forests MDPI

Article

DASR-Net: Land Cover Classification Methods for Hybrid Multiattention Multispectral High Spectral Resolution Remote Sensing Imagery

Xuyang Li [1], Xiangsuo Fan [1,2,*], Jinlong Fan [3], Qi Li [1], Yuan Gao [1] and Xueqiang Zhao [4,5]

[1] School of Automation, Guangxi University of Science and Technology, Liuzhou 545006, China; 221077062@stdmail.gxust.edu.cn (X.L.); liqi@gxust.edu.cn (Q.L.); gaoyuan@gxust.edu.cn (Y.G.)
[2] Guangxi Collaborative Innovation Centre for Earthmoving Machinery, Guangxi University of Science and Technology, Liuzhou 545006, China
[3] National Satellite Meteorological Center, China Meteorological Administration, Beijing 100081, China; fanjl@cma.gov.cn
[4] School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China; zhaoxq28@mail2.sysu.edu.cn
[5] China Water Resources Pearl River Planning Surveying and Designing Co., Ltd., Guangzhou 510610, China
* Correspondence: 100002085@gxust.edu.cn

Abstract: The prompt acquisition of precise land cover categorization data is indispensable for the strategic development of contemporary farming practices, especially within the realm of forestry oversight and preservation. Forests are complex ecosystems that require precise monitoring to assess their health, biodiversity, and response to environmental changes. The existing methods for classifying remotely sensed imagery often encounter challenges due to the intricate spacing of feature classes, intraclass diversity, and interclass similarity, which can lead to weak perceptual ability, insufficient feature expression, and a lack of distinction when classifying forested areas at various scales. In this study, we introduce the DASR-Net algorithm, which integrates a dual attention network (DAN) in parallel with the Residual Network (ResNet) to enhance land cover classification, specifically focusing on improving the classification of forested regions. The dual attention mechanism within DASR-Net is designed to address the complexities inherent in forested landscapes by effectively capturing multiscale semantic information. This is achieved through multiscale null attention, which allows for the detailed examination of forest structures across different scales, and channel attention, which assigns weights to each channel to enhance feature expression using an improved BSE-ResNet bilinear approach. The two-channel parallel architecture of DASR-Net is particularly adept at resolving structural differences within forested areas, thereby avoiding information loss and the excessive fusion of features that can occur with traditional methods. This results in a more discriminative classification of remote sensing imagery, which is essential for accurate forest monitoring and management. To assess the efficacy of DASR-Net, we carried out tests with 10m Sentinel-2 multispectral remote sensing images over the Heshan District, which is renowned for its varied forestry. The findings reveal that the DASR-Net algorithm attains an accuracy rate of 96.36%, outperforming classical neural network models and the transformer (ViT) model. This demonstrates the scientific robustness and promise of the DASR-Net model in assisting with automatic object recognition for precise forest classification. Furthermore, we emphasize the relevance of our proposed model to hyperspectral datasets, which are frequently utilized in agricultural and forest classification tasks. DASR-Net's enhanced feature extraction and classification capabilities are particularly advantageous for hyperspectral data, where the rich spectral information can be effectively harnessed to differentiate between various forest types and conditions. By doing so, DASR-Net contributes to advancing remote sensing applications in forest monitoring, supporting sustainable forestry practices and environmental conservation efforts. The findings of this study have significant practical implications for urban forestry management. The DASR-Net algorithm can enhance the accuracy of forest cover classification, aiding urban planners in better understanding and monitoring the status of urban forests. This, in turn, facilitates the development of effective

check for updates

Citation: Li, X.; Fan, X.; Fan, J.; Li, Q.; Gao, Y.; Zhao, X. DASR-Net: Land Cover Classification Methods for Hybrid Multiattention Multispectral High Spectral Resolution Remote Sensing Imagery. *Forests* **2024**, *15*, 1826. https://doi.org/10.3390/f15101826

Academic Editor: Rafael María Navarro-Cerrillo

Received: 28 September 2024
Revised: 17 October 2024
Accepted: 18 October 2024
Published: 19 October 2024

forest conservation and restoration strategies, promoting the sustainable development of the urban ecological environment.

Keywords: multiscale hollow attention; BSE-ResNe; dual-channel parallel architecture; sustainable forestry

1. Introduction

Modern remote sensing (RS) technology provides an extraordinary volume of Earth observation information, including a diverse range of satellite imagery and LiDAR datasets. These essential tools are vital for worldwide environmental oversight and are key in a variety of uses such as identifying land cover types, tracking alterations in landscapes, and observing and evaluating the impact of natural calamities [1–4]. The precision of land cover classification is of particular significance in the context of forest management and conservation, where detailed and accurate information is essential for a multitude of forestry-related tasks. Forests are complex and dynamic ecosystems that require meticulous monitoring to support activities like refined agricultural delineation, earth observation, regional environmental protection, and urban planning [5–8]. Moreover, the integration of satellite imagery and LiDAR data provides a comprehensive view of forest structure [9], which is crucial for understanding the habitat suitability for wildlife, assessing the impact of climate change on forest ecosystems, and planning sustainable forestry practices. The precision in land cover classification empowers scholars and decision-makers to make educated choices about preservation initiatives, afforestation schemes, and the reduction in biodiversity decline in wooded regions. As a result, progress in remote sensing techniques for categorizing land cover significantly enhances comprehensive environmental surveillance and is directly relevant to the sustainable administration and conservation of forest environments.

The classification of remote sensing imagery entails the recognition and grouping of individual pixels or areas within the image according to their spectral attributes, thereby allocating them to distinct categories. This procedure includes choosing suitable feature parameters to sort image components into separate, nonintersecting classification domains. In land cover classification, expert-deciphered category features, pixel-level category features, and object-oriented features can be utilized for classification, and these feature types can complement each other. Conventional approaches predominantly rely on surface attributes like pixel dimensions, form, hue, and texture for the categorization of images [10–14]. The raw spectral features of an image can be singleband or multiband images. Since multiband images contain richer information, they usually achieve better classification results. However, it is not the case that the higher the number of bands is, the better the classification effect. Having too many bands will not provide richer information but will lead to data redundancy [15]. A common practice is to obtain a vegetation index by linear or nonlinear operations, which reflects the image characteristics of two or more bands. Often, several raw bands and various index bands are combined for classification. Moreover, machine learning techniques including support vector machines (SVMs) [16], random forests (RFs) [17], K-means clustering (K-Means) [18], and K-nearest neighbor (KNN) algorithms [19] are applied for the classification of remote sensing imagery. Nonetheless, these standard methods require enhancement to address more intricate challenges.

Attributed to their ability to engage in hierarchical learning, these techniques are proficient in depicting complex nonlinear associations [20], which are instrumental in tasks such as categorization, the integration of data, and the reduction in dimensions [21–23]. In the domain of land cover classification, deep learning has achieved promising outcomes, particularly with models like U-Net [24], capable of yielding robust classification outcomes despite there being a smaller dataset for training. Several investigations have utilized adjusted loss functions and augmentation strategies to enhance model resilience against class imbalance issues in datasets [25]. Moreover, advanced deep learning approaches have been tailored for specialized imaging tasks, including a refined U-Net model for medical and brain tumor segmentation purposes [26–28]. When dealing with intricate or diverse data types, U-Net and its variants might exhibit a tendency towards biased feature extraction, which can result in significant discrepancies in predictive outcomes [29].

To overcome the hurdles associated with the intricate and high-definition nature of remote sensing imagery, both convolutional neural networks (CNNs) [30] and recurrent neural networks (RNNs) [31] have become prevalent tools in the realm of remote sensing image analysis. Research by Y. LeCun et al. [32] highlights that CNNs are particularly sensitive to variations in the rotation and scale of input images. Moreover, CNNs are not adept at capturing extended spatial relationships, leading to the integration of RNNs for tasks that demand long-term dependency management. A. Graves [33] has pointed out that RNNs grapple with issues like gradient disappearance and explosion during the processing of long-duration dependencies. The challenge of efficiently transmitting information across longer sequences can hinder the RNN's ability to learn these dependencies.

In addition, the absence of parallel processing in RNNs is a notable constraint. However, when performing deep neural network training, the gradient vanishing/exploding problem and network degradation problem are often encountered, which can negatively affect the network's training effectiveness and generalization ability. To address these problems, the residual neural network (ResNet) [34] introduces residual connectivity, which allows the network to learn a deeper level of feature representation. ResNet is able to train deeper networks and achieve higher accuracy after solving the problems of gradient vanishing and information loss. However, traditional ResNet does not fully consider the correlation and importance between feature channels, which may lead to the model's poor utilization of all feature channels [35]. To address this issue, the squeeze-and-excitation residual neural network (SE-ResNet) has been introduced to enhance the efficacy of remote sensing image processing. It incorporates the squeeze-and-excitation module, which allows for the dynamic refinement and prioritization of various channels within the feature map. This module can fine-tune the significance of each channel in the feature map, taking into account the overall channel content. It adaptively allocates weights by considering the global context of each channel, enabling the network to focus more on the channels that are crucial for the task at hand and to diminish the reliance on less relevant channels [36]. This ResNet-derived strategy has advanced the progress of semantic segmentation in remote sensing imagery. Rather than discarding ResNet, our aim is to devise a novel architecture that retains its benefits.

For the field of remote sensing imagery analysis, the adoption of the transformer architecture, particularly the Vision Transformer (ViT), has seen a surge in interest among scholars [37]. The transformer's self-attention mechanism allows it to effectively encode contextual relationships across the data, making it adept at handling the sequential nature of remote sensing images. The inclusion of positional encodings in the transformer design addresses the need to maintain the spatial order of pixels. The ViT model, pioneered by Google [38], has demonstrated the potential of transformer-based approaches for image classification tasks. In a related development, the work by Hong et al. [39] introduced the SpectralFormal network, which capitalizes on the spectral information within hyperspectral images to generate discriminative spectral embeddings, leading to improved classification results. Additionally, recent studies [40,41] have proposed ViT variants that employ weight-sharing strategies, allowing for consistent feature extraction across different parts of the image, thereby optimizing the use of available data. The advantage of this processing is that it enables the transformer to utilize the data more fully. Inspired by these excellent works, we use the DASR-Net algorithm that fuses the attention mechanism of the dual attention network (DAN) in parallel to the transformer to improve the DASR-Net algorithm of the ResNet network. The DASR-Net framework is applied to multispectral remote sensing data for land cover classification.

