



Technical Note

MADANet: A Lightweight Hyperspectral Image Classification Network with Multiscale Feature Aggregation and a Dual Attention Mechanism

Bing Cui ¹, Jiayang Wen ¹ , Xiukai Song ² and Jianlong He ^{2,*}

¹ College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China; cuibinge@sdust.edu.cn (B.C.); jiayangwen@sdust.edu.cn (J.W.)

² Shandong Key Laboratory of Marine Ecological Restoration, Shandong Marine Resource and Environment Research Institute, Yantai 264006, China; songxiukai@shandong.cn

* Correspondence: hejianlong@shandong.cn

Abstract: Hyperspectral remote sensing images, with their continuous, narrow, and rich spectra, hold distinct significance in the precise classification of land cover. Deep convolutional neural networks (CNNs) and their variants are increasingly utilized for hyperspectral classification, but solving the conflict between the number of model parameters, performance, and accuracy has become a pressing challenge. To alleviate this problem, we propose MADANet, a lightweight hyperspectral image classification network that combines multiscale feature aggregation and a dual attention mechanism. By employing depthwise separable convolution, multiscale features can be extracted and aggregated to capture local contextual information effectively. Simultaneously, the dual attention mechanism harnesses both channel and spatial dimensions to acquire comprehensive global semantic information. Ultimately, techniques such as global average pooling (GAP) and full connection (FC) are employed to integrate local contextual information with global semantic knowledge, thereby enabling the accurate classification of hyperspectral pixels. The results from the experiments conducted on representative hyperspectral images demonstrate that MADANet not only attains the highest classification accuracy but also maintains significantly fewer parameters compared to the other methods. Experimental results show that our proposed framework significantly reduces the number of model parameters while still achieving the highest classification accuracy. As an example, the model has only 0.16 M model parameters in the Indian Pines (IP) dataset, but the overall accuracy is as high as 98.34%. Similarly, the framework achieves an overall accuracy of 99.13%, 99.17%, and 99.08% on the University of Pavia (PU), Salinas (SA), and WHU Hi LongKou (LongKou) datasets, respectively. This result exceeds the classification accuracy of existing state-of-the-art frameworks under the same conditions.

Keywords: hyperspectral image classification; multiscale feature aggregation; dual attention mechanism



Citation: Cui, B.; Wen, J.; Song, X.; He, J. MADANet: A Lightweight Hyperspectral Image Classification Network with Multiscale Feature Aggregation and a Dual Attention Mechanism. *Remote Sens.* **2023**, *15*, 5222. <https://doi.org/10.3390/rs15215222>

Academic Editor: Paul Scheunders

Received: 18 September 2023

Revised: 24 October 2023

Accepted: 1 November 2023

Published: 3 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Hyperspectral technology, a very important branch of remote sensing technology, emerged in the 1980s and brought remote sensing technology into a new stage of development. Currently, hyperspectral technology is widely used in geological exploration, precision agriculture, environmental monitoring, oceanography, and other fields [1–5]. In contrast to natural images, hyperspectral images (HSIs) encompass numerous continuous spectral bands with high spectral resolution within a single scene, thereby affording a wealth of spectral and spatial information pertaining to ground objects. However, due to problems such as dimensional disasters and mixed pixels, hyperspectral image classification is challenging [6–9].

The rapid development of deep learning has provided powerful algorithms for solving these tasks [10]. Ever since AlexNet [11] secured victory in the 2012 ImageNet challenge,

convolutional neural networks (CNNs) have garnered considerable attention. In subsequent years, the field of image classification witnessed the emergence of several classic CNN models, notably GoogleNet, VGGNet, ResNet, and DenseNet [12–15], and many researchers have also conducted studies on hyperspectral image classification based on these models [16–24]. However, to enhance accuracy, most of these models incorporate an extensive number of hidden layers and training parameters, which limits their suitability for deployment in resource-constrained mobile and embedded applications, particularly on satellite and airborne platforms [25]. In recent years, many academic researchers have shifted their focus to building lightweight and efficient CNN for applications and implementations in mobile and embedded devices, such as MobileNet [26–28] and ShuffleNet [29,30], due to limitations in computing power, memory, power consumption, and parameter size. Compared to existing research, light CNN sacrifices some model accuracy to reduce the consumption of limited memory and computational resources. Therefore, the challenge of maintaining high accuracy with fewer parameters has become an urgent issue [31]. Multiscale feature extraction provides an effective solution to retain more relevant information within a limited number of parameters. Numerous experiments have demonstrated that multiscale features have a significant positive impact on classification performance [32–36].

