mathematics

MDPI

Article

Clustering-Based Class Hierarchy Modeling for Semantic Segmentation Using Remotely Sensed Imagery

Lanfa Liu [1,2], Song Wang [2], Zichen Tong [2] and Zhanchuan Cai [1,3,*]

[1] State Key Laboratory of Lunar and Planetary Sciences, Macau University of Science and Technology, Macau 999078, China
[2] Key Laboratory for Geographical Process Analysis and Simulation of Hubei Province, College of Urban and Environmental Sciences, Central China Normal University, Wuhan 430079, China
[3] School of Computer Science and Engineering, Macau University of Science and Technology, Macau 999078, China
* Correspondence: zccai@must.edu.mo

Abstract: Land use/land cover (LULC) nomenclature is commonly organized as a tree-like hierarchy, contributing to hierarchical LULC mapping. The hierarchical structure is typically defined by considering natural characteristics or human activities, which may not optimally align with the discriminative features and class relationships present in remotely sensed imagery. This paper explores a novel cluster-based class hierarchy modeling framework that generates data-driven hierarchical structures for LULC semantic segmentation. First, we perform spectral clustering on confusion matrices generated by a flat model, and then we introduce a hierarchical cluster validity index to obtain the optimal number of clusters to generate initial class hierarchies. We further employ ensemble clustering techniques to yield a refined final class hierarchy. Finally, we conduct comparative experiments on three benchmark datasets. Results demonstrating that the proposed method outperforms predefined hierarchies in both hierarchical LULC segmentation and classification.

Keywords: data-driven hierarchy; hierarchical segmentation; LULC; remote sensing

MSC: 68T07

check for updates

Academic Editors: Volodymyr Ponomaryov, Vladimir Lukin, Bogdan Smolka and Beatriz P. García Salgado

Received: 22 December 2024
Revised: 14 January 2025
Accepted: 18 January 2025
Published: 21 January 2025

Citation: Liu, L.; Wang, S.; Tong, Z.; Cai, Z. Clustering-Based Class Hierarchy Modeling for Semantic Segmentation Using Remotely Sensed Imagery. *Mathematics* **2025**, *13*, 331. https://doi.org/10.3390/math13030331

1. Introduction

Land use/land cover (LULC) mapping is an important variable in the study of Earth's surface properties, playing a crucial role in understanding global environmental change [1]. The rapid development of remote sensing technologies has advanced our ability to characterize diverse land cover classes and derive precise land use information. In particular, hyperspectral and high spatial resolution (HSR) imaging is capable of capturing rich spectral and spatial information, which can better distinguish surface features and objects, enabling finer discrimination of these two properties. This abundant semantic information presents the opportunity for fine-grained LULC mapping and also bring challenges such as large intra-class variance and class imbalance [2,3]. Therefore, understanding the content in remote sensing images has become an increasingly urgent practical need.

LULC nomenclatures are commonly organized as tree-like hierarchical structures [4], which vary across different products due to differences in spatial scale, data sources, and application requirements. For instance, China's land use/cover datasets (CLUDs) [5] organize land cover into 6 primary classes (cropland, forest, grassland, water, built-up area, and barren) and further subdivide into 25 secondary classes to capture specific types within each primary class. The MODIS Collection 5 Global Land Cover Type product [6] employs a three-level hierarchical framework with 6 top-level classes and 32 detailed land cover classes. The CORINE system divides land cover classes into a hierarchical structure of three levels, with 5 primary classes (artificial surfaces, agricultural land, forest and semi-natural land, wetlands, and water), 25 secondary classes, and 44 detailed classes [7]. These multi-level hierarchies enable detailed and scalable representation of LULC information for various applications.

LULC mapping methodologies can be categorized into flat and hierarchical approaches based on their consideration of class hierarchical relationships [8]. Flat methods directly classify each pixel in the imagery without considering hierarchical relationships between classes. Support vector machine (SVM) [9], neural networks [10], and multinomial logistic regression [11] were traditionally used. Recently, deep learning methods have gained popularity due to their ability to automatically learn complex patterns and relationships in data [12]. U-Net is one of the most popular deep learning-based semantic segmentation algorithms for LULC mapping [13]. A fully convolutional network (FCN) is employed to perform pixel-wise semantic segmentation of remotely sensed imagery, demonstrating the capability in LULC mapping [14,15]. The emergence of vision transformer (ViT) has further advanced the field, and various transformer-based semantic segmentation models are proposed for LULC mapping [16,17]. In contrast, hierarchical methods take advantage of structured relationships between classes, which have shown potential for enhancing LULC mapping accuracy. Recent studies [18,19] have explored the extent to which hierarchical methods can improve the accuracy of land cover mapping, and comparative studies were conducted using the Sentinel-2 dataset with random forest (RF). A three-stage hierarchical framework is proposed to map peatland sub-classes using multi-sensor data [20]. The class hierarchy is considered in the loss function for training to achieve consistent hierarchical land use classification [21]. HierU-Net [22] is proposed to improve land cover segmentation by incorporating tree-like hierarchical information of land cover classes with U-Net.

However, the hierarchical structures are typically predefined based on domain expertise, which may not optimally reflect the discriminative patterns present in remotely sensed imagery. The potential of data-driven approaches for class hierarchy modeling remains unexplored in LULC semantic segmentation. Attempts at data-driven hierarchy construction have shown promise in image classification tasks. The authors of [23] identified the hierarchical structure by performing spectral cluster on the confusion matrix (CM) generated by preliminary validation, which is flexible without relying on input features and successfully adopted in HD-CNN for hierarchical classification [24]. A hierarchical cluster validity index (HCVI) [25] is developed to select the optimal number of clusters for K-means clustering, and a reasonable hierarchical structure is built by using deep features extracted by a VGG16 model. However, the authors of [26] pointed out that the performance of HCVI is still facing the unstable issue. Furthermore, there are methods for constructing hierarchies by means of building visual trees from the perspective of class similarity, e.g., HAP [27], JHCSL [28], and MTHL [29].

