	0.140	1.524	0.875	0.935	0.911	2.831
Trinamix	Static	CP	1.712	0.101	1.723	0.841	0.925	1.112	2.509
AgroCares	Static	IVTD	4.558	−0.709	4.540	0.577	0.862	11.412	1.537
	Moving	IVTD	4.141	0.010	4.176	0.651	0.857	11.125	1.692
NEOSpectra	Static	IVTD	4.040	−0.315	4.061	0.668	0.911	7.270	1.735
	Moving	IVTD	**3.549**	−0.262	3.570	**0.743**	0.927	5.942	**1.974**
	Turntable	IVTD	3.550	−0.277	**3.569**	0.743	0.854	11.650	1.974
Trinamix	Static	IVTD	4.142	−0.180	4.173	0.651	0.912	7.068	1.692
AgroCares	Static	NDFD	8.355	−1.046	8.360	−0.019	0.497	30.314	0.991
	Moving	NDFD	7.406	−0.232	7.465	0.200	0.645	21.191	1.118
NEOSpectra	Static	NDFD	6.515	−0.590	6.543	0.381	0.838	10.082	1.271
	Moving	NDFD	5.943	−0.164	5.991	0.485	0.952	3.016	1.393
	Turntable	NDFD	**5.544**	−0.348	**5.579**	**0.552**	0.816	11.202	**1.493**
Trinamix	Static	NDFD	7.295	−0.515	7.338	0.223	0.649	21.140	1.135

Table 6. *Cont.*

Instrument	Mode	Variable	RMSE	Bias	SE	R^2	Slope	Intercept	RPD
AgroCares	Static	aNDF	4.380	−0.501	4.388	0.841	1.018	−0.462	2.506
	Moving	aNDF	3.808	−0.865	3.739	0.880	1.020	−0.186	2.883
NEOSpectra	Static	aNDF	3.832	−0.713	3.797	0.878	1.050	−1.935	2.864
	Moving	aNDF	3.494	−0.174	3.519	0.899	0.991	0.666	3.141
	Turntable	aNDF	3.304	−0.404	3.307	0.909	0.991	0.872	3.322
Trinamix	Static	aNDF	**3.180**	0.189	**3.201**	**0.916**	1.010	−0.705	**3.452**

SE—Standard Error, RPD—Ratio of Performance to Deviation, ADF—Acid Detergent Fiber, ADL—Acid Detergent Lignin, CP—Crude Protein, IVTD—in vitro Total Digestibility, NDFD—Neutral Detergent Fiber Digestibility, aNDF—actual Neutral Detergent Fiber.

Table 7. Performance of the best model for each predicted variable on the validation set.

Instrument	Mode	Variable	RMSE	SE	R^2	RPD	Success (R^2 [23])	Classification (RPD [24])
NEOSpectra	Moving	CP	1.412	1.422	0.892	3.042	Moderately Successful	Good
	Moving	IVTD	3.549	3.570	0.743	1.974	Not Useful	Very poor
	Turntable	ADF	2.207	2.217	0.866	2.736	Moderately Successful	Fair
	Turntable	ADL	1.794	1.800	0.502	1.417	Not Useful	Very poor
	Turntable	NDFD	5.544	5.579	0.552	1.493	Not Useful	Very poor
Trinamix	Static	aNDF	3.180	3.201	0.916	3.452	Successful	Very good

RMSE—Root Mean Squared Error, SE—Standard Error, RPD—Ratio of Performance to Deviation, ADF—Acid Detergent Fiber, ADL—Acid Detergent Lignin, CP—Crude Protein, IVTD—in vitro Total Digestibility, NDFD—Neutral Detergent Fiber Digestibility, aNDF—actual Neutral Detergent Fiber.