To address the problems of insufficient U-Net feature expression capability and the inconspicuous CNN differentiation of non-high-resolution images when classifying remote sensing images under different scale targets, a new semantic segmentation network framework for RS images, DASR-Net, is proposed, where one branch is a VIT network that uses the DAN attention mechanism for VIT networks, and the other strand is a modified version of ResNet that forms a parallel dual encoder structure. The DAN module, a pivotal part of our VIT network, is crafted to extract features with enhanced segmentation powers. Utilizing this module, we integrate features derived from a refined ResNet architecture in a parallel branch. This fusion strategy allows for a more comprehensive grasp of diverse image characteristics, enabling precise pixel-level classification in remote sensing imagery.

The primary advancements presented in this study are summarized as follows:

(1) The attention mechanism of DAN is constructed to focus on fully utilizing the correlation between features and the importance of channels, thus alleviating the problem of insufficient feature expressiveness when classifying remote sensing images under different scale targets. In addition, DAN compensates for the problem of excessive computational cost of VIT due to its global modeling capability.

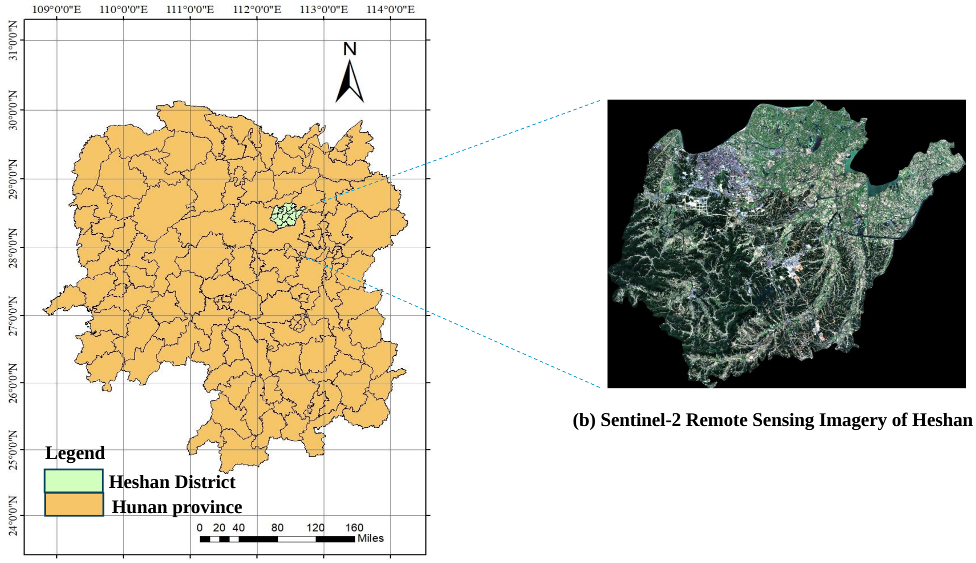
(2) A BSE-ResNet network is constructed to allow information to propagate more freely through the network. This architecture enables BSE-ResNet to adeptly capture the nuanced features within the original image, concurrently reducing the loss of fine details.

(3) The dual-channel parallel architecture is implemented to address structural discrepancies, aiming to prevent information loss and an overabundance of feature fusion. This architecture is particularly effective in recognizing feature types that exhibit high similarity, ensuring that subtle differences are preserved and accurately identified during the classification process.

2. Materials and Methods

2.1. Study Area Overview

Heshan District, nestled in the northern-central part of Hunan Province, lies at the western bank of Dongting Lake and the lower reaches of the Zishui River. It spans a geographical range from approximately 28°16′ N to 28°53′ N in latitude and from 112°11′ E to 112°43′ E in longitude. The research area is illustrated in Figure 1.



(b) Sentinel-2 Remote Sensing Imagery of Heshan District

(a) Heshan District Cartography

Figure 1. Sentinel-2 remote sensing image of the study area. (**a**) Heshan District cartography: displays the geographical layout of Heshan District in Hunan Province. (**b**) Sentinel-2 remote sensing imagery of Heshan District: Shows the different land cover types and their spatial distribution.

2.2. Preliminary Processing of Remote Sensing Imagery

On-site land cover data are vital for generating robust training and validation samples for remote sensing image analysis. The reliability of these data is key to the accuracy of classification results. During October 2021, we conducted field research to examine various land use categories. We targeted aquaculture and agricultural plots over 100 square meters for sampling, aiming to enhance the quality of training and validation samples.

The research presented herein utilizes a 10 m resolution multispectral image captured by the Sentinel-2 satellite on 6 October 2021. To prepare the data, we initiated preprocessing with atmospheric correction via the Sen2Cor tool to eliminate atmospheric interference from the imagery. Subsequent steps included image enhancement and the creation of a mask for feature isolation. In addition, to maintain consistent resolution, we resampled the bands using SNAP8.0 software to ensure that they had the same 10 m resolution. For the needs of this study, we selected four standard bands—Band 2, Band 3, Band 4, and Band 8—for our analysis, utilizing ENVI5.3 software to synthesize these bands. These bands span the visible and near-infrared spectrum, proving valuable for distinguishing and classifying features. During the sample data collection phase, we delineated regions of interest (ROIs) and assembled a corresponding sample library in the Heshan District of Yiyang City, informed by prior knowledge and field investigations, as depicted in Figure 2. Through these remote sensing data processing and analytical techniques, we secured high-caliber training and validation samples, underpinning the classification precision and the reliability of feature type identification within the study region.

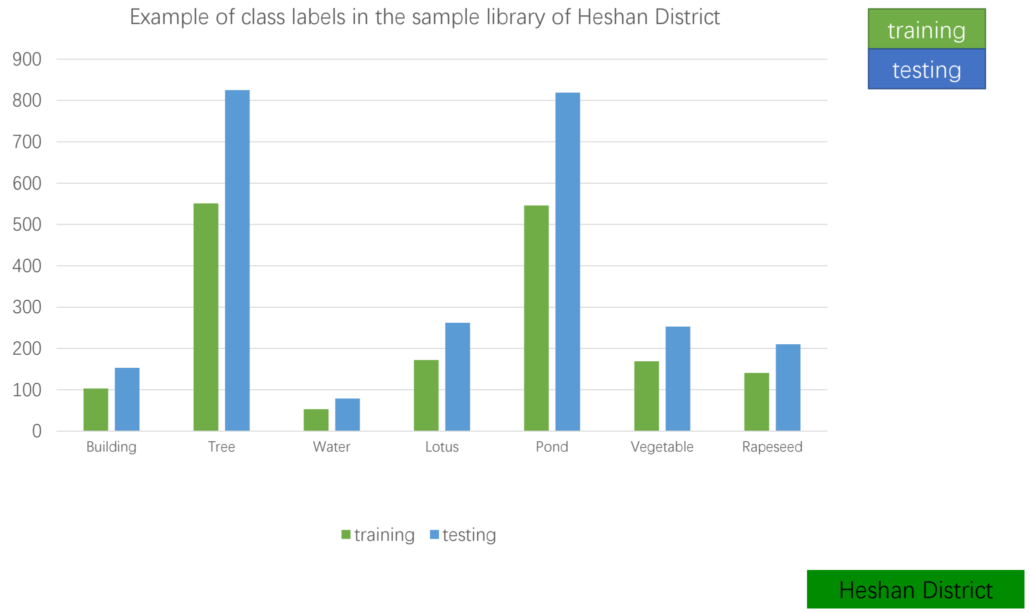


Figure 2. Dataset training and testing sample ratio.

2.3. DASR-Net Model

Aiming at the problem that the perception ability of the traditional model for targets of different scales needs to be improved, the expression ability of the features is not strong, and the differentiation is not obvious, this work presents the DASR-Net, a novel architecture that merges the DAN attention mechanism with a refined residual neural network, as shown in Figure 3. The DASR-Net is designed to include NDVI, NDWI, and BSE-ResNet, and incorporates the DAN attention mechanism within a transformer framework. The NDVI and NDWI spectral data are fused at the transformer's input stage. Encoder-derived features from the transformer are fused with those from the optimized residual network, and the concatenated features are compressed by a dense layer. Classification is performed using the RELU activation and a 1 × 1 convolutional layer, facilitating detailed pixel-wise classification for multispectral imagery.

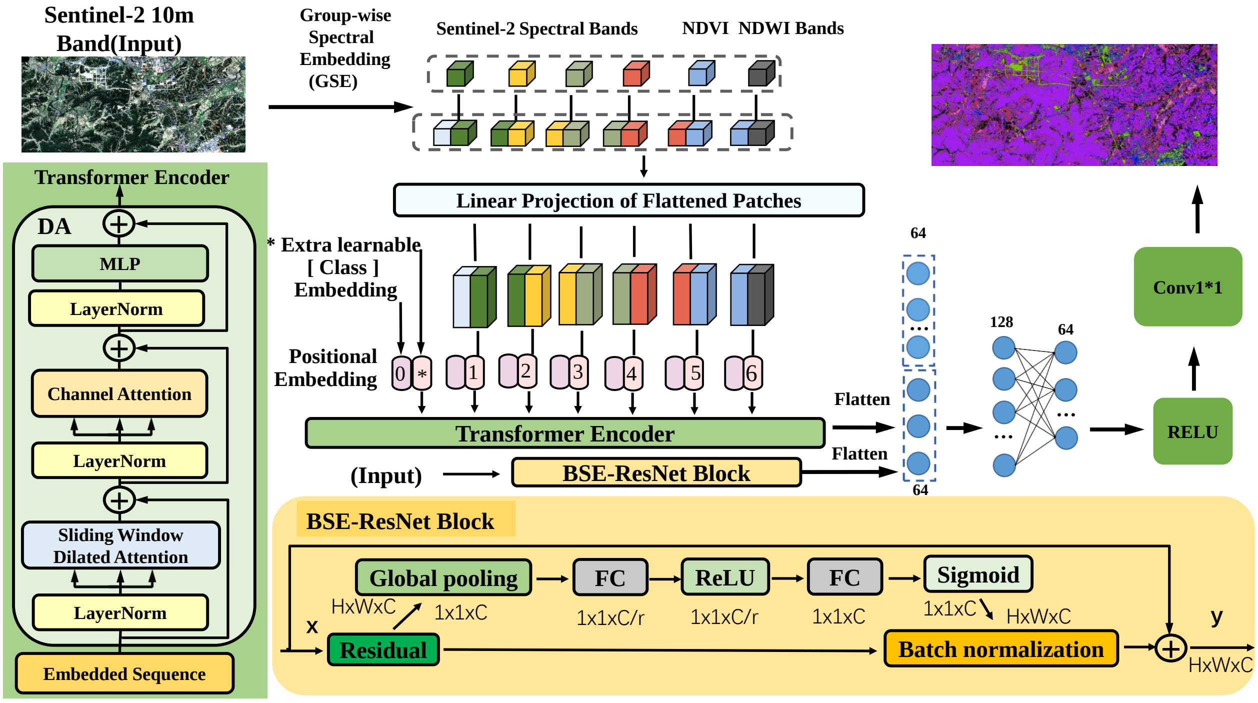


Figure 3. Schematic of the DASR-Net architecture for the hyperspectral multispectral image classification task.