The hierarchy generation-related literature mainly focuses on classification tasks, and it is challenging when applied to dense pixel-wise segmentation involving large-scale pixels. In this paper, we contribute to a cluster-based class hierarchy modeling framework, which, to the best of our knowledge, is the first practical hierarchy generation workflow in LULC semantic segmentation. Inspired by CM [23], initial data-driven hierarchies are generated through spectral clustering on confusion matrices obtained by a flat model. We introduce a robust method for determining optimal cluster numbers by introducing HCVI and further integrate with ensemble clustering techniques, enhancing the stability and reliability of the generated hierarchies. The proposed method potentially advances the state of the art in hierarchical LULC semantic segmentation by introducing a novel class hierarchy modeling framework by incorporating ensemble techniques and mitigating class imbalance.

2. Methodology

Given a set of n fine classes $\mathcal{X} = \{x_1, x_2, \cdots, x_n\}$, our goal is to model a class hierarchy that will be able to group the most detailed classes into primary classes. We performed spectral clustering based on the confusion matrix from the flat segmentation or classification results. While cluster results can be independently sampled from K clusters, represented as $\mathcal{C} = \{C_1, \cdots, C_k\}$, where K_i is the cluster number for the i-th BP, $1 \leq \pi_i(x_j) \leq K_i$, and $1 \leq i \leq r, 1 \leq j \leq n$. HCVI is introduced to assess the clustering results from a perspective of imbalance and to obtain a basic partition (BP) by iterating through each cluster center. A set of BPs is obtained and fed into an ensemble clustering algorithm to achieve the final clustering result as the class hierarchy. An illustration of the proposed method is given in Figure 1. The details of our approach are described below.

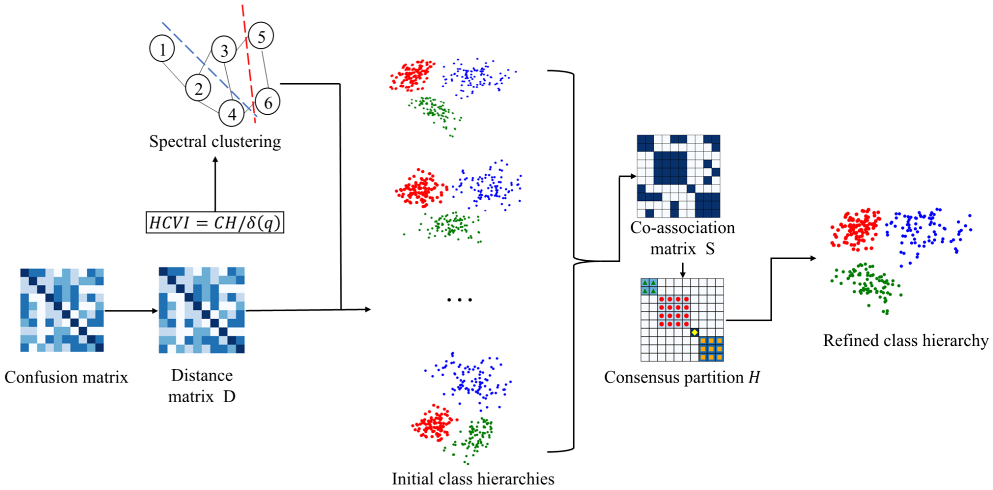

Figure 1. Illustration of the proposed hierarchy generation method.

Initial class hierarchy discovery. U-Net has been widely used and proven to be effective in image segmentation tasks. It has a unique architecture that combines convolutional and deconvolutional layers, allowing it to capture both local and global features of the input image. We firstly employed a U-Net as the flat approach to conduct semantic segmentation experiments to obtain a confusion matrix F and construct a distance matrix D by the following equation:

$$D = \frac{1}{2}[(I - F) + (I - F)^T], \tag{1}$$

where F stands for the obtained confusion matrix and I stands for the unit matrix. In the formula, $(I - F)$ and $(I - F)^T$ are used to construct a symmetrical matrix that can facilitate the calculation of spectral clustering and ensures that the distance measure is consistent. Each entry D_{ij} in matrix D represents the distance or dissimilarity between class i and class j. Spectral clustering does not rely on assumptions that the data lie in convex sets or depend on the Euclidean distance metric, making it well suited for capturing non-linear data structures. This advantage makes it an appropriate choice compared to some other clustering methods, like K-Means, when clustering on confusion matrices. We perform the spectral cluster algorithm on matrix D based on the number of clusters to group secondary classes into primary classes, thereby creating a two-level hierarchy that maps multiple fine classes into a single coarse class.

Optimal number of clusters. However, the hierarchies we obtained in the previous step were obtained without choosing the optical number of clusters, which will affect the performance of the clustering algorithm. We follow HCVI by introducing a balance parameter $\delta(q)$ and the Calinski–Harabasz index (CH) [30] to assess the quality of the generated hierarchy. $\delta(q)$ takes into account the effects of sample data imbalance within the data and the imbalance of clustering classes in a comprehensive way. The index is formulated to evaluate the hierarchical clustering, as given by the following formula:

$$\delta(q) = \frac{1}{q}\sum_{q=1}^{k}((\frac{r_k - r_E}{r_E})^2 + 1)((\frac{m_k - m_E}{m_E})^2 + 1), \tag{2}$$

where the parameter of $\sum_{q=1}^{k}((\frac{r_k - r_E}{r_E})^2 + 1)$ is to indicate category balance in the clustering, the parameter of $\sum_{q=1}^{k}((\frac{m_k - m_E}{m_E})^2 + 1)$ is to indicate sample balance, q is the number of clusters, r_k is the number of fine-grained classes within superclass k, r_E is the average of the number of fine-grained classes within all superclasses, m_k is the number of samples within superclass k, and m_E is the average of the number of samples within all superclasses. The CH index reflects the degree of goodness of the clustering by the ratio of the inter-cluster distance within the clusters to the intra-cluster distance, which is calculated by the formula as