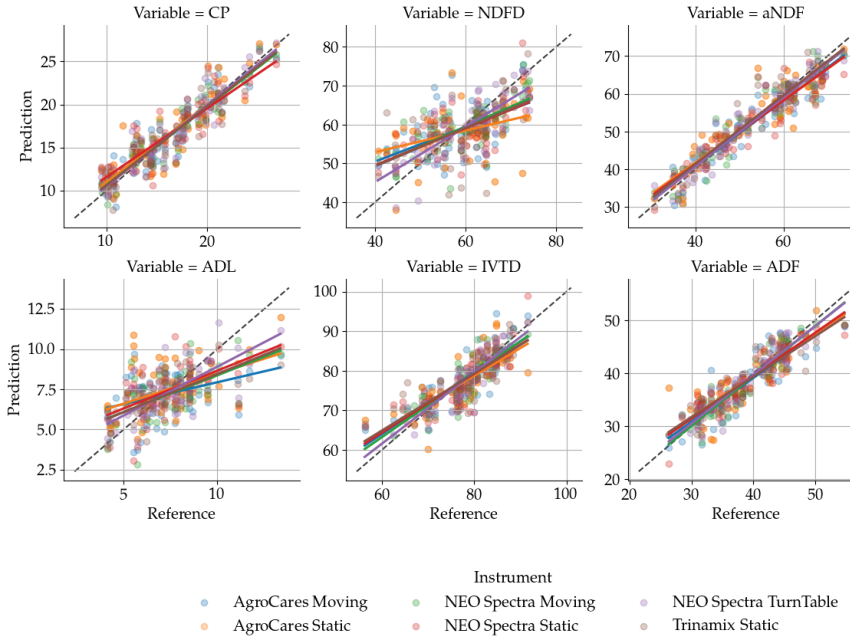


Figure 3. Comparative evaluation of three handheld spectrometers and methods used to predict nutritional content in feed samples. The different colors and shapes represent readings from moving, static, or turntable methods of using the AgroCares, NEO Spectra, and TrinamiX instruments. Each dot represents the pair of reference data and the prediction using the calibrated PLSR model from the validation set (n = 60). The regression lines for each method showcase the accuracy and precision in predicting the content of Crude Protein (CP), Neutral Detergent Fiber (aNDF), Acid Detergent Fiber (ADF), Acid Detergent Lignin (ADL), Neutral Detergent Fiber Digestibility (NDFD), and in vitro Total Digestibility (IVTD). The dashed black line represents a 1:1 agreement between the reference and predicted values.

When calibrated exclusively with static scans, TrinamiX had a better RMSE, SE, and RPD than the NEOSpectra for predicting CP and ADL. These findings concur with the insights of Feng et al. [7], highlighting the enhanced spectral quality afforded by moving scans due to their capacity to encapsulate a more generalized and homogeneous representation of the samples. A closer examination of the calibration data delineated in Table 5 reveals a performance hierarchy within the same instrument, with the order of efficacy being turntable > moving > static. This sequence also correlates with the increasing number of latent variables that can be utilized in the NEOSpectra scanning process, thereby suggesting that sliding scans not only improve the spectral quality but also allow for a better model calibration.

Utilizing the same scanning pattern—whether static, moving, or turntable—tends to yield comparable calibration performance across different devices as evidenced by the minimal variation in the R^2 and the RPD values, usually within the same class range of success, according to Table 2. This is consistent across most variables, with the notable exception of Neutral Detergent Fiber (aNDF), where the Trinamix instrument in static mode achieved an RPD of 3.453, surpassing those of AgroCares at 2.506 and NEOSpectra at 2.864. The findings thus suggest that the methodology of spectral data acquisition is more important than the choice of the handheld instrument.

Figure 4 presents normalized boxplots of prediction errors, facilitating a more nuanced comparison of calibration performance. Notably, CP, ADF, and aNDF demonstrate the most favorable results, characterized by minimal bias and RMSE, with most prediction errors falling within one standard deviation. Conversely, ADL predictions are less accurate, exhibiting a multitude of outliers as reflected in Table 7, indicating a disparity between the predicted and observed values.

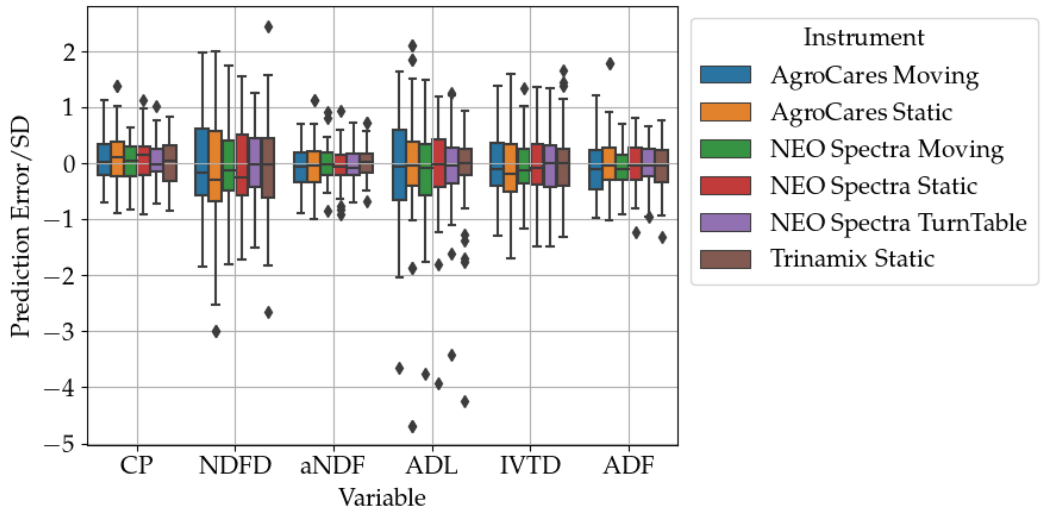


Figure 4. The chart presents the normalized distribution of prediction errors on the validation set for six forage quality variables—CP, NDFD, aNDF, ADL, IVTD, and ADF—obtained using different spectral scanning instruments and methods. Each boxplot shows the median, quartiles, and outliers for the prediction error standard deviation (SD) of each method.