Attention mechanisms have demonstrated substantial promise in enhancing the performance of CNNs by effectively suppressing redundant information within feature maps and extracting meaningful features. Consequently, attention mechanisms have been widely adopted in contemporary deep CNN architectures [37–41]. The SE block [37] is an attention mechanism that effectively enhances accuracy by modeling the correlation between feature maps and reinforcing the important feature maps. CBAM [38] improves the SE block by introducing two different pooling operations, GAP and GMP, and adds a spatial attention module to enhance information interaction in the space. This provides an effective method for extracting and utilizing spatial and spectral features for hyperspectral image classification but also presents new challenges in terms of computational resources and data requirements. The nonlocal (NL) block efficiently captures long-range feature dependencies by modeling the global context through the self-attention mechanism [39]. The dual attention mechanism proposed by DANet [40] applies the idea of NL to both the spatial and channel domains, using image pixels and feature maps as query statements to model the context to improve the global feature representation capability of the network. The superiority of DANet to model context has been demonstrated in several other works [42,43].

This paper introduces a lightweight hyperspectral image classification method founded on multiscale feature aggregation and a dual attention mechanism. First, two multiscale aggregation units (MA units) based on multiscale products were proposed to effectively combine local spectrum and spatial features at different scales to solve the “adjacent pixel effect” issue in hyperspectral data. Specifically, the MA units use cross-scale convolution kernels to extract multiscale features, which are then nonlinearly fused through the Hadamard product. Second, dual attention units (DA units) were introduced to capture global dependencies in both spectral and spatial dimensions, aiming to obtain feature representations with lower intra-class divergence. Finally, the feature maps produced by the MA units and the DA units are connected to create a multiscale spatial-spectral feature map, which is then classified using a Softmax classifier.

The principal contributions of this study are delineated as follows:

- (1) A lightweight hyperspectral image classification method is proposed based on multiscale feature aggregation and a dual attention mechanism, which can achieve good performance quickly with low computational consumption.
- (2) We introduce an applicable unit: the multiscale aggregation unit (MA unit). The MA unit first captures multiscale spatial context information through multilayer deep convolution and then performs nonlinear fusion using the Hadamard product to cope with the “neighboring pixel effect” in hyperspectral images. This unit provides new perspectives and methods for hyperspectral image feature extraction.

- (3) To assess the generalization and advantages of the proposed method, experiments were conducted across various scenarios encompassing three agricultural contexts and one urban setting. The experimental results consistently illustrate that our method outperforms other state-of-the-art approaches in all tested scenarios.

2. Related Work

2.1. ShuffleNet

ShuffleNet is an efficient CNN model proposed by KuangShi Technology. ShuffleNet V1 [29] uses a 1×1 group convolution to reduce the computational complexity and number of parameters. To overcome the drawbacks brought about by group convolutions, ShuffleNet V1 introduced a novel channel shuffle operation to help channel information exchange. ShuffleNet V2 [30], an evolution of ShuffleNet V1, proficiently mitigates memory access costs by maintaining a consistent number of input and output channels within the convolutional layer. Furthermore, during module construction, it employs a channel-splitting method to curtail the number of convolutional channels, resulting in a concurrent reduction in both computational complexity and parameter count. Compared with ShuffleNet V1, ShuffleNet V2 is faster and more accurate on the same computational budget. The structure of the ShuffleNet V2 basic block is shown in Figure 1.