$$CH = \frac{tr(B_k)(N-K)}{tr(W_k)(K-1)}, \tag{3}$$

$$B_k = \sum_{q=1}^{k} n_q(c_q - c_e)(c_q - c_e)^T, \tag{4}$$

$$W_k = \sum_{q=1}^{k}\sum_{x \in c_q}(x - c_q)(x - c_q)^T, \tag{5}$$

where N is the total number of samples, K is the number of classes formed by clustering, c_e represents the class centroid, n_q represents the number of samples in class q, and c_q represents the data set of q. The HCVI is calculated by the formula as

$$HCVI = CH/\delta(q). \tag{6}$$

Ensemble clustering for class hierarchy. Although we are able to generate class hierarchy efficiently using cluster algorithms, the category hierarchies obtained suffer from instability due to the robustness problems inherent in the cluster algorithms. Therefore, the ensemble clustering method [31] is introduced to obtain a refined class hierarchy.

We perform the cluster algorithm 100 times to obtain a set of BPs. We denote $\Pi = \{\pi_1, \pi_2, \cdots, \pi_r\}$ as r BPs, each of which divides $\mathcal{X}$ into K_i crisply partitioned and maps $\mathcal{X}$ into a label set $\pi_i = \{\pi_i(x_1), \pi_i(x_2), \cdots, \pi_i(x_n)\}$. Ensemble clustering methods summarize r BPs as a co-association matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ which calculates the number of times two instances occur in the same cluster based on Π. It can be defined as

$$\mathbf{S}(x_p, x_q) = \sum_{i=1}^{r} \delta(\pi_i(x_p), \pi_i(x_q)), \tag{7}$$

where $x_p, x_q \in \mathcal{X}$ and $\delta(a, b)$ are 1 if $a = b$; 0 otherwise. Obviously, $\mathbf{S}$ could be normalized by $\mathbf{S} = \mathbf{S}/r$. Its trace minimization form is calculated by following

$$\min_{\mathbf{H}} \operatorname{tr}(\mathbf{H}^\top \mathbf{L}_s \mathbf{H}) \quad \text{s.t.} \quad \mathbf{H}^\top \mathbf{H} = \mathbf{I}, \tag{8}$$

where $\mathbf{L}_s = \mathbf{I} - \mathbf{D}_s^{-1/2}\mathbf{S}\mathbf{D}_s^{-1/2}$ is the normalized Laplacian matrix of $\mathbf{S}$, with degree matrix $\mathbf{D}_s \in \mathbb{R}^{n\times n}$ being a diagonal matrix whose j^{th} diagonal element is the sum of the j^{th} row of $\mathbf{S}$, and $\mathbf{H} \in \mathbb{R}^{n\times K}$ is defined as the scaled partition matrix of π:

$$H_{jk} = \begin{cases} \frac{1}{\sqrt{|C_k|}}, & \text{if } x_j \in C_k \text{ in } \pi, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

The final ensemble clustering result is represented as $\mathbf{H}$, which will be applied to the hierarchical mapping.

3. Results and Discussion

3.1. Experiment Setup

To evaluate our proposed clustering-based class hierarchy modeling framework, we conducted experiments on three benchmark datasets for hierarchical semantic segmentation (GID-15 and HierToulouse) and hierarchical classification (DFC18). These three datasets differ in their location, size, and class structure, and are suitable for a comprehensive evaluation of the proposed mechanism. All the experiments are implemented using Pytorch on a workstation with a NVIDIA RTX 4090 GPU.

The GID-15 dataset primarily covers urban and rural areas in 60 different cities in China, with a total annotated area exceeding 50,000 square kilometers. It contains 25,200 patches of 512 × 512 pixels, and we divide them into two subsets, with 20,160 patches used for training and 5040 patches for testing. For the GID-15 dataset, we consider the predefined hierarchy as organized by [32]. It contains a total of 15 of the most detailed and 5 primary land cover classes.

The HierToulouse dataset is a high-resolution remote sensing dataset focused on the urban area of Toulouse, France. It contains 11,528 paired image patches of 512 × 512 pixels, 8624 images for training, 1452 images for validation, and another 1452 images for testing.

The DFC18 dataset is provided by the 2018 Data Fusion Contest, which includes spectral data with 48 bands ranging from 380 to 1050 nm with 1 m spatial resolution. It covers a real urban scene in and around the University of Houston and contains a total number of 20 land use classes. We organized it as a two-level hierarchy by referring to the OCS GE LULC system, as shown in Figure 2. The predefined hierarchy contains six primary classes, namely vegetation, natural surfaces, buildings, transportation, public facilities, and others, corresponding to the class codes in the dataset as 1–5, 6–7, 8–9, 10–16, 17–18, and 19–20. We followed [33] to sample the data for the DFC18 dataset.

To assess hierarchical classification and segmentation performance, we use four metrics including Overall Accuracy (OA) and Kappa coefficient for classification and mIoU and FWIoU for semantic segmentation. The formulas are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}, \tag{10}$$

$$Kappa = \frac{2 * TP}{2 * TP + FP + FN}, \tag{11}$$

$$mIoU = \frac{1}{k+1}\sum_{i=0}^{k} \frac{TP}{TP + FP + FN}, \tag{12}$$

$$FWIoU = \frac{\sum_c P_c \cdot \mathrm{IoU}_c}{\sum_c P_c}, \tag{13}$$

where TP, TN, FP, FN, P_c, and IoU_c represent true positive, true negative, false positive, false negative, weight of class c, and IoU score of class c, respectively.