2.3.1. Groupwise Spectral Embedding (GSE)

The spectral data within the remote sensing imagery, obtained from diverse locations, indicate the absorption characteristics at various wavelengths. It is vital to detect the subtle variations in these spectral signatures for effective feature classification. Even though multispectral images possess a more limited band range than hyperspectral images, there is still a need for stronger correlation among the available bands.

Suppose that we input a sequence of 1D pixels $x = [x_1, x_2, x_3, \cdots, x_m] \in \mathbb{R}^{1\times m}$. The input of the transformer is obtained by the calculation of Equation (1).

$$A = wx \tag{1}$$

where $w \in \mathbb{R}^{d\times 1}$ represents the linear mapping that is applied across all spectral channels in the sequence, with $A \in \mathbb{R}^{d\times m}$ aggregating the resulting feature vectors. The Generalized Spectral Embedding (GSE) is expressed in Equation (2):

$$\dot{A} = \mathrm{WX} \tag{2}$$

In which $W \in \mathbb{R}^{d\times n}$signifies the linear transformation matrix, $X \in \mathbb{R}^{n\times m}$ denotes the matrix of spectral attributes, and n indicates the count of adjacent spectral bands. We divided the pixel sequence into six 1 × 1 sequences and generated six d × 1 sequences in the BSE-RseNet branch according to different neighboring band settings. We chose the optimal value of d = 2, determined by the experimental accuracy and the precision of the prediction map. Within the transformer branch, we employed half of each pair of contiguous sequences for fusion, thereby enhancing the interband correlation. For a comprehensive illustration, see Figure 4.

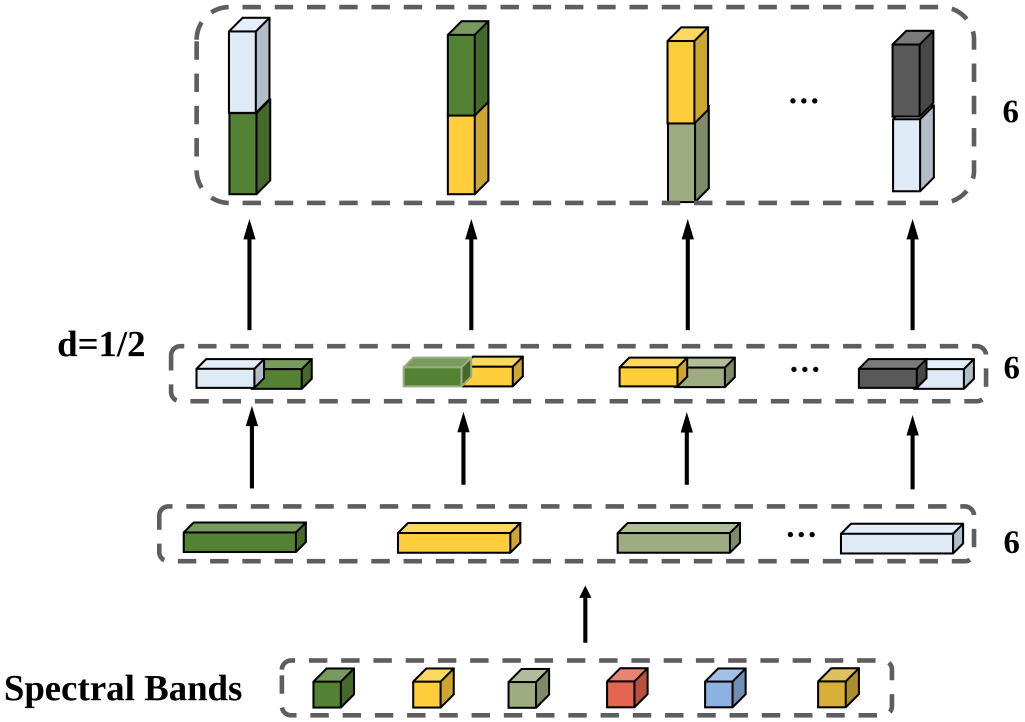


Figure 4. Visual representation of the evolution in feature embedding via collective spectral embedding.

2.3.2. Dual Attention Network (DAN)

The dual attention network enhances the feature representation of agricultural remote sensing images by taking the output of DilateFormer as the input and calculating the weight of each channel using the channel attention mechanism. The significance of various channels in feature representation can be modified through the channel attention allocation mechanism. By conducting weighted multiplication with the initial input, a weighted feature representation is derived. Such a feature representation has a stronger segmentation capability and further enhances the performance of the whole model.

First, DilateFormer adopts the "Dilated Spatial Encoding" method to expand the receptive field of the transformer to effectively capture local and global contextual information in remote sensing images. This approach incorporates multiscale cavity convolution within the transformer architecture to enlarge the receptive field, thereby enhancing the feature extraction capabilities for agricultural remote sensing imagery. Additionally, DilateFormer introduces an adaptive sliding window dilated attention (SWDA) mechanism for adjusting the attention weights between each pixel according to the surrounding pixels to address complex background and target situations in agricultural remote sensing images, as shown in Figure 5.

DilateFormer provides an effective solution to solve the long-range dependence problem in agricultural remote sensing. At the same time, the model maintains good computational efficiency and adapts to inputs of different scales and resolutions, and the DilateFormer algorithm improves the ability to recognize the differences between various remote sensing feature categories by introducing the channel attention module. The channel attention module can automatically adjust the importance of different channels to distinguish specific feature classes more accurately. This is achieved by learning and utilizing the information redundancy between bands in remote sensing images, which helps to reduce the influence of redundant information and thus improves the robustness and generalization ability of the model. The DilateFormer model connects the output slices and then performs feature aggregation through a linear layer. This approach successfully solves the long-range dependency problem while maintaining good computational efficiency. Precision and robustness are vital for remote sensing image segmentation. By appending a channel attention module following the DilateFormer, the interrelation between the features and the significance of the individual channels can be effectively harnessed, thereby advancing the accuracy and robustness of the remote sensing image segmentation models.

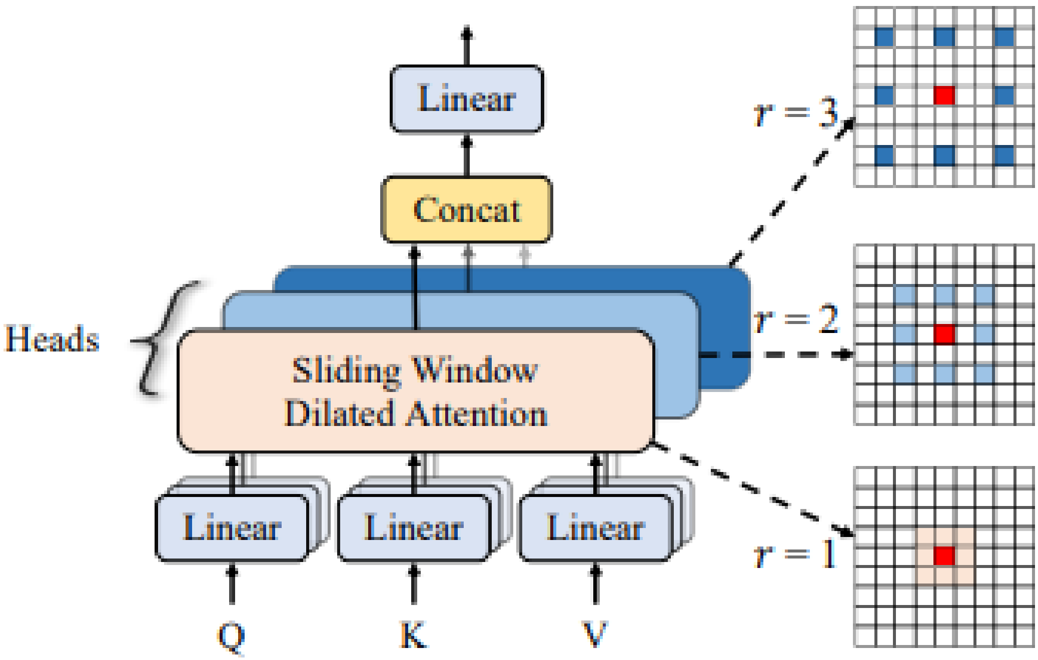


Figure 5. DilateFormer attention mechanism.

Employing the DAN model with its channel attention mechanism significantly enhances the precision and consistency of segmentation in agricultural remote sensing images. This approach provides an advanced technique for the analysis of such imagery and is poised to further its impact in the agricultural remote sensing as technology progresses.

DAN mechanism represents an innovative network architecture that enhances feature extraction by integrating two synergistic attention mechanisms: spatial attention and channel attention. This strategy significantly boosts the model's proficiency in handling hyperspectral and multispectral data. The spatial attention mechanism is aimed at identifying key regions within the input image that are crucial for classification tasks. It accomplishes this through an attention map that highlights areas deserving focus while suppressing irrelevant or noisy regions. By concentrating on areas containing valuable information, the spatial attention mechanism helps to improve the signal-to-noise ratio in the feature maps and reduces interference from noise. Once these significant regions are identified, the spatial attention adjusts their weights to enhance the feature representation, allowing the network to capture more details of these areas. On the other hand, the channel attention mechanism adaptively recalibrates feature responses by assigning different weights to various feature channels. In hyperspectral and multispectral data, each channel represents a different spectral band, which may contain varying amounts of information. Channel attention aids in identifying which channels contain more discriminative information. By amplifying the feature responses of these channels, the accuracy of classification can be improved. Through channel weighting, the channel attention mechanism can modulate the intensity of the feature maps, ensuring that the network focuses on channels rich in information and overlooks those with less. The DAN, with its dual attention mechanism, performs feature fusion across both spatial and channel dimensions, more effectively extracting the useful information from the data.