The standard error of laboratory (SEL) values reported by Cherney et al. (2021) [3] for the wet chemistry of the same samples indicated errors that are an order of magnitude lower than the root mean square error (RMSE) values from our results presented in Table 7 ($SEL_{aNDF} \leq 0.66$, $SEL_{ADF} \leq 0.70$, $SEL_{ADL} \leq 0.30$, $SEL_{IVTD} \leq 0.76$, $SEL_{NDFD} \leq 2.36$, and $SEL_{CP} \leq 0.44$).

Compared to the existing literature, our calibration results are compatible with the previously reported findings. For undried samples, [3] evaluated the Aurora instrument for haylage, corn silage, and Total Mixed Ration, while [7] assessed the NEOSpectra for corn silage, alfalfa, grass, and mixed alfalfa and grass silage. Both studies utilized moving scans. As shown in Figure 5, our results with haylage samples exhibit similar characteristics to

the best metrics from the literature, further supporting the notion that dynamic scans often outperform other methods. Additionally, our findings indicate that ADF, aNDF, and CP achieve moderately successful-to-excellent calibrations, whereas IVTD, NDFD, and ADL do not yield useful models according to the [23] R^2 criteria.

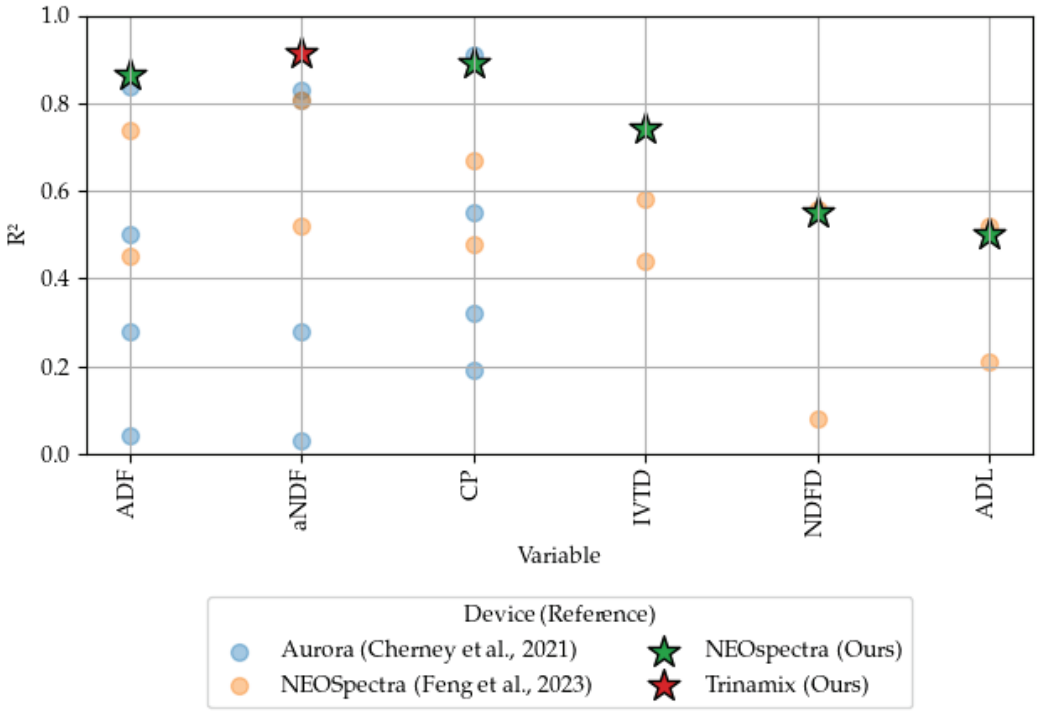


Figure 5. Undried data R^2 comparison with Aurora [3], calibrated for haylage, corn silage, and Total Mixed Ration, and NEOSpectra [7] calibrated for grass, alfalfa, and mixed silage forages. Both references were sampled using moving scans. The dots represents the metrics obtained by the references, and the stars represents the metrics obtained by our best model.

As one of the goals of portable NIRS is to be used for in-field forage quality estimation, it is important to understand how the prediction model performance is affected when using undried samples. Since water has high absorption bands at 1400 and 1900 nm, it creates interference in the raw spectrum of the material in the NIR region [25]. Figure 6 illustrates how our model performance compares to that of models calibrated on dried samples. It is evident that the water content of the samples affects the performance of the models to varying degrees for all forage quality metrics studied. ADF, aNDF, and CP are less impacted by water, exhibiting similar R^2 values to those of dried materials. In contrast, IVTD, NDFD, and ADL are severely affected by the water content, resulting in lower R^2 values. Further investigation into how water influences the PLSR loading factors is necessary to better understand whether these effects can be mitigated.