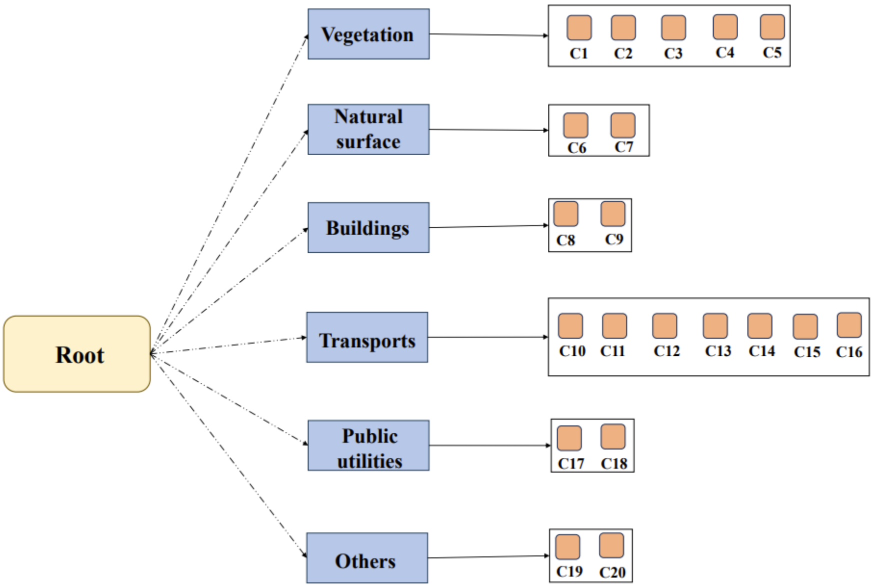

Figure 2. The predefined hierarchy for the DFC18 dataset.

3.2. Experiment Results on the HierToulouse Dataset

To validate the effectiveness of our data-driven hierarchy approach for LULC segmentation tasks, we conducted hierarchical segmentation experiments on the GID-15 and HierToulouse datasets using HierU-Net by considering three hierarchies: predefined, CM, and ours. The loss curves of the three hierarchies during model training is shown in Figure 3.

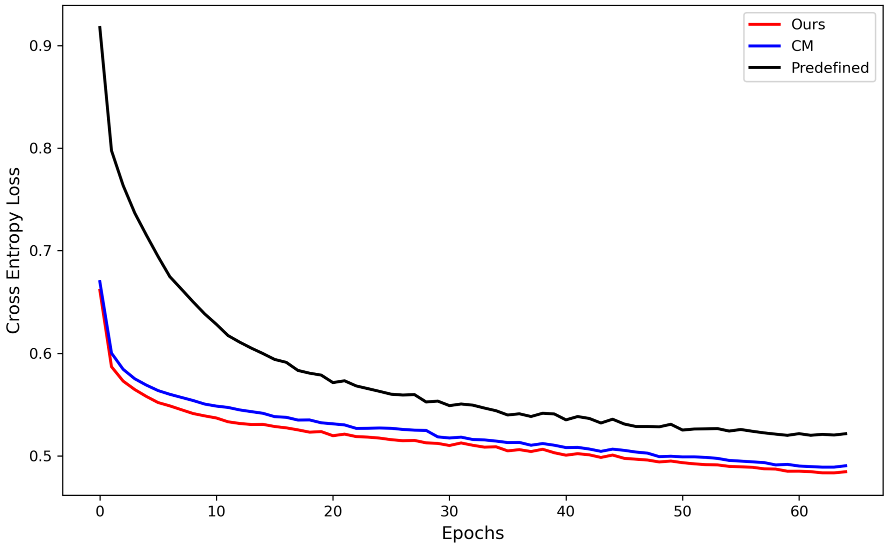

Figure 3. Loss curves during model training.

The IoU scores for each class, mIoU, OA, and FWIoU scores achieved by HierU-Net with different hierarchies on the HierToulouse dataset are presented in Table 1. All three methods demonstrate good performance, with our proposed hierarchy achieving the highest score on mIoU, OA, and FWIoU. On the mIoU score, ours achieves the highest score of 45.7%, representing a significant improvement of 2.99% over the predefined hierarchy method and 1.29% over the CM-based hierarchy method. In terms of the OA score, our proposed method achieves the highest value of 88.98%, outperforming the predefined hierarchy by 1.80% and the CM-based hierarchy by 0.79%. On the FWIoU score, our proposed method achieves the highest value of 81.97%, outperforming the predefined hierarchy by 4.21% and the CM-based hierarchy by 1.32%. For per-class comparison, our proposed hierarchy consistently outperforms the other two methods, suggesting the proposed hierarchy is better suited to model class relationships for semantic segmentation. The predefined hierarchy has the worst results, particularly for *Construction* and *Non-Construction* classes, comparatively.

Table 1. Hierarchical segmentation performance on the HierToulouse dataset.

Class Names	Predefined	CM	Ours	Proportion of Pixels (%)
No. of clusters	4	4	3	
$\delta(q)$	2.1448	1.9445	1.7459	
Construction	69.69	75.3	75.39	13.52
Non-construction	64.65	69.59	70.07	12.06
Mineral material	30.61	29.26	39.29	1.88
Water surface	74.85	78.72	80.5	2.19
Broad leaved forest	57.40	58.53	58.67	10.89
Shrubbery	24.67	23.14	27.55	1.60
Herage	87.26	88.48	88.53	57.50
mIoU	42.71	44.41	45.7	
FWIoU	77.76	80.62	81.97	
OA	87.18	88.19	88.98	

Figure 4 presents a visual comparison of the outcomes obtained using HierU-Net with all three hierarchies. These three methods achieve satisfactory segmentation results. for *Construction area* and *Water Surface*, the segmentations produced by all three methods align well with the LC labels. In comparison, the predefined hierarchy exhibits the worst performance. For *Non-construction area*, the segmentation results of the predefined hierarchy are worse compared to CM and ours, as reflected by the boundaries and integrity of roads.