2.3.3. BSE-ResNet

This study introduces a detailed shallow-structured BSE-ResNet model that excels in capturing global contextual details across various scales, with an emphasis on channel information. Within its deep layers, abundant in semantic content, the model assesses channel-wise class relationships, aiming to enhance intraclass cohesion and interclass semantic distinction, thereby boosting the model's overall generalization capabilities.

The SE module's concept involves training parameters that can modify and prioritize channel significance within a feature map. By considering the overall channel data, it assigns weights in real-time, allowing the network to focus on channels crucial for the task at hand and to diminish the impact of less relevant channels. Moreover, in the realm of agricultural remote sensing, the BSE-ResNet model can be applied to the examination and interpretation of crop imagery. Utilizing the model's shallow architecture and fine detail, it can effectively encapsulate the global context about crop health and soil conditions, emphasizing channel relevance. This enhances the precision of crop growth analysis and aids in making informed agricultural decisions, as depicted in the Figure 6.

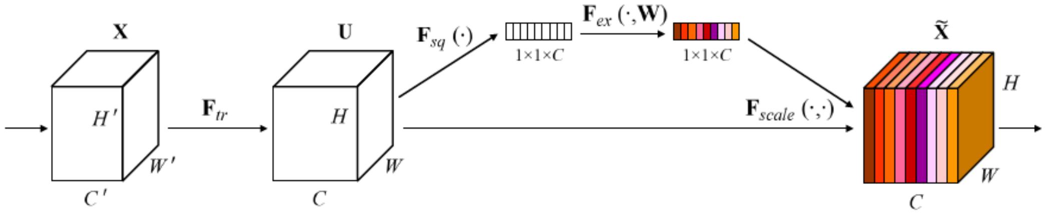


Figure 6. SE module.

The SE module is of great significance in applying remote sensing in agriculture. The module mainly consists of two phases: the squeeze phase and the excitation phase. In the squeeze phase, the SE module converts the feature maps of each channel into a single value through the global average pooling operation to achieve the purpose of integrating the channel feature information. This method not only concentrates on particular feature subsets within individual channels but also merges comprehensive contextual information. During the excitation stage, the SE module acquires the channel weights via two fully connected layers. The initial fully connected layer serves to downscale the dimensionality, thereby enhancing the module's computational efficiency. The second fully connected layer is used to learn the channel weights by mapping the individual values obtained earlier to a weight vector equal to the number of input channels. Finally, a weighted summation operation is performed on the features of each channel based on the learned weight vectors, and the weighted feature map is obtained as the output of the SE module. SE-ResNet is a deep learning model that combines the residual network (ResNet) and squeeze-and-excitation (SE) modules. The model borrows the idea of residual connectivity in ResNet, which solves the gradient vanishing and exploding problems in deep networks by jump connectivity and improves the propagation and retention of features.

In agricultural applications, the scale operation is important to adjust the distribution of the output feature maps of each residual block in the SE-ResNet network. The scale operation scales and translates the feature maps by learning the learnable parameters (gamma and beta) to fit the feature distributions of different layers and channels. In addition, batch normalization can also achieve similar functions as scale operation and can optimize the stability and convergence of the network through the learned parameters. To further improve the generalization performance of the model and to increase stability and accelerate convergence during training, we introduce regularization techniques and use batch normalization in BSE-ResNet to reduce the risk of overfitting. This BSE-ResNet model combined with a regularization technique has broad application prospects in agricultural remote sensing, as shown in Figure 7.

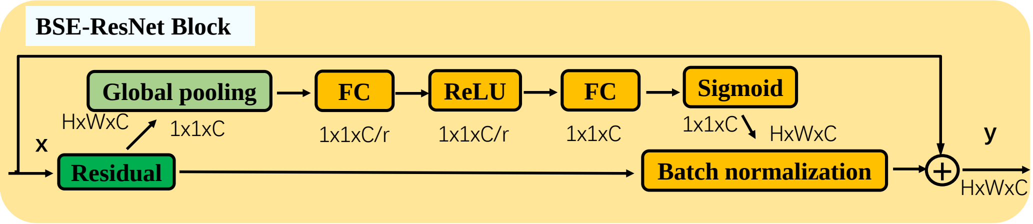


Figure 7. Schematic diagram of the improved BSE-ResNet structure based on the residual structure.

2.3.4. Assessment System

The three assessment criteria derived from the confusion matrix for the validation of multispectral pixel classification are as follows: total accuracy (OA), mean precision (AA), and the Kappa statistic.

(1) The calculation formula of OA is given in Equation (3) below. Give the Formula (3),

$$OA = \frac{Tq + Tp}{Tq + Fq + Tp + Fp} \tag{3}$$

where T_q is true positive, the number of true examples, F_q is false positive, the number of false-positive examples, T_p is true negative, the number of true negative examples, and F_q is false negative, the number of false-negative examples.

(2) AA is the average precision, which is a more accurate evaluation index in agricultural classification. Its calculation Formula (4) is as follows.

$$AA = \frac{1}{Y}\frac{Tq + Tp}{Tp + Fq + Tq + Fp} \tag{4}$$

where Y denotes the number of categories.

(3) The Kappa statistic evaluates the level of agreement between the actual and predicted classifications of an agricultural model in practical testing scenarios. It is calculated using the following Formula (5).

$$Kappa \ = \frac{P_x - P_y}{1 - P_y} \tag{5}$$

P_x denotes the exact match of the observed data, while P_y indicates the probability that the classifier will yield a concordant prediction, aligning the classification outcome with the actual Ground Truth.

3. Results

3.1. Feature Elimination Analysis

The results from the experimental assessments on the dataset for the research area confirm the effectiveness of the network design put forth in this manuscript. The findings are detailed in Table 1. The baseline ViT model achieved an overall accuracy of 94.63%, indicating its suitability for multispectral image categorization. The integration of the DAN module with ViT (ViT+ DAN) improved the overall accuracy to 96.76%, validating the enhancement brought by the DAN mechanism in channel interaction. The SE-ResNet module's addition to ViT showed a better OA than the standalone ViT, with concurrent improvements in the average accuracy and Kappa score, hinting at SE ResNet's potential to refine ViT's categorization for specific classes in multispectral imagery. The combination of ViT with the DAN module, including GSE band data, led to an average accuracy of 95.19%, underscoring the advantage of this feature fusion for multispectral image pixel classification. The SE-ResNet model supplemented with the GSE module also saw a boost in OA. Altering ViT's attention to DAN's, in parallel with SE-ResNet, and then merging the features, resulted in an overall accuracy of 95.27%. Nevertheless, the DASR-

Net outperformed with an OA of 96.36% when all the proposed modules were combined. This demonstrates that the modules developed in this study are effective for classifying agricultural land cover and contribute to refining the classification accuracy.

Table 1. Results of ablation experiments on the study area dataset using DASR-Net with different module combinations. (Bold represents the best in the same type).

Different Methods	Different Module			Metric			Time (s)
	GSE	DAN	SEReNet	OA (%)	AA (%)	Kappa	
ViT	×	×	×	94.63	91.44	0.9303	**984.1**
DASR -Net	✓	×	×	94.89	91.57	0.9336	1173.24
DASR -Net	×	✓	×	95.09	91.10	0.9362	1004.17
DASR -Net	×	×	✓	95.09	91.17	0.9362	1480.55
DASR -Net	×	✓	✓	95.19	91.62	0.9375	1442.84
DASR -Net	✓	✓	×	95.22	92.05	0.9379	1131.63
DASR -Net	✓	×	✓	95.27	92.43	0.9385	1681.18
DASR -Net	✓	✓	✓	**96.36**	**94.47**	**0.9527**	1109.14

To assess the influence of training sample size on experimental outcomes, we randomly drew samples ranging from 10% to 90% from the Heshan District dataset for training and validation, with the remainder used for testing. Table 2 illustrates the results, indicating that a higher sample proportion does not necessarily lead to improved accuracy and that the noise is randomly distributed. Consequently, we opt for a 40% sample size for our experiment to attain the optimal classification performance.

Table 2. The results of the model proposed in this paper under different proportions. (Bold represents the best in the same type).

Ratio of Training	Class No.							Metrics			Time (s)
	1	2	3	4	5	6	7	OA (%)	AA (%)	Kappa	
10%	84.00	**99.27**	92.30	95.34	**99.26**	80.95	**97.14**	95.82	92.61	0.9457	**784.48**
20%	86.27	97.45	**96.15**	94.25	98.90	**90.47**	**97.14**	96.19	94.38	0.9506	1013.04
30%	86.84	97.81	87.17	95.41	98.53	88.88	93.33	95.61	92.57	0.9430	1128.57
40%	**96.07**	99.09	94.23	**96.00**	98.53	84.52	92.85	**96.36**	**94.47**	**0.9527**	1109.14
50%	84.37	98.69	87.87	91.78	98.38	79.62	90.85	94.24	90.23	0.9249	1207.31
60%	95.31	92.05	93.90	86.92	97.93	87.34	90.07	98.90	89.32	0.9380	1124.42
70%	92.17	98.54	93.47	91.83	98.74	88.13	94.28	96.05	93.89	0.9487	1109.14
80%	87.74	98.45	88.57	90.57	98.90	87.83	95.35	95.59	92.49	0.9426	1078.08
90%	90.43	98.38	89.83	93.90	98.77	86.54	94.60	95.87	93.21	0.9464	826.18

Given that the Sentinel-2 satellite captures the data of Mt. Heshan with 13 spectral bands while other satellites may not provide as many, despite all of them including RGB and NIR bands, researchers often select a subset of 4 common bands (RGB + NIR) from the 13 to enhance the dataset's generalizability for their experimental analysis. To investigate if utilizing additional Sentinel-2 bands aids in the precise classification of land cover types within the study area, we initially included the visible and near-infrared (VNIR) bands in our experiment. Subsequently, we also incorporated the shortwave infrared (SWIR) bands, along with all the available bands for comparative purposes. The findings are presented in Table 3.

Table 3. The results of the two models across different spectral bands. (Bold represents the best in the same type).