Figures 7 and 8 display the first two latent variables that contain most of the explained variance of the PLS models (Appendix A). It is evident that the main absorption bands of water play a role in the loading factors of the latent variables, particularly at the 1904 nm band, where water has a significant absorption peak. These effects arise from the interactions of water with the O-H groups present in carbohydrates, fats, and proteins, which can form hydrogen bonds with most types of fiber. These results are consistent with the findings obtained from studies on small grains [25].

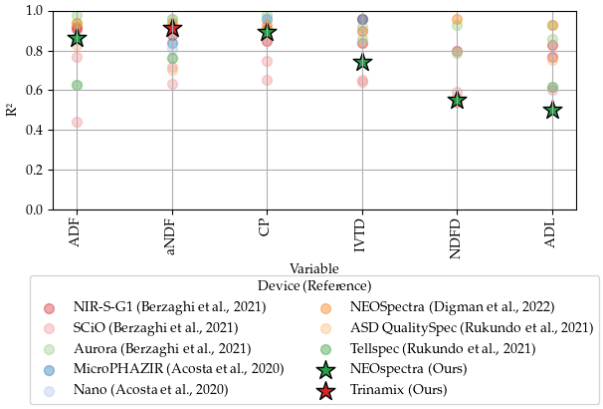


Figure 6. Comparison of R^2 values from models calibrated on dried samples (literature) versus our model calibrated on undried samples. Tellspec and ASD QualitySpec [2] were calibrated for grass. NEOSpectra [4] was calibrated for grass, alfalfa, and mixed silage forages. Nano and MicroPHAZIR [6] were calibrated for grass forages. NIR-S-G1, SCiO, and Aurora [8] were calibrated for alfalfa and grass forages. The dots represent the metrics obtained from the references, and the stars represent the metrics obtained by our best model.

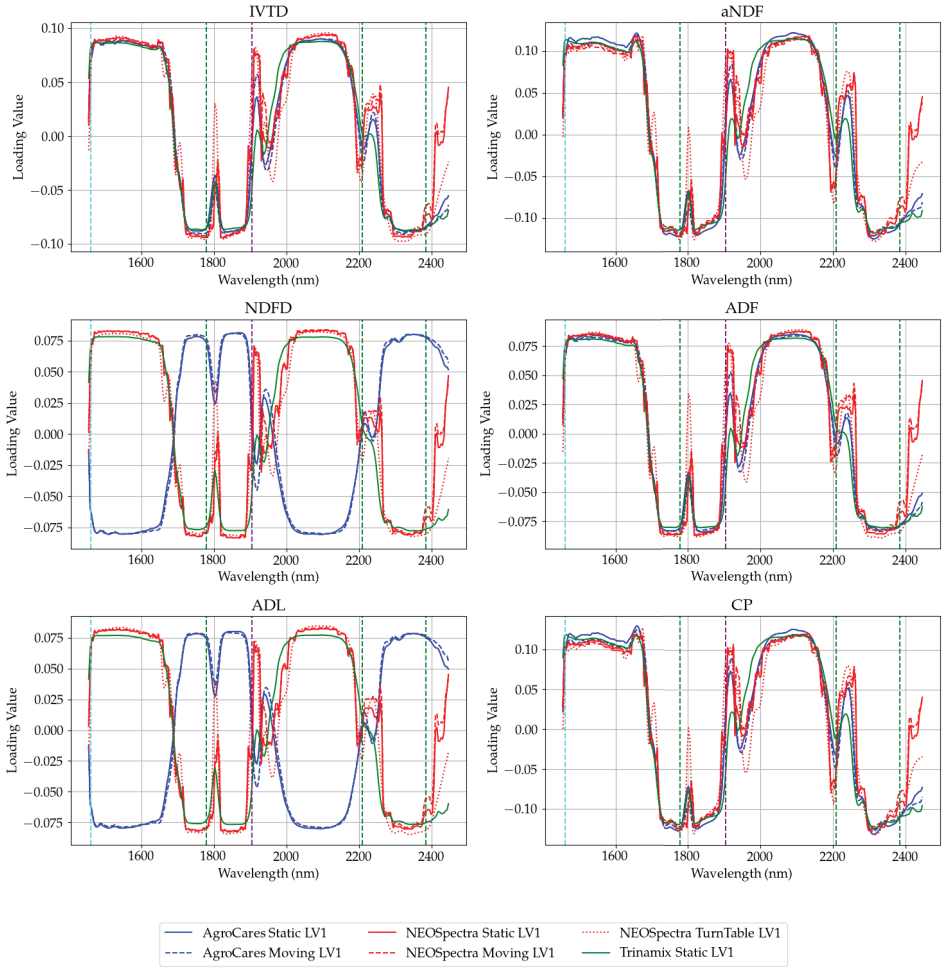


Figure 7. Loading values of the first latent variable of the spectrum (first derivative). Instruments are divided by color (AgroCares—blue, NEOSpectra—red, and Trinamix—green). The scan modes are divided by the line style (continuous—static, dashed—moving, and dotted—turntable). The vertical lines are the water absorption bands. Very small absorption bands (1778, 2208, and 2384 nm) are in green. The large absorption band (1460 nm) is illustrated in cyan. The very large absorption band (1904 nm) is shown in purple.