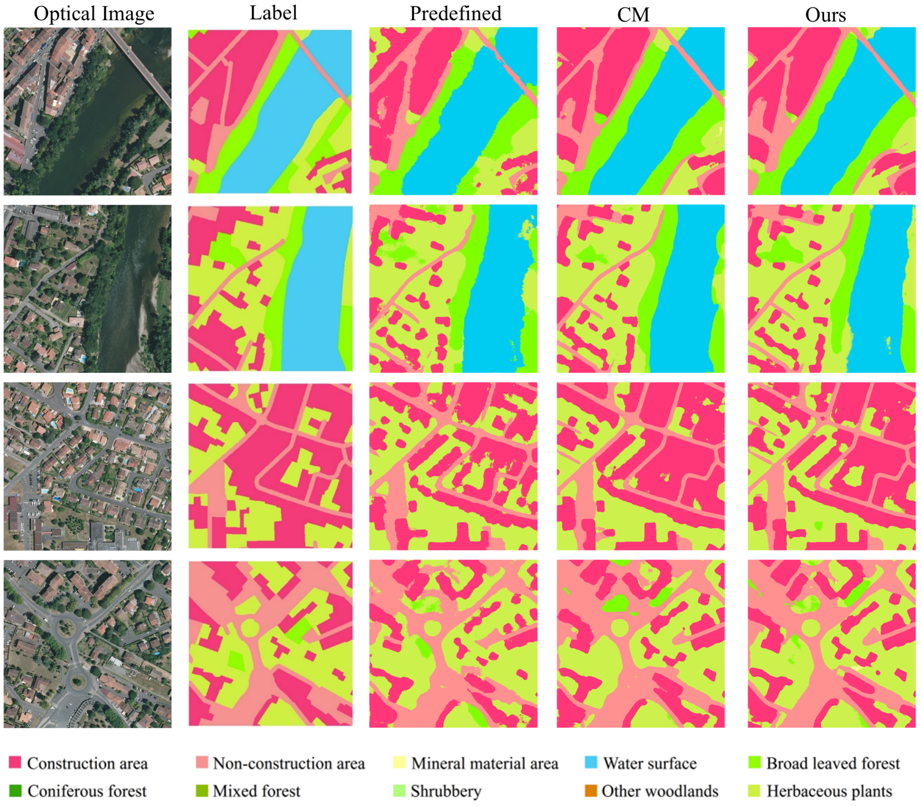

Figure 4. Segmentation results with HierU-Net framework on the HierToulouse dataset (From **left** to **right** are Optical image, LC label, results from predefined hierarchy, CM, Ours).

3.3. Experiment Results on the GID-15 Dataset

Table 2 summarizes the IoU scores for each class on the GID-15 dataset, mIoU, OA, and FWIoU scores obtained with three different hierarchies. Our proposed method achieves the highest mIoU of 54.53%, outperforming the predefined hierarchy by 1.08% and the CM-based hierarchy by 1.06%. On the OA score, our proposed method achieves the highest value of 77.22%, outperforming the predefined hierarchy by 2.84% and the CM-based hierarchy by 1.31%. Similarly, for the FWIoU score, our proposed method achieves the highest value of 67.97%, outperforming the predefined hierarchy by 3.23% and the CM-based hierarchy by 1.86%.

Table 2. Hierarchical semantic segmentation performance on the GID-15 dataset.

Class Names	Predefined	CM	Ours	Proportion of Pixels (%)
No. of clusters	5	5	5	
$\delta(q)$	2.0768	1.89	1.4258	
Irrigated land	71.58	49.23	47.87	2.97
Paddy field	73.89	75.26	75.67	38.92
Dry cropland	15.78	63.9	62.58	11.06
Artificial meadow	21.75	30.8	35.39	0.92
Arbor forest	86.69	68.79	60.5	8.67
Shrub land	74.92	37.34	65.6	4.26
Natural meadow	36.27	40.36	52.45	1.09
Garden land	0	35.29	33.97	0.30
Industrial land	61.4	63.88	61.45	2.90
Urban residential	72.37	69.07	75.34	5.68
Rural residential	63.5	66.91	67.63	4.76
Traffic land	58.48	58.42	61.35	2.48
River	57.64	49.62	39.56	4.60
Lake	78.24	65.41	70.23	9.80
Pond	28.93	27.72	23.44	1.60
mIoU	53.45	53.47	54.53	
FWIoU	64.74	66.11	67.97	
OA	74.38	75.91	77.22	

When examining per-class IoU scores, our method performs particularly well on certain classes. Specifically, it achieves the highest IoU score in six classes, including frequently occurring classes *Rural residential*, *Urban residential*, and *Traffic land*, demonstrating that our framework is able to effectively capture the characteristics of these classes and improve the segmentation results. There are also some classes where the performance difference between the three methods is not significant, such as IoU scores for *Industrial residential*, ranging from 61.4% to 63.88%.

Figure 5 illustrates segmentation results obtained using HierU-Net with both data-driven and predefined hierarchies. All three methods performed well on *Urban residential, Industrial land, Traffic land, and Irrigated land*. Compared with CM and predefined hierarchies, ours achieved clear boundary of *Traffic land*. By modeling the hierarchical relationships between classes, semantic segmentation with our proposed class hierarchy achieved improved performance on easily confused classes, particularly between *Urban residential* and *Industrial land*.