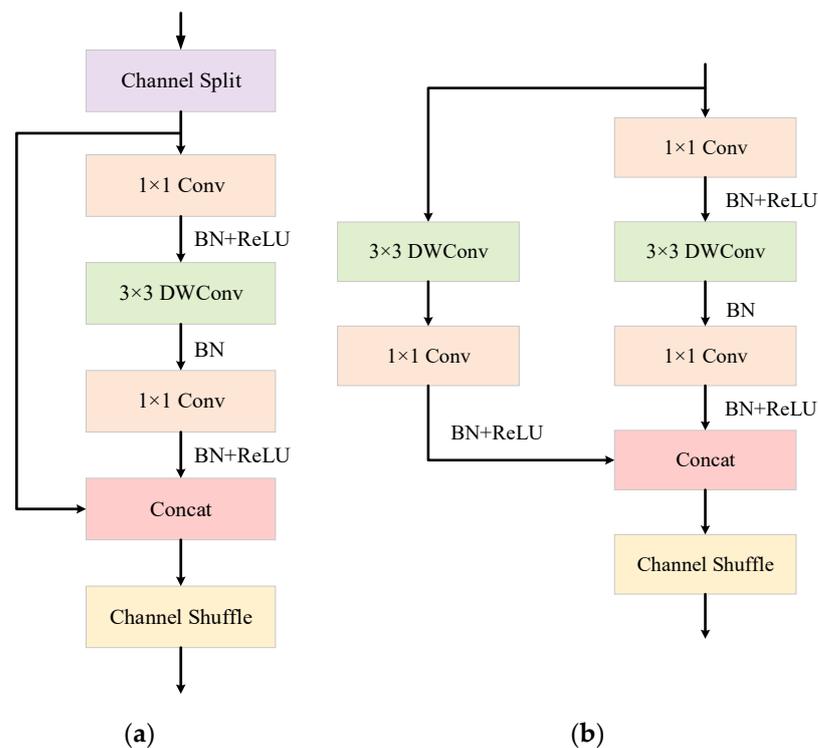


Figure 1. The structure of the ShuffleNet V2 basic block. (a) Basic unit (stride = 1). (b) Spatial down sampling (stride = 2). DWConv: depthwise convolution.

2.2. Attention Mechanism

In 2019, Jun Fu's team introduced DANet [40], a dual attention mechanism that adaptively captures feature dependencies in both spatial and channel dimensions, thereby enhancing the network's global feature representation. Position attention provides a similarity matrix by computing the relationship between each pixel and all other pixels and then multiplies this similarity matrix by the original matrix to update the pixels at each location in the original feature map. Channel attention performs a similar transformation to update each channel. The outputs of these two attention modules are combined to further enhance the feature representation. These attention modules are shown in Figure 2.

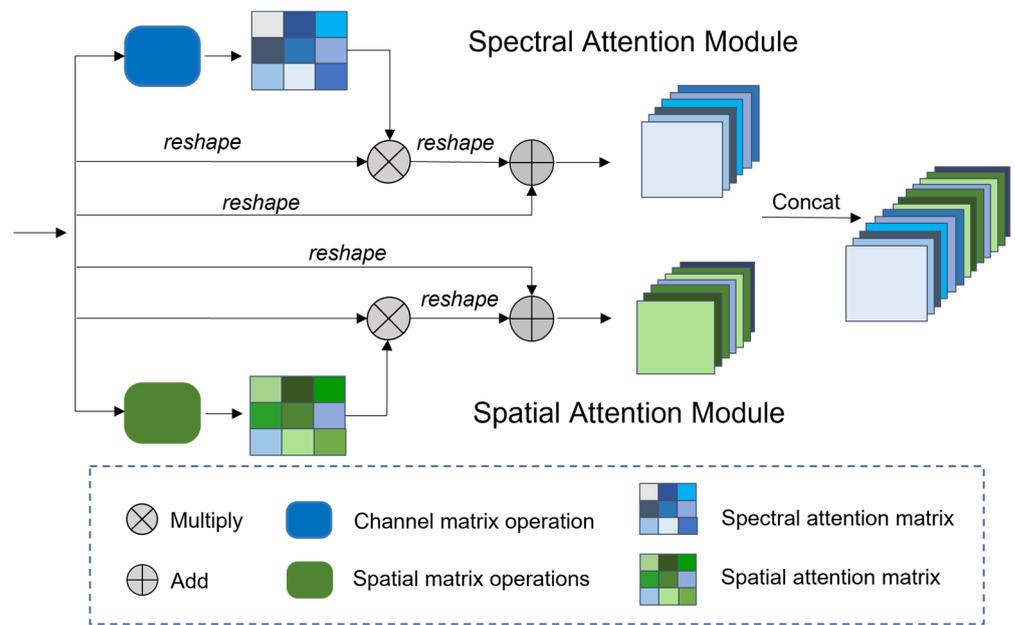


Figure 2. Dual attention unit (DA unit).

3. Methodology

3.1. Overall Architecture

Figure 3 depicts a schematic representation of the MADANet method. We take the Indian Pines dataset as an example to illustrate the process in detail. First, PCA is applied to reduce the spectral dimensions and suppress the band noise in the original HSI. Second, this HSI, after dimensionality reduction, is segmented into 3D image cubes centered around labeled pixels. Third, a convolution operation and a maximum pooling operation, with a size of 3×3 , are applied successively to each 3D image cube. Afterward, the 3D cubes are transported to two parallel branches, namely, the MA and DA branches. To retain more relevant information at a limited depth, we use hierarchical layers composed of three MA units in the bottom MA branch. The DA unit in the top DA branch is used to improve the global feature representation capability of the network. Consequently, we will obtain discriminative feature maps for multiple classes. Following the aforementioned operations, these feature maps are then fed into the FFC module to derive the HSI classification result.