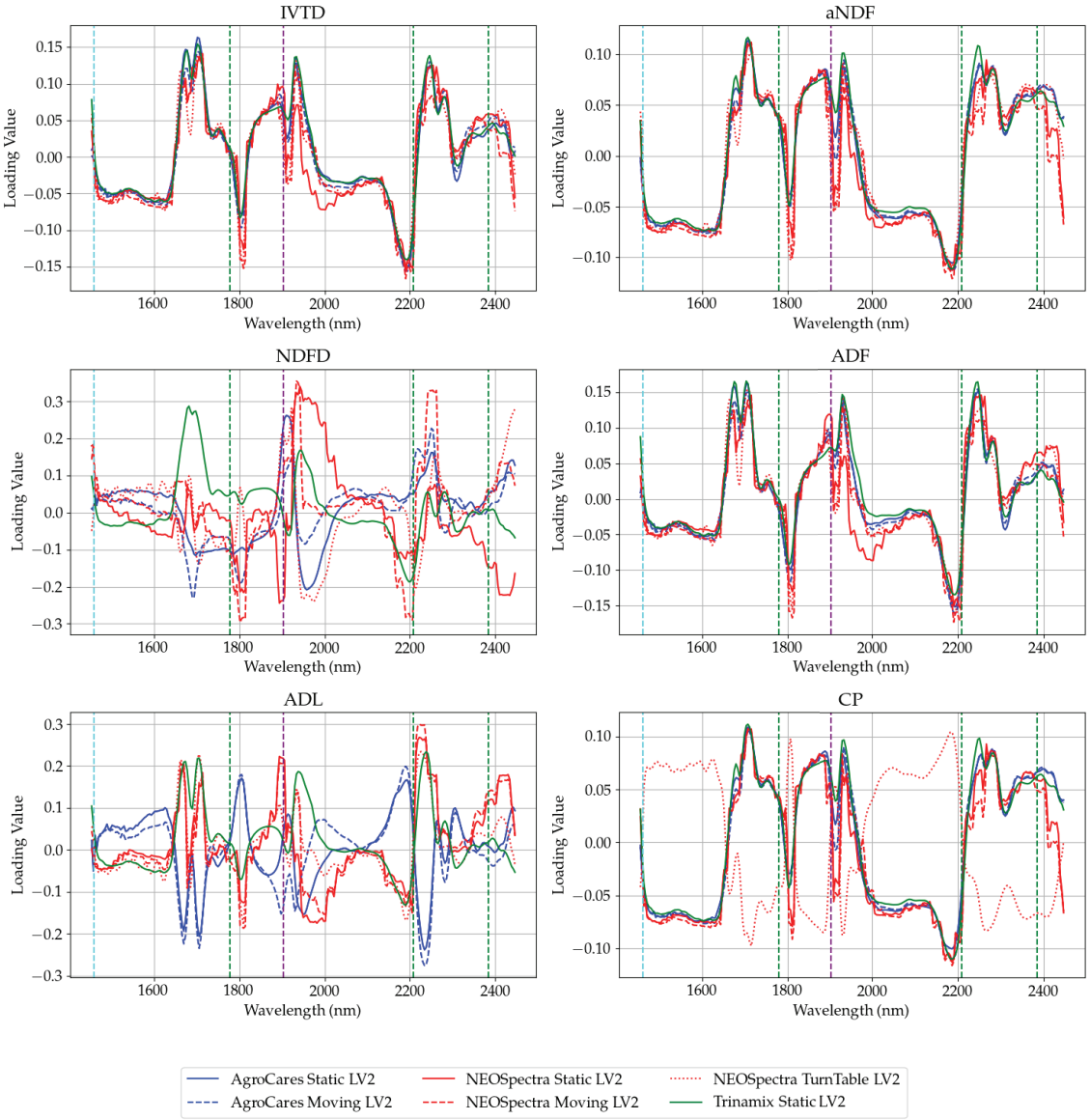


Figure 8. Loading values of the second latent variable of the spectrum (first derivative). Instruments are divided by color (AgroCares—blue, NEOSpectra—red, and Trinamix—green). The scan modes are divided by the line style (continuous—static, dashed—moving, and dotted—turntable). The vertical lines are the water absorption bands. Very small absorption bands (1778, 2208, and 2384 nm) are in green. The large absorption band (1460 nm) is illustrated in cyan. The very large absorption band (1904 nm) is in purple.