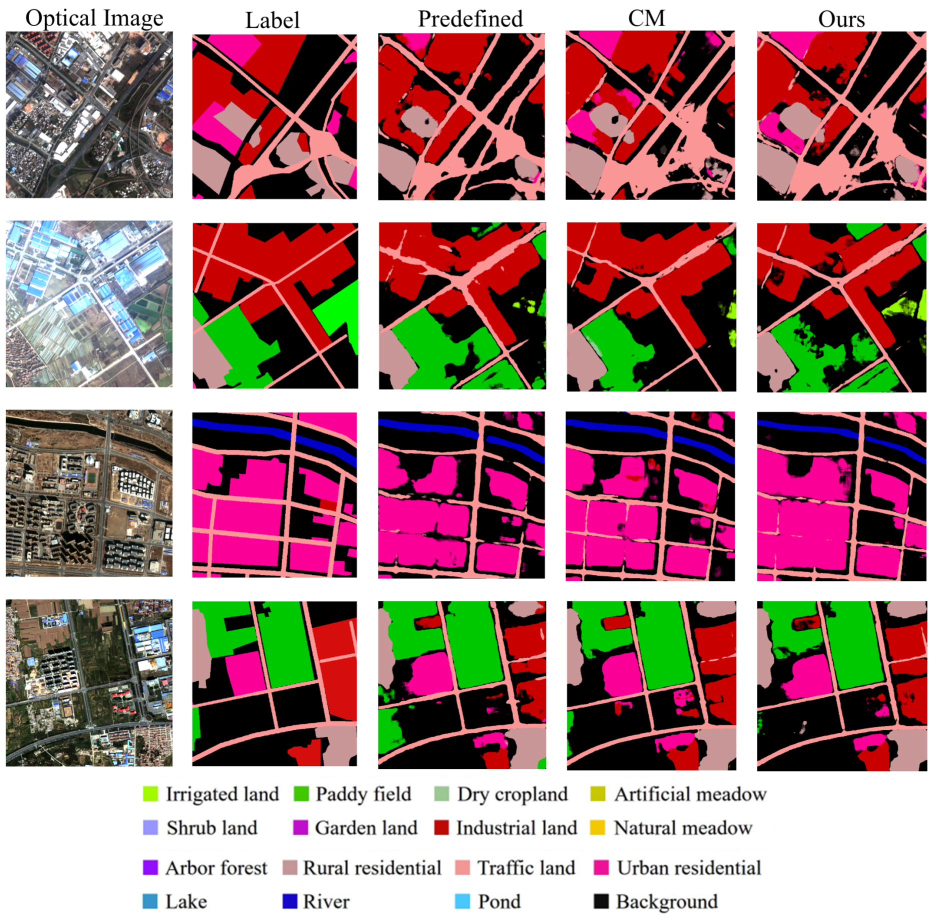

Figure 5. Segmentation results with HierU-Net framework on the GID dataset (From **left** to **right** are Optical image, LC label, results from predefined hierarchy, CM, Ours).

3.4. Evaluation on Hierarchical Classification

To further demonstrate the flexibility of our proposed method, we investigate its applicability to hierarchical classification tasks. We conduct experiments on the DFC18 dataset using HD-CNN, employing four distinct hierarchies: predefined, CM, HCVI, and our proposed hierarchy. Additionally, comparative analysis is performed with three flat classification methods, namely SVM, RF, and CNN.

Table 3 presents performance comparisons of flat and hierarchical methods on the DFC18 dataset. The results indicate that hierarchical classification methods consistently outperform flat approaches and highlight the advantages of utilizing structured relationships among classes. Our proposed method achieves the highest overall accuracy of 53.64% and a Kappa coefficient of 0.511, demonstrating its effectiveness in classification tasks. Compared to the CM-based method, our approach improves overall accuracy by 1.46%. Similarly, it outperforms the HCVI-based method by 0.22% and the predefined hierarchy by 0.87%. When comparing our hierarchical method to flat classifications (SVM, RF, and CNN), the advantages become even more obvious. Figure 6 illustrates classification results obtained using flat and hierarchical methods. Visual comparison shows that the hierarchical way reduces some errors, particularly among classes prone to be confused such as *evergreen trees* and *deciduous trees*.

Table 3. Hierarchical and flat classification performance on the DFC18 dataset

Hierarchy	$\delta(q)$	No. of Clusters	Method	Accuracy (%)	Kappa
CM	2.1730	6	HD-CNN	52.18	0.495
HCVI	2.0249	6	HD-CNN	53.42	0.507
Ours	1.9731	6	HD-CNN	53.64	0.511
Predefined	2.5919	6	HD-CNN	52.77	0.489
	-	-	SVM	42.38	0.393
-	-	-	RF	43.65	0.401
	-	-	CNN	33.42	0.293

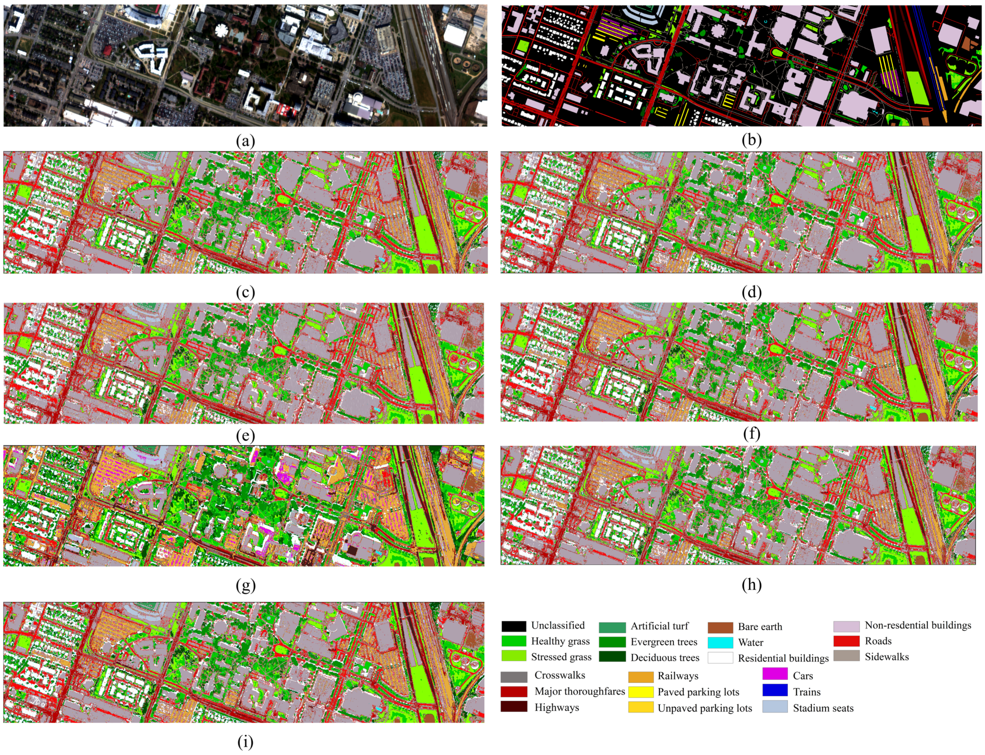
(a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 6. Classification results of flat and hierarchical classification with different hierarchies: (**a**) RGB image; (**b**) Ground truth; (**c**) Ours; (**d**) CM; (**e**) HCVI; (**f**) Predefined; (**g**) SVM; (**h**) RF; (**i**) CNN.