Different Bands (Method)	Class No.							Metrics			Time (s)
	1	2	3	4	5	6	7	OA (%)	AA (%)	Kappa	
4 bands (ViT)	86.59	97.50	88.04	87.25	98.42	87.11	**95.10**	94.63	91.44	0.9303	984.10
4 bands (DASR-Net)	**92.17**	98.54	**93.47**	91.83	**98.74**	**88.13**	94.28	**96.36**	**94.47**	**0.9527**	1109.14
4 bands + VNIR8 (ViT)	87.15	97.92	88.04	**95.09**	97.06	81.35	93.87	94.50	91.50	0.9286	**693.47**
4 bands + VNIR8 (DASR-Net)	88.26	**98.75**	86.95	93.79	98.53	85.42	93.87	95.52	92.23	0.9417	810.30
4 bands + SWIR 7(ViT)	86.03	98.44	86.95	94.44	98.63	79.32	92.65	94.70	90.93	0.9310	986.77
4 bands + SWIR7 (DASR-Net)	91.62	98.54	**93.47**	**95.09**	98.53	85.76	93.06	95.95	93.73	0.9306	1122.55
Full bands (ViT)	89.38	98.33	82.60	93.79	98.32	83.38	93.87	95.06	91.39	0.9358	862.45
Full bands (DASR-Net)	86.59	98.54	89.13	90.52	98.53	85.08	91.83	94.89	91.46	0.9336	1003.46

3.2. Multi-Method Comparison

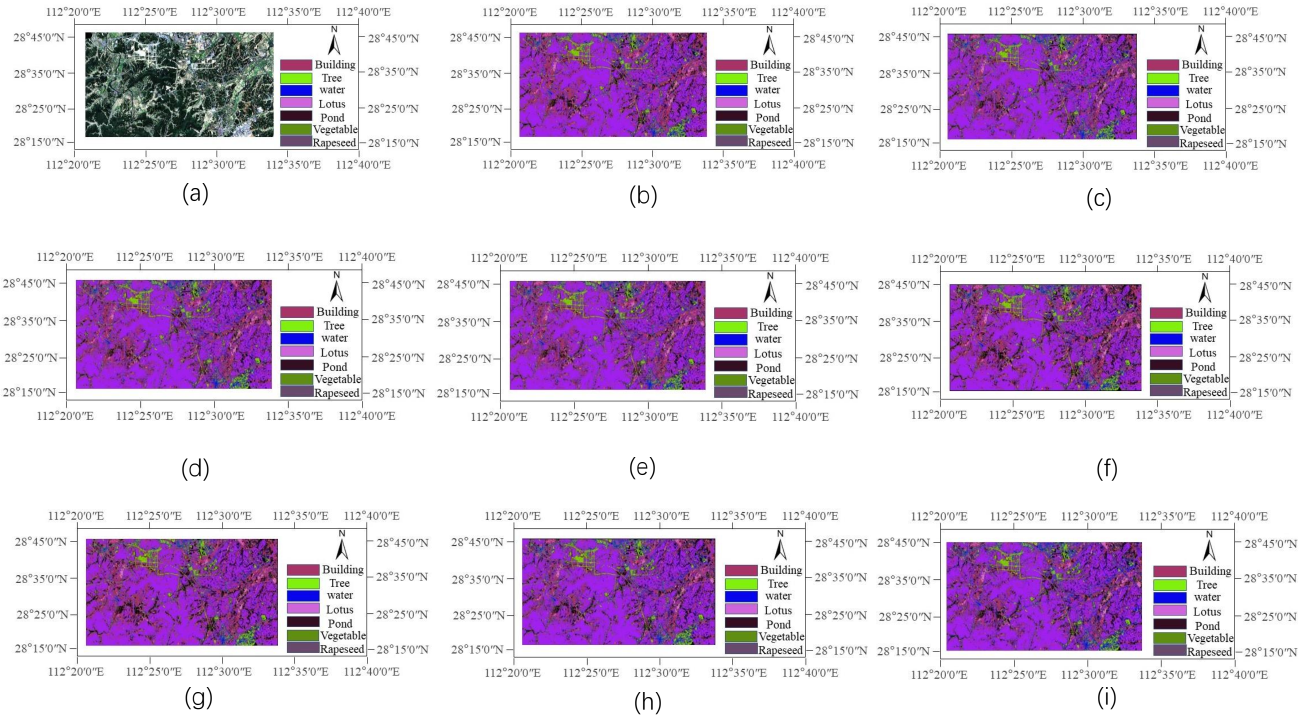
3.2.1. Comparative Analysis of Multispectral Data

The outcomes of the quantitative classification using the three combined metrics—OA, AA, and Kappa coefficient—along with the category-specific accuracies for the dataset of the Heshan District study area are presented in Table 4. Among the tested methods, CNN yielded the poorest overall performance, with lower OA, AA, and Kappa scores compared to the other models. Specifically, the accuracies for water and rapeseed were notably low at 8.69% and 8.57%, respectively. This is likely due to the limited number of bands in multispectral imagery, leading to underfitting when using individual bands, which hampers the CNN's ability to extract meaningful features, whereas hyperspectral imagery provides 200 bands of information. Despite this, CNN excels in classifying hyperspectral images compared to most other methods. The conventional classifiers, SVM and KNN, yielded reasonably good results with OAs of 93.95% and 93.72%, respectively, but struggled with the classification of the building category. RNN, ViT, SF, and DASR-Net are all deep learning-based spectral sequence classification techniques, and their performances were closely aligned, highlighting the strength of deep learning in processing sequential data. DASR-Net outperformed the other comparative models in terms of OA, AA, and Kappa, and it also demonstrated superior performance across various categories, including building, tree, and vegetable.

Table 4. The results of various algorithms. (Bold represents the best in the same type).

C N.	Different Methods							
	SVM	KNN	RF	CNN	RNN	ViT	SF	DASR-Net
1	75.32	70.12	74.02	39.10	87.15	86.59	91.06	**92.17**
2	98.06	98.06	98.30	**98.96**	98.33	97.50	98.54	98.54
3	75.00	87.50	87.50	8.69	85.86	88.04	83.69	**93.47**
4	**96.21**	94.69	93.18	80.06	91.50	87.25	89.54	91.83
5	98.78	97.07	98.04	95.39	97.38	98.42	**99.16**	98.74
6	80.31	82.67	84.25	66.44	84.40	87.11	84.06	**88.13**
7	93.39	**95.28**	93.39	8.57	94.69	95.10	93.46	94.28
OA (%)	93.95	93.72	94.18	77.13	94.66	94.63	95.12	**96.36**
AA (%)	88.16	89.35	89.82	51.72	91.34	91.44	91.36	**94.47**
Kappa	0.9211	0.9184	0.9244	0.6971	0.9308	0.9303	0.9365	**0.9527**
time (s)	139.65	122.64	**38.35**	299.07	350.1	984.10	1047.64	1109.14

The classification maps for Heshan District generated by various models are displayed in Figure 8. The study area within the square box has been annotated based on prior knowledge and outdoor sampling data. In Figure 8, SVM misclassifies rapeseed entirely as vegetables, leading to evident confusion. The CNN model exhibits inferior classification outcomes across the entire study area, struggling to distinguish between categories with subtle differences within the dataset. The ViT, SF, and DASR-Net have better overall classification results, but locally, DASR-Net has better results. DASR-Net can better distinguish between rapeseed and vegetables with more minor intragroup differences and does not misclassify vegetables, and there is some improvement in the overall OA, AA, and Kappa, as well as better graphing than ViT, which is of great value.



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

Figure 8. Results of different algorithms. (**a**) Image. (**b**) SVM. (**c**) KNN. (**d**) RF. (**e**) CNN. (**f**) RNN. (**g**) Transformer (ViT). (**h**) SF. (**i**) DASR-Net.

In the right border of the prediction map where the various algorithms are compared, under the blue category label for "water" in each algorithm, only our proposed algorithm is able to accurately classify this "water" category, while the other algorithms fail to do so. This clearly demonstrates the superiority of our algorithm in marginal areas. We speculate that this is due to the DAN module in our algorithm.

3.2.2. Comparative Analysis of Hyperspectral Data

To evaluate whether the model introduced in this study yields superior qualitative outcomes in terms of generalizability and applicability for classifying agricultural land cover in hyperspectral imagery, we employ three agricultural-related hyperspectral datasets: the WHU-Hi-HanChuan dataset, the HyRANK-Loukia dataset, and the Pavia University dataset. These are used to benchmark against other contemporary and exemplary models within the field to ascertain the qualitative outcomes. Presented here are the findings from the HyRANK-Loukia, Pavia University, and WHU-Hi-HanChuan datasets.

(1) The HyRANK-Loukia dataset includes 176 spectral bands covering a wide range from visible to near-infrared with an image size of 249 × 945 pixels. It contains 13,503 manually tagged pixels for 14 land cover types. This is shown in Table 5 below.

(2) The Pavia University dataset, captured by the ROSIS sensor(ROSIS (Reflective Optics System Imaging Spectrometer) is developed and operated by the University Research Center of Iceland — the Centre for Remote Sensing (CRS) at the University of Reykjavik) in 2003 at the University of Pavia, Italy, measures 610 × 340 pixels with a 1.3 m GSD. It covers a spectral range of 430–860 nm, initially with 115 bands, 22 of which

are water vapor absorption bands and are excluded, leaving 93 bands; the details are in Table 6.

(3) The WHU-Hi-HanChuan dataset was gathered on 17 June 2016, between 17:57 and 18:46 in Hanchuan, Hubei, China, using a Leica Aibot X6 drone(The Leica Aibot X6 is manufactured by Aibotix GmbH, a company founded in 2010 with its headquarters located in Kassel, Germany) with a 17 mm Nano-Hyperspec sensor(The Nano-Hyperspec sensor is manufactured by Headwall Photonics, which is located in Fitchburg, MA, USA). The session occurred under clear skies, with a temperature around 30 °C and 70% humidity. The area, a mix of urban and agricultural, featured buildings, water, and farmlands with seven crop types. The drone flew at 250 m, capturing 1217 × 303 pixel images across 274 bands (400–1000 nm) with a spatial resolution of about 0.109 m. The dataset, shown in Table 7, includes many shadowed areas due to the angle of the low afternoon sun.

Table 5. Detailed division of label samples in the HyRANK-Loukia dataset. (The background color is the corresponding color for each category during classification).