Collectively, these analyses underscore the significance of the scanning pattern over the specific technology or instrument used. The consistency in data acquisition methodology emerges as a critical factor in the calibration performance, influencing the robustness of predictive models more substantially than the hardware utilized. Furthermore, the lower performance of certain variables is likely due to the water's electromagnetic absorption and interaction with the undried sample material.

4. Conclusions

This study systematically explores the influence of the scanning methodology and instrument design on the efficacy of spectroscopic models in forage analysis. Our findings are derived from a set of 600 ensiled forage samples collected across New York state. As detailed in Tables 5 and 6, and summarized in Table 7, we highlight the importance

of the role the spectral acquisition technique plays over the specific technical features of handheld NIRS devices.

The consistency observed across the instruments when identical scanning patterns were employed underscores the methodological influence over technology. Specifically, the NEOSpectra instrument, when employed in a dynamic mode, demonstrates a significant advantage in the predictive accuracy for all variables. This suggests that the precision and reliability of predictions are more heavily contingent upon applying robust and consistent scanning protocols.

Based on the analysis of Figure 2 and Table A1, we recommend limiting the number of latent variables to 7–10 to avoid overfitting and ensure future model performance. Although our results indicate that 11–20 LVs can have a small improvement in the unseen validation set, the explained variance of these variables does not significantly improve the results enough to justify using them. Therefore, a careful balance must be struck between model complexity and predictive stability.

Comparative analyses, particularly for undried forage, have aligned with the findings from the literature, confirming the validity of our models within the expected performance parameters. Moreover, the results have revealed that scanning modes incorporating movement tend to enhance the homogeneity of the sample representation, which is critical in achieving high-quality spectral data. When comparing dried and grounded material calibrations, we can see that fiber constituents and protein are less impacted by water absorption. However, there remains a knowledge gap in understanding the water interactions of the undried and unground forage constituents, specifically how water interaction affects their NIR spectral characteristics. The loadings obtained through Partial Least Squares Regression of the NIR spectra highlight the critical role of variance at the wavelengths associated with O-H absorptions in constructing models for these materials. The behavior of water within complex agricultural substances is expected to differ from that of liquid water.

Figures 3 and 4 provide visual confirmation of the comparative and error distribution analyses, respectively, illustrating the nuanced performance across different forage constituents and underscoring the models that exhibit both high accuracy and those with room for improvement.

In summary, this research affirms the importance of the scanning pattern in developing robust near-infrared spectroscopic models. It contributes valuable insights that may guide practitioners in selecting the most suitable instruments and modes for forage quality assessment. As the field advances, future studies should further refine these methodologies, optimizing the balance between technological innovation and practical application for enhanced forage analysis. The evaluation of embedded NIR sensors in agricultural machinery to predict forage quality and properties is one of the paths forward in undried forage research.

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

Funding: This work was supported in part by the USDA National Institute of Food and Agriculture, AFRP project 1027162.

Institutional Review Board Statement: Not applicable for studies not involving humans or animals.

Informed Consent Statement: Not applicable for studies not involving humans or animals.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

Appendix A. Explained Variance

Table A1. Explained variance (%) by each latent variable (LV) for different instruments and target variables. The values represent the individual contribution of each LV to the total explained variance.