3.5. Advantages and Disadvantages

Unlike predefined hierarchies that may be constrained by expert assumptions, we proposed a data-driven approach to generate hierarchical structures. Instead of clustering directly using image features, we adopted the idea of CM by performing clustering on confusion matrices obtained by a flat model, which alleviate the challenge of large-scale dense pixels. For the class imbalance issue, we considered spectral clustering to generate initial hierarchies and further refine with ensemble clustering. The objective function of spectral clustering penalizes unbalanced partitions, hence encouraging balanced trees. By analyzing the balancing parameter of $\delta(q)$, we can observe that our approach achieves better balanced class hierarchy. The larger the parameter of $\delta(q)$, the more imbalanced it is. Specifically, for the HierToulouse experiment, the balancing parameter for our method is 1.75, while the predefined hierarchy has a value of 2.14 and the CM method has a value of 1.94. This indicates the advantage of our approach in effectively mitigating the class imbalance problem compared to other methods.

The proposed approach is particularly advantageous in LULC mapping using remotely sensed imagery, where the spectral properties of the detailed classes may vary depending on the region and the sensor. The comparative experiments on three datasets showed that our proposed hierarchy generation framework consistently leads to better performance, as evidenced by higher mIoU score in semantic segmentation and increased accuracy in classification. We observed considerable improvement in accuracy for classes with small samples, which is meaningful in the context of real-world applications. For instance, in the HierToulouse dataset, the *Shurberry* class, which has the smallest number of pixels with a proportion of 1.6%, exhibited an improvement in accuracy from 24.67% (predefined hierarchy) to 27.55% with our proposed hierarchy. The $Artificial\ meadow$ class in the GID-15 dataset, having a proportion of 0.92%, also showed substantial improvements in mIoU scores. Compared to the predefined hierarchy, our proposed approach and the CM-based approach increased the mIoU score for $Artificial\ meadow$ to 35.39% and 30.8%, respectively.

Our method has a dependency on the dataset with fine-grained classes. Experiments were conducted mainly in urban and peri-urban areas, and the application to agriculture areas has not been explored. Additionally, another disadvantage of our method is the increased computational complexity. In certain circumstances, there is a need to make trade-offs between accuracy and computational efficiency.

4. Conclusions

In this study, we propose a cluster-based class hierarchy modeling framework for LULC mapping using remotely sensed imagery. The hierarchical structure is constructed through three stages, ensuring the hierarchy aligns closely with the intrinsic characteristics of the data. We conducted experiments on three benchmark datasets, utilizing HierU-Net for hierarchical segmentation and HD-CNN for hierarchical classification. Our proposed hierarchy consistently outperforms predefined hierarchy, demonstrating its success in LULC mapping; it also has the potential to make a broader contribution to scene understanding with structured semantic knowledge. By capturing the relationships between different classes, the hierarchical model can help in better understanding the context and semantics of the scene. The approach can also be applied to other domains where hierarchical relationships exist in the data. For example, in medical imaging, hierarchical models can be used to represent different levels of anatomical structures or disease classifications.

Future work could explore hybrid approaches that integrate the strengths of different hierarchies by considering the use of mixture of experts (MoE), which is a technique that combines multiple expert models to make predictions. Each expert model specializes in a particular subset of the input space, and a gating mechanism is used to determine which experts to use for a given input. For example, we can have one set of experts for the predefined hierarchy and another set of experts for the data-driven hierarchy. The gating mechanism can then be used to determine which experts to use based on the input data. This can help to capture the advantages of both predefined and data-driven hierarchies, leading to more accurate and reliable semantic segmentation results.

Author Contributions: Conceptualization, L.L. and S.W.; methodology, S.W.; experiments, S.W.; formal analysis, S.W. and Z.T.; data curation, S.W. and Z.T.; writing—original draft preparation, S.W. and L.L.; writing—review and editing, Z.C.; supervision, Z.C.; funding acquisition, Z.C. and L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 42271482, in part by the International Science and Technology Cooperation Program of Hubei Province under Grant 2022EHB019, in part by Guangzhou Development District International Cooperation Project (Grant No. 2023GH01) and in part by a grant from the Macao Young Scholars Program (AM2022013).

Data Availability Statement: All data are listed in the paper.