No.	Name	Train.	Val.	Test.
1	Dense Urban Fabric	14	15	259
2	Mineral Extraction Sites	3	4	60
3	Non Irrigated Arable Land	27	27	488
4	Fruit Trees	4	4	71
5	Olive Groves	70	70	1261
6	Broad-leaved Forest	11	11	201
7	Coniferous Forest	25	25	450
8	Mixed Forest	54	53	965
9	Dense Sclerophyllous Vegetation	190	189	3414
10	Sparce Sclerophyllous Vegetation	140	140	2523
11	Sparcely Vegetated Areas	21	20	363
12	Rocks and Sand	24	25	438
13	Water	69	70	1254
14	Coastal Water	23	22	406
	Total	675	675	12,153

Table 6. Provides a detailed breakdown of the labeled sample distribution for the Pavia University dataset. (The background color is the corresponding color for each category during classification).

No.	Name	Train.	Val.	Test.
1	Asphalt	331	332	5568
2	Meadows	932	933	16,784
3	Gravel	105	105	1889
4	Trees	153	153	2758
5	Painted metal sheets	67	67	1211
6	Bare Soil	251	252	4526
7	Bitumen	67	66	1197
8	Self-Blocking Bricks	184	184	3314
9	Shadows	48	47	852
	Total	2138	2139	38,099

Table 7. Provides a detailed breakdown of the labeled sample distribution for the WHU-Hi-HanChuan dataset. (The background color is the corresponding color for each category during classification).

No.	Name	Train.	Val.	Test.
1	Strawberry	2236	2237	40,262
2	Cowpea	1137	1138	20,478
3	Soybean	514	515	9258
4	Sorghum	268	267	4818
5	Water spinach	60	60	1080
6	Watermelon	227	226	4080
7	Greens	295	295	5313
8	Trees	899	899	16,180
9	Grass	473	473	8522
10	Red roof	526	526	9464
11	Gray roof	845	846	15,220
12	Plastic	184	184	3311
13	Bare soil	456	456	8204
14	Road	928	928	16,704
15	Bright object	57	57	1022
16	Water	3770	3770	67,861
Total		12,875	12,877	231,777

Table 8 presents the classification outcomes for various models on the hyperspectral dataset, including quantitative metrics such as OA, AA, and Kappa coefficient. The bold figures in each row indicate the top performance for each category. Overall, the VIT model yields the least favorable results, with lower OA, AA, and Kappa scores compared to the other models. Figures 7–9 display the classification maps produced by different models on the HyRANK-Loukia, Pavia University, and WHU-Hi-HanChuan hyperspectral datasets, respectively.

Table 8. The performance of different algorithms on the three datasets. (Bold represents the best performance of this category).

High-Resolution Evaluation		Different Methods							
		SVM	KNN	RF	CNN	RNN	ViT	SF	DASR-Net
HyRANK-Loukia	OA (%)	73.63	79.42	77.59	84.15	78.07	75.82	77.31	**85.11**
	AA (%)	74.42	80.76	80.41	85.53	80.19	78.15	79.56	**86.56**
	Kappa	0.7141	0.7769	0.7625	0.8280	0.7625	0.7383	0.7541	**0.8386**
	inference (ITS)	306.12	322.32	**76.45**	437.52	196.33	230.18	321.37	221.07
WHU-Hi-HanChuan	OA (%)	55.32	60.56	69.66	71.74	53.27	50.64	75.38	**77.99**
	AA (%)	49.08	71.40	76.77	78.03	53.10	56.12	81.20	**85.46**
	Kappa	0.4916	0.5564	0.6576	0.6787	0.4673	0.4486	0.7192	**0.7508**
	inference (ITS)	180.93	179.65	**39.51**	240.57	114.84	137.58	201.27	135.16
Pavia University	OA (%)	71.97	70.83	69.28	81.93	78.35	68.83	74.95	**84.15**
	AA (%)	76.65	79.92	80.01	**86.21**	84.05	77.69	83.30	85.53
	Kappa	0.6320	0.6323	0.6196	0.7628	0.7223	0.6018	0.6797	**0.7901**
	inference (ITS)	380.12	506.32	**146.61**	520.72	330.23	224.08	377.77	231.01

Overall, conventional machine learning algorithms such as SVM, KNN, and RF do not excel across the three hyperspectral datasets, placing them in the mid- to lower-range of classification performance based on the OA, AA, and Kappa metrics. In contrast, deep learning approaches, particularly the traditional RNN, exhibit remarkable performance in the WHU-Hi-HanChuan and Pavia University datasets. RNN surpasses traditional classifiers in terms of qualitative metrics, highlighting the advantage of deep learning in land cover classification tasks. Furthermore, CNNs excel in extracting spatial features from extensive and continuous spectral data like hyperspectral imagery. Consequently, CNNs also achieve strong results on the three datasets, with the WHU-Hi-HanChuan dataset showcasing the highest accuracy for CNNs, second only to our proposed method.

Among them, inference (ITS) represents the scale of the time-series throughput during inference, which is a metric used to measure the time required for the model to make a single prediction, with the unit being seconds. It can be observed that, apart from the RF, our model is the fastest in terms of inference speed, indicating that our model exhibits excellent performance in terms of inference time efficiency.

Compared with other methods, its classification accuracy on the three hyperspectral datasets is undoubtedly optimal, which is also applicable to agricultural growing areas, as shown in Figures 9–11.

In Figures 9–11, a comparison of the algorithms against the Ground Truth is provided. In Figure 9, on the left side, algorithms c and d clearly show a confusion between the red category "Dense Urban Fabric" and the yellow category "Non Irrigated." In the central area, only our proposed algorithm is able to accurately distinguish the small red "Dense Urban Fabric" points without misclassifying them as green "Coniferous Forest" points. In Figure 10, in the center of the image, on the blue block representing "Painted Metal Sheets", there are misclassifications by algorithms c, d, e, h, and i, with varying degrees of confusion with the orange "Bitumen" small points. In the lower yellow block representing "Self-Blocking Bricks", misclassification and confusion are present in algorithms c, d, e, and f, with some "Meadows" small points incorrectly classified. In Figure 11, apart from our proposed algorithm, all other algorithms exhibit misclassification where the pink-bordered category "Soybean" incorrectly includes green "Grass" small points within its edges. This clearly demonstrates the superiority of our algorithm in terms of classification accuracy.

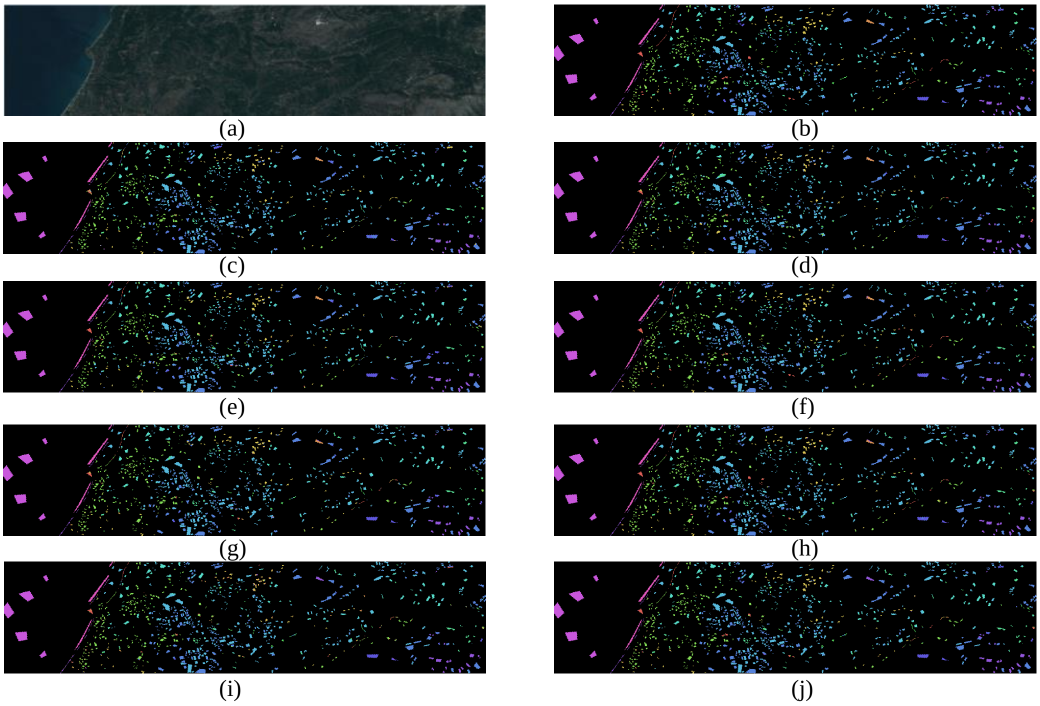


Figure 9. Classification maps of HyRANK-Loukia dataset ((**a**) Image. (**b**) Ground Truth. (**c**) SVM. (**d**) KNN. (**e**) RF. (**f**) CNN. (**g**) RNN. (**h**) Transformer (ViT). (**i**) SF. (**j**) DASR-Net).

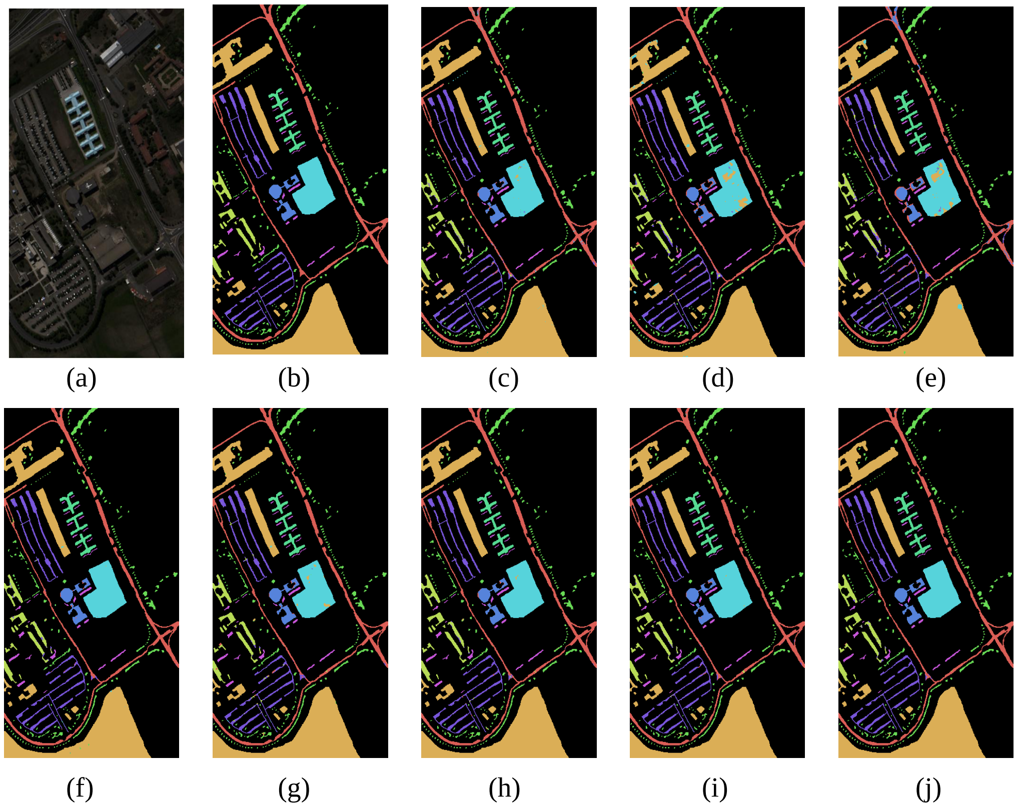


Figure 10. Classification maps of Pavia University dataset ((**a**) Image. (**b**) Ground Truth. (**c**) SVM. (**d**) KNN. (**e**) RF. (**f**) CNN. (**g**) RNN. (**h**) Transformer (ViT). (**i**) SF. (**j**) DASR-Net).