Instrument	Variable	LV1	LV2	LV3	LV4	LV5	LV6	LV7	LV8	LV9	LV10
AgroCares Static	ADF	69.64	9.99	1.89	5.29	2.66	2.74	1.73	2.22	0.50	0.80
	ADL	75.36	3.10	3.93	1.58	4.79	2.72	1.73	3.06	0.49	0.36
	CP	34.57	40.04	3.19	4.97	5.35	2.54	2.17	1.89	0.99	0.91
	IVTD	66.68	12.45	1.59	4.12	3.94	3.21	1.77	1.49	1.28	0.83
	NDFD	75.40	4.20	3.90	1.83	2.45	2.76	2.93	2.62	0.80	0.68
	aNDF	38.45	35.27	3.75	5.49	4.13	3.71	2.43	2.03	0.56	0.75
AgroCares Moving	ADF	72.73	11.85	1.76	5.23	2.02	1.75	0.52	1.52	0.26	0.26
	ADL	79.69	3.66	4.60	1.87	2.07	3.65	0.94	1.01	0.17	0.29
	CP	37.47	42.25	2.13	7.73	1.68	3.31	1.06	1.09	0.51	0.21
	IVTD	67.76	15.55	1.78	6.10	1.02	1.57	1.75	1.66	0.29	0.25
	NDFD	78.51	2.07	4.56	2.58	1.23	6.11	1.23	1.18	0.18	0.33
	aNDF	42.04	38.25	2.75	5.86	3.70	1.75	0.70	1.75	0.49	0.21
NEOSpectra Static	ADF	67.24	13.98	7.18	3.96	2.41	0.70	1.23	0.42	0.53	0.32
	ADL	74.43	5.09	6.78	5.11	3.41	0.64	1.10	0.63	0.44	0.31
	CP	36.26	42.49	6.67	5.75	1.79	1.20	0.63	1.28	1.00	0.45
	IVTD	62.96	17.68	5.75	5.34	1.61	1.39	1.68	0.50	0.59	0.34
	NDFD	73.31	2.32	9.77	6.12	1.84	1.48	1.91	0.47	0.43	0.35
	aNDF	36.99	38.83	10.39	4.16	3.23	0.52	1.70	0.45	0.90	0.40
NEOSpectra Moving	ADF	67.69	12.72	4.56	7.72	1.77	1.13	1.12	0.46	0.55	0.28
	ADL	73.90	4.19	8.23	5.41	2.66	0.82	0.45	0.84	1.19	0.28
	CP	35.14	39.27	3.32	13.64	2.10	0.80	1.78	0.98	0.46	0.40
	IVTD	60.10	15.80	2.24	13.97	1.72	1.27	1.47	0.54	0.60	0.28
	NDFD	72.91	1.90	7.88	9.07	2.23	0.87	1.05	1.08	0.73	0.25
	aNDF	38.82	34.14	3.75	15.01	2.22	1.30	0.78	0.85	0.33	0.59
NEOSpectra TurnTable	ADF	65.80	16.21	4.45	6.11	2.14	1.01	0.48	0.87	0.46	0.26
	ADL	72.69	6.89	6.00	3.80	5.30	1.02	0.49	0.31	1.06	0.19
	CP	30.11	50.20	6.68	3.97	2.36	1.91	0.48	1.09	0.58	0.32
	IVTD	59.73	20.88	2.94	7.73	2.67	1.44	0.48	1.05	0.46	0.29
	NDFD	75.94	2.90	7.23	2.55	4.24	3.00	0.50	0.75	0.55	0.19
	aNDF	33.29	45.87	8.35	3.29	2.71	1.56	0.31	1.25	0.46	0.41
Trinamix Static	ADF	74.94	10.20	2.82	2.63	1.33	2.96	1.53	0.70	0.27	0.59
	ADL	79.72	5.00	2.72	1.96	1.78	3.62	1.22	1.35	0.65	0.30
	CP	42.12	37.73	5.05	1.99	2.13	2.41	1.72	3.17	0.45	1.02
	IVTD	70.04	13.63	2.41	3.13	1.39	2.30	0.73	3.34	0.57	0.55
	NDFD	80.21	3.67	3.60	1.17	2.53	2.12	2.01	1.96	0.25	0.54
	aNDF	45.70	33.24	5.78	2.67	1.75	4.03	2.68	0.53	0.70	0.81

ADF—Acid Detergent Fiber, ADL—Acid Detergent Lignin, CP—Crude Protein, IVTD—in vitro Total Digestibility, NDFD—Neutral Detergent Fiber Digestibility, aNDF—actual Neutral Detergent Fiber.

References

1. Gislon, G.; Ferrero, F.; Bava, L.; Borreani, G.; Prà, A.D.; Pacchioli, M.T.; Sandrucci, A.; Zucali, M.; Tabacco, E. Forage systems and sustainability of milk production: Feed efficiency, environmental impacts and soil carbon stocks. *J. Clean. Prod.* **2020**, *260*, 121012. [CrossRef]
2. Rukundo, I.R.; Danao, M.G.C.; Mitchell, R.B.; Masterson, S.D.; Weller, C.L. Comparing the use of handheld and benchtop NIR spectrometers in predicting nutritional value of forage. *Appl. Eng. Agric.* **2021**, *37*, 171–181. [CrossRef]
3. Cherney, J.H.; Digman, M.F.; Cherney, D.J. Handheld NIRS for forage evaluation. *Comput. Electron. Agric.* **2021**, *190*, 106469. [CrossRef]
4. Digman, M.F.; Cherney, J.H.; Cherney, D.J.R. The Relative Performance of a Benchtop Scanning Monochromator and Handheld Fourier Transform Near-Infrared Reflectance Spectrometer in Predicting Forage Nutritive Value. *Sensors* **2022**, *22*, 658. [CrossRef] [PubMed]
5. Cozzolino, D.; Labandera, M. Determination of dry matter and crude protein contents of undried forages by near-infrared reflectance spectroscopy. *J. Sci. Food Agric.* **2002**, *82*, 380–384. [CrossRef]
6. Acosta, J.J.; Castillo, M.S.; Hodge, G.R. Comparison of benchtop and handheld near-infrared spectroscopy devices to determine forage nutritive value. *Crop Sci.* **2020**, *60*, 3410–3422. [CrossRef]
7. Feng, X.; Cherney, J.H.; Cherney, D.J.R.; Digman, M.F. Practical Considerations for Using the NeoSpectra-Scanner Handheld Near-Infrared Reflectance Spectrometer to Predict the Nutritive Value of Undried Ensiled Forage. *Sensors* **2023**, *23*, 1750. [CrossRef] [PubMed]