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

References

1. Gómez, C.; White, J.C.; Wulder, M.A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 55–72. [CrossRef]
2. He, S.; Yang, H.; Zhang, X.; Li, X. MFTransNet: A Multi-Modal Fusion with CNN-Transformer Network for Semantic Segmentation of HSR Remote Sensing Images. *Mathematics* **2023**, *11*, 722. [CrossRef]
3. Guo, S.; Yang, Q.; Xiang, S.; Wang, S.; Wang, X. Mask2Former with Improved Query for Semantic Segmentation in Remote-Sensing Images. *Mathematics* **2024**, *12*, 765. [CrossRef]
4. Lei, Z.; Li, H.; Zhao, J.; Jing, L.; Tang, Y.; Wang, H. Individual Tree Species Classification Based on a Hierarchical Convolutional Neural Network and Multitemporal Google Earth Images. *Remote Sens.* **2022**, *14*, 5124. [CrossRef]
5. Liu, J.; Kuang, W.; Zhang, Z.; Xu, X.; Qin, Y.; Ning, J.; Zhou, W.; Zhang, S.; Li, R.; Yan, C.; et al. Spatiotemporal characteristics, patterns, and causes of land-use changes in China since the late 1980s. *J. Geogr. Sci.* **2014**, *24*, 195–210. [CrossRef]
6. Sulla-Menashe, D.; Gray, J.M.; Abercrombie, S.P.; Friedl, M.A. Hierarchical mapping of annual global land cover 2001 to present: The modis collection 6 land cover product. *Remote Sens. Environ.* **2019**, *222*, 183–194. [CrossRef]
7. Commission of the European Communities. CORINE Land Cover. Available online: http://www.eea.europa.eu/publications/COR0-landcover (accessed on 29 October 2024).
8. Fenske, K.; Feilhauer, H.; Förster, M.; Stellmes, M.; Waske, B. Hierarchical classification with subsequent aggregation of heathland habitats using an intra-annual RapidEye time series. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *87*, 102036. [CrossRef]
9. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [CrossRef]
10. Zhong, Y.; Zhang, L. An adaptive artificial immune network for supervised classification of multi/hyperspectral remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 894–909. [CrossRef]
11. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4085–4098. [CrossRef]
12. Li, S.; Song, W.; Fang, L.; Chen, Y.; Ghamisi, P.; Benediktsson, J.A. Deep learning for hyperspectral image classification: An overview. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6690–6709. [CrossRef]
13. Solórzano, J.V.; Mas, J.F.; Gao, Y.; Gallardo-Cruz, J.A. Land Use Land Cover Classification with U-Net: Advantages of Combining Sentinel-1 and Sentinel-2 Imagery. *Remote Sens.* **2021**, *13*, 3600. [CrossRef]
14. Buttar, P.K.; Sachan, M.K. Land Cover Segmentation Using 3-D FCN-Based Architecture with Coordinate Attention. *IEEE Geosci. Remote. Sens. Lett.* **2024**, *21*, 1–5. [CrossRef]
15. Volpi, M.; Tuia, D. Dense Semantic Labeling of Subdecimeter Resolution Images with Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 881–893. [CrossRef]
16. Fan, J.; Shi, Z.; Ren, Z.; Zhou, Y.; Ji, M. DDPM-SegFormer: Highly refined feature land use and land cover segmentation with a fused denoising diffusion probabilistic model and transformer. *Int. J. Appl. Earth. Obs. Geoinf. 2024*, *133*, 104093. [CrossRef]
17. Wang, L.; Li, R.; Zhang, C.; Fang, S.; Duan, C.; Meng, X.; Atkinson, P.M. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *190*, 196–214. [CrossRef]
18. Demirkan, D.; Koz, A.; Düzgün, H. Hierarchical classification of sentinel 2-a images for land use and land cover mapping and its use for the corine system. *J. Appl. Remote Sens.* **2020**, *14*, 026524. [CrossRef]
19. Waśniewski, A.; Hościło, A.; Chmielewska, M. Can a hierarchical classification of sentinel-2 data improve land cover mapping? *Remote Sens.* **2022**, *14*, 989. [CrossRef]
20. Pontone, N.; Millard, K.; Thompson, D.K.; Guindon, L.; Beaudoin, A. A hierarchical, multi-sensor framework for peatland sub-class and vegetation mapping throughout the Canadian boreal forest. *Remote Sens. Ecol. Conserv.* **2024**, *10*, 500–516. [CrossRef]
21. Yang, C.; Rottensteiner, F.; Heipke, C. A hierarchical deep learning framework for the consistent classification of land use objects in geospatial databases. *ISPRS J. Photogramm. Remote Sens.* **2021**, *177*, 38–56. [CrossRef]
22. Liu, L.; Tong, Z.; Cai, Z.; Wu, H.; Zhang, R.; Le Bris, A.; Olteanu-Raimond, A.M. HierU-Net: A hierarchical semantic segmentation method for land cover mapping. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–14. [CrossRef]
23. Bengio, S.; Weston, J.; Grangier, D. Label embedding trees for large multi-class tasks. In Proceedings of the 23rd International Conference Neural Information Processing Systems, Vancouver, BC, Canada, 7–10 December 2010; pp. 163–171.
24. Yan, Z.; Zhang, H.; Piramuthu, R.; Jagadeesh, V.; DeCoste, D.; Di, W.; Yu, Y. HD-CNN: Hierarchical deep convolutional neural networks for large scale visual recognition. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 2740–2748.
25. Zheng, Y.; Chen, Q.; Fan, J.; Gao, X. Hierarchical convolutional neural network via hierarchical cluster validity based visual tree learning. *Neurocomputing* **2020**, *409*, 408–419. [CrossRef]
26. Huang, H.; Wang, Y.; Hu, Q. Building hierarchical class structures for extreme multi-class learning. *Int. J. Mach. Learn. Cybern.* **2023**, *14*, 2575–2590. [CrossRef]

27. Zheng, Y.; Fan, J.; Zhang, J.; Gao, X. Hierarchical learning of multi-task sparse metrics for large-scale image classification. *Pattern Recognit.* **2017**, *67*, 97–109. [CrossRef]
28. Qu, Y.; Lin, L.; Shen, F.; Lu, C.; Wu, Y.; Xie, Y.; Tao, D. Joint hierarchical category structure learning and large-scale image classification. *IEEE Trans. Image Process.* **2017**, *26*, 4331–4346. [CrossRef]
29. Fan, J.; Zhou, N.; Peng, J.; Gao, L. Hierarchical learning of tree classifiers for large-scale plant species identification. *IEEE Trans. Image Process.* **2015**, *24*, 4172–4184. [CrossRef] [PubMed]
30. Caliński, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **1974**, *3*, 1–27. [CrossRef]
31. Tao, Z.; Liu, H.; Li, S.; Fu, Y. Robust spectral ensemble clustering. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, Indianapolis, IN, USA, 24–28 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 367–376.
32. Tong, X.; Xia, G.S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Land-cover classification with high resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* **2020**, *237*, 111322. [CrossRef]
33. 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]

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.