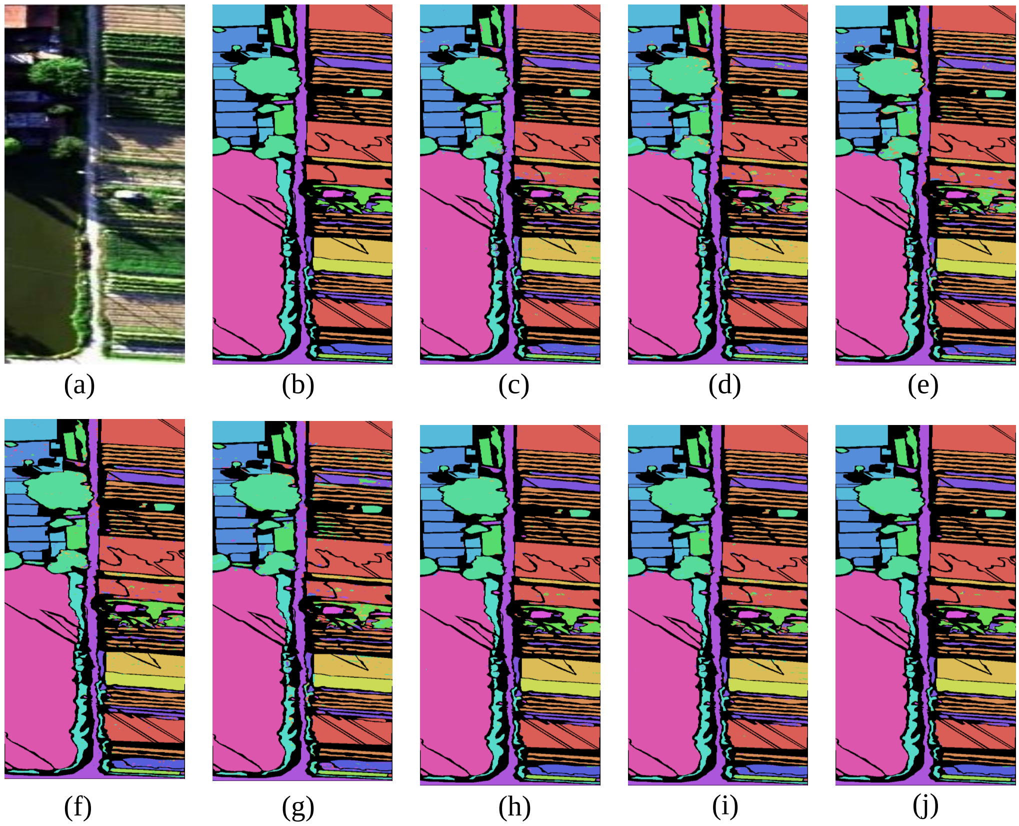


Figure 11. Classification maps of WHU-Hi-HanChuan dataset ((**a**) Image. (**b**) Ground Truth. (**c**) SVM. (**d**) KNN. (**e**) RF. (**f**) CNN. (**g**) RNN. (**h**) Transformer (ViT). (**i**) SF. (**j**) DASR-Net).

3.3. Analysis of Trees and Forests

Vegetation coverage serves as a pivotal indicator of the health and vitality of forest ecosystems. In the domain of remote sensing image processing, the analytical examination of forested areas and individual tree species is of paramount significance. The nuanced extraction and interpretation of remote sensing imagery afford us a wealth of information pertaining to the spatial distribution of forest resources, the physiological status of tree growth, and the extent of vegetation coverage, all of which are indispensable for the management of forestry resources, the surveillance of ecological conditions, and the investigation of climatic variations.

Within multispectral datasets, the delineation of tree populations is discernible, and the algorithmic approach developed by our research team exhibits superior performance in the classification of trees, exhibiting a notably reduced incidence of misclassification when contrasted with other methodologies. This enhanced discriminative accuracy is instrumental in the delineation of ecological function zones and the evaluation of environmental quality, providing a robust foundation for ecological planning and conservation strategies.

Moreover, the analysis of vegetation coverage trends contributes to the revelation of the temporal dynamics and evolutionary trajectories of regional ecological systems. In the context of the hyperspectral HyRANK-Loukia dataset, the precise classification of various species, including Fruit Trees and Mixed Forests exemplify the advanced capabilities of remote sensing technology in discerning subtle ecological distinctions. The integration of this classification with ancillary data such as topographical features and soil characteristics facilitates the advancement of precision forestry practices, thereby enhancing the efficacy of forest management and sustainability efforts.

Advanced analytical methods used in remote sensing image processing improve our knowledge of forest ecosystems and provide stakeholders with the essential tools for making data-driven decisions in forestry conservation and environmental management.

4. Discussion

Topography affects land suitability for crops and is key to forest health and diversity. Land class distribution correlates with local agriculture and forest structure, often clustering in specific areas. Our research introduces an efficient land cover classification method using hyperspectral and multispectral imagery, crucial for forest management. Remote sensing can face challenges like poor perception and feature expression at different scales, particularly in complex forests. Our DASR-Net model captures contextual information in a token-based framework, enriching features for more precise forest feature extraction through an adaptive attention mechanism.

Conventional classification techniques may not fully account for pixel interdependencies, resulting in less effective classification in intricate forest scenes. In contrast, deep learning models and our DASR-Net can effectively grasp the complex spatial and spectral characteristics of forest environments. This study evaluates the pros and cons of eight approaches—SVM, KNN, RF, CNN, RNN, ViT, SF, and DASR-Net—especially for multispectral and hyperspectral forest imagery. The findings show that our DASR-Net outperforms in classification, notably in differentiating forest types, assessing forest health, and identifying cover changes.

This progress holds significant guidance for the localized application of precision forestry practices, offering robust data backing for activities like forest inventory, biodiversity evaluation, and the monitoring of unauthorized logging. The improved classification precision delivered by DASR-Net serves as an invaluable resource for policymakers and forestry experts.

The reason why DASR-Net outperforms other models is primarily due to the DAN's spatial attention component, which focuses on the most relevant spatial areas of the data, highlighting regions that are crucial for classification while suppressing less important or noisy areas. This selective focus helps to improve the signal-to-noise ratio and enhances the network's ability to extract meaningful features from complex scenes. Additionally, the channel attention component of DAN adaptively recalibrates channel feature responses by assigning different weights to different channels. In hyperspectral and multispectral data, each channel represents a different spectral band, and this recalibration helps to emphasize the channels that contain more discriminative information.

The introduction of the DASR-Net algorithm has made significant contributions to the field of remote sensing and land cover classification, particularly in the realm of forest management and conservation. Unlike traditional methods that often struggle with the complex boundaries between feature categories and the similarities among different forest types, the DASR-Net algorithm is specifically designed with a dual attention mechanism to address these challenges. This innovation enables the refined analysis of forest areas, aiding in a clearer understanding of forest structure and condition. Previous studies may have overlooked the importance of capturing multiscale semantic information. DASR-Net meticulously examines forest structures at different scales through its multiscale spatial attention, providing a more comprehensive perspective for monitoring environmental changes and protecting biodiversity. The channel attention mechanism within DASR-Net employs an improved BSE-ResNet bilinear method, assigning weights to each channel to enhance feature representation, which is a departure from traditional methods that may suffer from insufficient feature expression and, consequently, lower classification discriminability. The approach of DASR-Net makes the classification of remote sensing imagery more distinctive. The findings of this study have direct practical significance for urban forestry management. By improving the accuracy of forest cover classification, DASR-Net assists urban planners in developing effective conservation and restoration strategies, which is crucial for the sustainable development of urban ecological environments.

5. Conclusions

In light of this text, we aim to harness global contextual information within remote sensing images to enhance the recognition of geographical attribute features, with a particular focus on forested environments. We introduce DASR-Net, a semantic segmentation framework that integrates ViT with a BSE-ResNet, utilizing a dual-encoder structure to capture the intricate details of forest landscapes. The attention mechanism module within our DASR-Net, termed DAN, exploits the intercorrelations between features and the significance of channels to guide the encoder toward more discriminative feature extraction. This is especially critical for forest applications where subtle differences between tree species or forest conditions must be discerned. The meticulously designed BSE-ResNet network is intended to encapsulate the nuanced features in the original image, which is essential for accurate forest classification and monitoring. The dual-channel parallelism of our framework establishes an architecture for feature information exchange that transcends the limitations of individual network windows. It addresses the issue of indistinct differentiation between various scales and effectively recognizes feature types with high similarity, a common challenge in dense forest imagery. This approach plays a significant role in identifying and distinguishing complex forest structures, such as tree species, canopy density, and understory vegetation. Despite the numerous advantages of our DASR-Net, we acknowledge its shortcomings in the precise extraction of feature boundaries. This deficiency is particularly evident in the segmentation results, where the contours may not entirely conform to the actual shape of the forest features, and the boundary lines may lack smoothness. To rectify this, we are committed to further investigating encoding methods that specifically target boundary features, ensuring a more accurate delineation of forest boundaries. Regarding dataset utilization, we are dedicated to capitalizing on high-resolution remote sensing images with complex features and rich information, which is particularly relevant to forested areas. By expanding the application scope of our algorithm and enlarging the dataset, we aim to construct a comprehensive dataset of agriculture and forestry-related features from multi-source remote sensing images. This will not only enhance the robustness of our DASR-Net but will also contribute to more effective and precise forest management and conservation efforts. Looking ahead, we should optimize algorithms for boundary feature extraction to improve the precision and smoothness of feature boundaries. We plan to explore the model's generalization capabilities across different terrains and seasonal conditions. Furthermore, we aim to investigate how to apply the model to larger-scale remote sensing data, enabling more extensive forest monitoring and management. This will address the model's applicability to different datasets and larger-scale scenarios, ensuring its utility in a broader range of contexts.