8. Berzaghi, P.; Cherney, J.H.; Casler. Prediction performance of portable near infrared reflectance instruments using preprocessed dried, ground forage samples. *Comput. Electron. Agric.* **2021**, *182*, 106013. [CrossRef]
9. Barton, F.; Burdick, D. Prediction of forage quality with NIR reflectance spectroscopy. In *Proceedings of The XIV International Grassland Congress, Lexington, KY, USA, 15–24 June 1981*; CRC Press: Boca Raton, FL, USA, 2019; pp. 532–534.
10. Shenk, J.S.; Westerhaus, M.O.; Hoover, M.R. Analysis of Forages by Infrared Reflectance1. *J. Dairy Sci.* **1979**, *62*, 807–812. [CrossRef]
11. Beć, K.B.; Grabska, J.; Siesler, H.W.; Huck, C.W. Handheld near-infrared spectrometers: Where are we heading? *NIR News* **2020**, *31*, 28–35. [CrossRef]
12. Rego, G.; Ferrero, F.; Valledor, M.; Campo, J.C.; Forcada, S.; Royo, L.J.; Soldado, A. A portable IoT NIR spectroscopic system to analyze the quality of dairy farm forage. *Comput. Electron. Agric.* **2020**, *175*, 105578. [CrossRef]
13. Beć, K.B.; Grabska, J.; Huck, C.W. Principles and Applications of Miniaturized Near-Infrared (NIR) Spectrometers. *Chemistry* **2021**, *27*, 1514–1532. [CrossRef] [PubMed]
14. Yan, H.; De Gea Neves, M.; Noda, I.; Guedes, G.M.; Silva Ferreira, A.C.; Pfeifer, F.; Chen, X.; Siesler, H.W. Handheld Near-Infrared Spectroscopy: State-of-the-Art Instrumentation and Applications in Material Identification, Food Authentication, and Environmental Investigations. *Chemosensors* **2023**, *11*, 272. [CrossRef]
15. Bedord, L. John Deere Adds Manure Constituent Sensing to HarvestLab 3000. 2019. Available online: https://www.agriculture.com/news/technology/john-deere-adds-manure-constituent-sensing-to-harvestlab-3000 (accessed on 6 March 2024).
16. Chen, L.; Xing, L.; Han, L. Review of the application of near-infrared spectroscopy technology to determine the chemical composition of animal manure. *J. Environ. Qual.* **2013**, *42*, 1015–1028. [CrossRef] [PubMed]
17. Hart, L.; Huguenin-Elie, O.; Latsch, R.; Simmler, M.; Dubois, S.; Umstatter, C. Comparison of Spectral Reflectance-Based Smart Farming Tools and a Conventional Approach to Determine Herbage Mass and Grass Quality on Farm. *Remote Sens.* **2020**, *12*, 3256. [CrossRef]
18. Muncan, J.; Tsenkova, R. Aquaphotomics-from innovative knowledge to integrative platform in science and technology. *Molecules* **2019**, *24*, 2742. [CrossRef] [PubMed]
19. Tsenkova, R.; Munćan, J.; Pollner, B.; Kovacs, Z. Essentials of aquaphotomics and its chemometrics approaches. *Front. Chem.* **2018**, *6*, 363. [CrossRef] [PubMed]
20. Valentine, M.E.; Karayilanli, E.; Cherney, J.H.; Cherney, D.J. Comparison of in vitro long digestion methods and digestion rates for diverse forages. *Crop Sci.* **2019**, *59*, 422–435. [CrossRef]
21. Roelofs, R.; Shankar, V.; Recht, B.; Fridovich-Keil, S.; Hardt, M.; Miller, J.; Schmidt, L. A meta-analysis of overfitting in machine learning. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2019; pp. 9175–9185.
22. Mendes de Oliveira, D.; Pasquini, C.; Rita de Araújo Nogueira, A.; Dias Rabelo, M.; Lúcia Ferreira Simeone, M.; Batista de Souza, G. Comparative analysis of compact and benchtop near-infrared spectrometers for forage nutritional trait measurements. *Microchem. J.* **2023**, *196*, 109682. [CrossRef]
23. Malley, D.F.; McClure, C.; Martin, P.D.; Buckley, K.; McCaughey, W.P. Compositional Analysis of Cattle Manure During Composting Using a Field-Portable Near-Infrared Spectrometer. *Commun. Soil Sci. Plant Anal.* **2005**, *36*, 455–475. [CrossRef]
24. Williams, P. The RPD Statistic: A Tutorial Note. *NIR News* **2014**, *25*, 22–26. [CrossRef]
25. Williams, P. Influence of water on prediction of composition and quality factors: The aquaphotomics of low moisture agricultural materials. *J. Infrared Spectrosc.* **2009**, *17*, 315–328. [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

Sensors Editorial Office
E-mail: sensors@mdpi.com
www.mdpi.com/journal/sensors

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 978-3-7258-5954-2