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

Funding: This work is the result of the research project funded by the Guangxi Key Research and Development Program: AB24010312, funder: Qi Li; ESA-MOST Dragon Program: Dragon 6-95177, funder: Jinlong Fan.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: Author Xueqiang Zhao was employed by the company China Water Resources Pearl River Planning Surveying and Designing Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Lu, Y.; Yang, J.; Peng, M.; Li, T.; Wen, D.; Huang, X. Monitoring ecosystem services in the Guangdong-Hong Kong-Macao Greater Bay Area based on multi-temporal deep learning. *Sci. Total. Environ.* **2022**, *822*, 153662. [CrossRef] [PubMed]
2. Darem, A.A.; Alhashmi, A.A.; Almadani, A.M.; Alanazi, A.K.; Sutantra, G.A. Development of a map for land use and land cover classification of the Northern Border Region using remote sensing and GIS. *Egypt. J. Remote Sens. Space Sci.* **2023**, *26*, 341–350. [CrossRef]
3. Selmy, S.A.; Kucher, D.E.; Mozgeris, G.; Moursy, A.R.; Jimenez-Ballesta, R.; Kucher, O.D.; Fadl, M.E.; Mustafa, A.r.A. Detecting, analyzing, and predicting land use/land cover (LULC) changes in arid regions using landsat images, CA-Markov hybrid model, and GIS techniques. *Remote Sens.* **2023**, *15*, 5522. [CrossRef]
4. Munawar, H.S.; Mojtahedi, M.; Hammad, A.W.; Kouzani, A.; Mahmud, M.P. Disruptive technologies as a solution for disaster risk management: A review. *Sci. Total Environ.* **2022**, *806*, 151351. [CrossRef]
5. Bwambale, E.; Naangmenyele, Z.; Iradukunda, P.; Agboka, K.M.; Houessou-Dossou, E.A.; Akansake, D.A.; Bisa, M.E.; Hamadou, A.A.H.; Hakizayezu, J.; Onofua, O.E.; et al. Towards precision irrigation management: A review of GIS, remote sensing and emerging technologies. *Cogent Eng.* **2022**, *9*, 2100573. [CrossRef]
6. Li, J.; Pei, Y.; Zhao, S.; Xiao, R.; Sang, X.; Zhang, C. A review of remote sensing for environmental monitoring in China. *Remote Sens.* **2020**, *12*, 1130. [CrossRef]
7. Bourbonnais, M. Applications of geographic information systems, spatial analysis, and remote sensing in environmental impact assessment. In *Routledge Handbook of Environmental Impact Assessment*; Routledge: London, UK, 2022; pp. 201–220.
8. Levin, N.; Kyba, C.C.; Zhang, Q.; de Miguel, A.S.; Román, M.O.; Li, X.; Portnov, B.A.; Molthan, A.L.; Jechow, A.; Miller, S.D.; et al. Remote sensing of night lights: A review and an outlook for the future. *Remote Sens. Environ.* **2020**, *237*, 111443. [CrossRef]
9. Asner, G.P.; Knapp, D.E.; Kennedy-Bowdoin, T.; Jones, M.O.; Martin, R.E.; Boardman, J.W.; Field, C.B. *Carnegie Airborne Observatory: In-Flight Fusion of Hyperspectral Imaging and Waveform Light Detection and Ranging for THREE-Dimensional Studies of Ecosystems*; SPIE: Kuala Lumpur, Malaysia, 2007; Volume 1, p. 013536.
10. Szeliski, R. *Computer Vision: Algorithms and Applications*; Springer Nature: Berlin/Heidelberg, Germany, 2022.
11. Panagakis, Y.; Kossaifi, J.; Chrysos, G.G.; Oldfield, J.; Nicolaou, M.A.; Anandkumar, A.; Zafeiriou, S. Tensor methods in computer vision and deep learning. *Proc. IEEE* **2021**, *109*, 863–890. [CrossRef]
12. Umbaugh, S.E. *Digital Image Processing and Analysis: Computer Vision and Image Analysis*; CRC Press: Boca Raton, FL, USA, 2023.
13. Al-Kaff, A.; Martin, D.; Garcia, F.; de la Escalera, A.; Armingol, J.M. Survey of computer vision algorithms and applications for unmanned aerial vehicles. *Expert Syst. Appl.* **2018**, *92*, 447–463. [CrossRef]
14. Shenoy, A.; Thillaiarasu, N. A survey on different computer vision based human activity recognition for surveillance applications. In Proceedings of the 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 29–31 March 2022; pp. 1372–1376.

15. Li, H.; Cui, J.; Zhang, X.; Han, Y.; Cao, L. Dimensionality reduction and classification of hyperspectral remote sensing image feature extraction. *Remote Sens.* **2022**, *14*, 4579. [CrossRef]
16. Liu, G.; Wang, L.; Liu, D.; Fei, L.; Yang, J. Hyperspectral image classification based on non-parallel support vector machine. *Remote Sens.* **2022**, *14*, 2447. [CrossRef]
17. Ayerdi, B.; Romay, M.G. Hyperspectral image analysis by spectral–spatial processing and anticipative hybrid extreme rotation forest classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 2627–2639. [CrossRef]
18. Lin, T.H.; Li, H.T.; Tsai, K.C. Implementing the Fisher's Discriminant Ratio in ak-Means Clustering Algorithm for Feature Selection and Data Set Trimming. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 76–87. [CrossRef] [PubMed]
19. Alimjan, G.; Sun, T.; Liang, Y.; Jumahun, H.; Guan, Y. A new technique for remote sensing image classification based on combinatorial algorithm of SVM and KNN. *Int. J. Pattern Recognit. Artif. Intell.* **2018**, *32*, 1859012. [CrossRef]
20. Huang, B.; Zhao, B.; Song, Y. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sens. Environ.* **2018**, *214*, 73–86. [CrossRef]
21. Calota, I.; Faur, D.; Datcu, M. DNN-based semantic extraction: Fast learning from multispectral signatures. In Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 3672–3675.
22. He, G.; Xing, S.; Xia, Z.; Huang, Q.; Fan, J. Panchromatic and multi-spectral image fusion for new satellites based on multi-channel deep model. *Mach. Vis. Appl.* **2018**, *29*, 933–946. [CrossRef]
23. Yang, M.; Ling, J.; Chen, J.; Feng, M.; Yang, J. Discriminative semi-supervised learning via deep and dictionary representation for image classification. *Pattern Recognit.* **2023**, *140*, 109521. [CrossRef]
24. Falk, T.; Mai, D.; Bensch, R.; Çiçek, Ö.; Abdulkadir, A.; Marrakchi, Y.; Böhm, A.; Deubner, J.; Jäckel, Z.; Seiwald, K.; et al. U-Net: Deep learning for cell counting, detection, and morphometry. *Nat. Methods* **2019**, *16*, 67–70. [CrossRef]
25. Chen, Y.; Ren, Q.; Yan, J. Rethinking and Improving Robustness of Convolutional Neural Networks: A Shapley Value-based Approach in Frequency Domain. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 324–337.
26. Lin, D.; Li, Y.; Nwe, T.L.; Dong, S.; Oo, Z.M. RefineU-Net: Improved U-Net with progressive global feedbacks and residual attention guided local refinement for medical image segmentation. *Pattern Recognit. Lett.* **2020**, *138*, 267–275. [CrossRef]
27. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Proceedings of the 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, 20 September 2018*; Proceedings 4; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–11.
28. Kumar, G.M.; Parthasarathy, E. Development of an enhanced U-Net model for brain tumor segmentation with optimized architecture. *Biomed. Signal Process. Control* **2023**, *81*, 104427.
29. Ding, Z.; Zhang, Y.; Zhu, C.; Zhang, G.; Li, X.; Jiang, N.; Que, Y.; Peng, Y.; Guan, X. CAT-Unet: An enhanced U-Net architecture with coordinate attention and skip-neighborhood attention transformer for medical image segmentation. *Inf. Sci.* **2024**, *670*, 120578. [CrossRef]
30. Li, J.; Cui, R.; Li, B.; Li, Y.; Mei, S.; Du, Q. Dual 1D-2D spatial-spectral cnn for hyperspectral image super-resolution. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3113–3116.
31. Hang, R.; Liu, Q.; Hong, D.; Ghamisi, P. Cascaded recurrent neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5384–5394. [CrossRef]
32. 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]
33. Ahmad, T.; Zhang, D. A data-driven deep sequence-to-sequence long-short memory method along with a gated recurrent neural network for wind power forecasting. *Energy* **2022**, *239*, 122109. [CrossRef]
34. Yang, J.H.; Wang, L.G.; Qian, J.X. Hyperspectral image classification based on spatial and spectral features and sparse representation. *Appl. Geophys.* **2014**, *11*, 489–499. [CrossRef]
35. Bello, I.; Fedus, W.; Du, X.; Cubuk, E.D.; Srinivas, A.; Lin, T.Y.; Shlens, J.; Zoph, B. Revisiting resnets: Improved training and scaling strategies. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22614–22627.
36. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
37. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
38. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
39. Hong, D.; Han, Z.; Yao, J.; Gao, L.; Zhang, B.; Plaza, A.; Chanussot, J. SpectralFormer: Rethinking hyperspectral image classification with transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5518615. [CrossRef]

40. Fan, X.; Li, X.; Yan, C.; Fan, J.; Yu, L.; Wang, N.; Chen, L. MARC-Net: Terrain Classification in Parallel Network Architectures Containing Multiple Attention Mechanisms and Multi-Scale Residual Cascades. *Forests* **2023**, *14*, 1060. [CrossRef]
41. Fan, X.; Li, X.; Yan, C.; Fan, J.; Chen, L.; Wang, N. Converging Channel Attention Mechanisms with Multilayer Perceptron Parallel Networks for Land Cover Classification. *Remote Sens.* **2023**, *15*, 3924. [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.