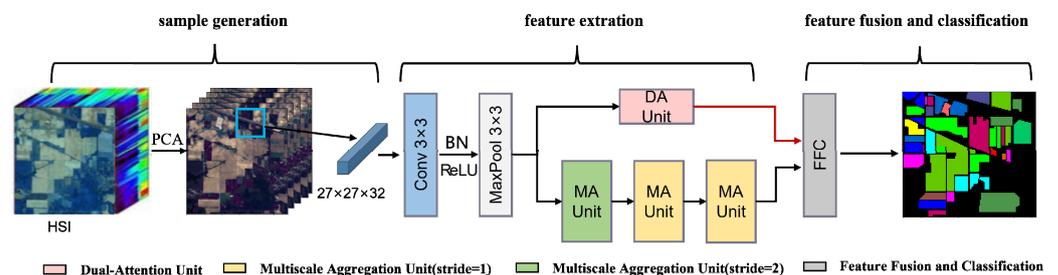


Figure 3. The structure of MADANet. Red narrows: downsampling operators.

3.2. MA Unit

By design, the MA unit pursues a strong feature extraction ability and is extremely lightweight. We use two types of feature aggregation units, including the basic unit (shown in Figure 4a) and the spatial downsampling unit (shown in Figure 4b). Multiple depth convolutions with different convolution kernel sizes (3×3 , 5×5 , 7×7) are used in the residual branch of the basic unit to capture multiscale spatial contextual information, and these multiscale features are subsequently fused by the Hadamard product. Additionally, batch normalization (BN) and ReLU activation are introduced to each convolutional layer

to alleviate the gradient disappearance problem. In the spatial downsampling unit, a 3×3 depthwise convolution of step 2 and a 1×1 convolution are added to the shortcut branch to ensure that the output size of the feature maps of both branches is the same. The output features of the two branches are merged using the Concat operation, followed by performing the Channel Shuffle operation to help with communicating information.

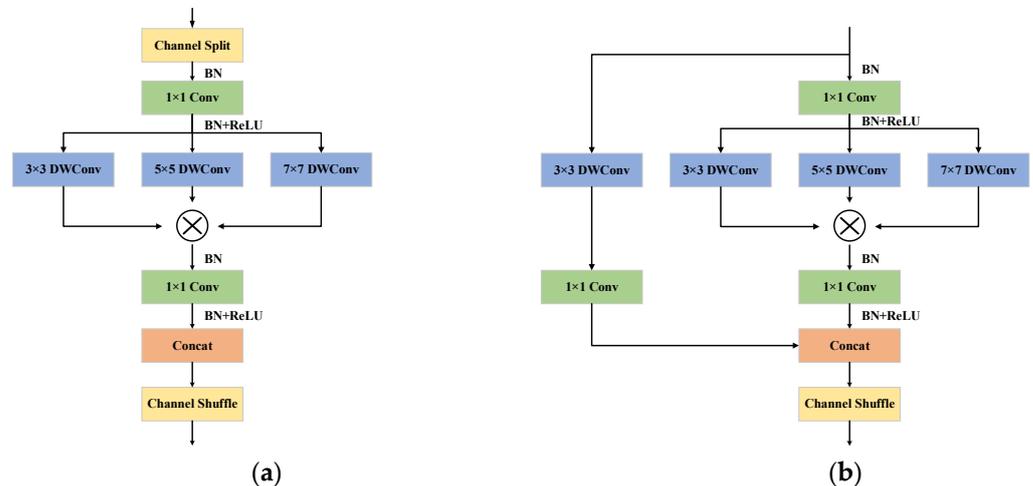


Figure 4. Multiscale aggregation unit. (a) Basic unit (stride = 1). (b) Spatial downsampling unit (stride = 2). DWConv: depthwise convolution.

3.3. Feature Fusion and Classification

The proposed feature fusion and classification module is shown in Figure 5. The obtained global and local features are concatenated and then fused by 1×1 convolution and batch normalization. The GAP operation is employed to condense spatial information, and the FC layer is utilized to reduce the dimensionality of the output channels to align with the number of surface categories. The Softmax function accepts the output of the FC layer and generates the classification probabilities, and finally, the predicted class is obtained by the Argmax function.

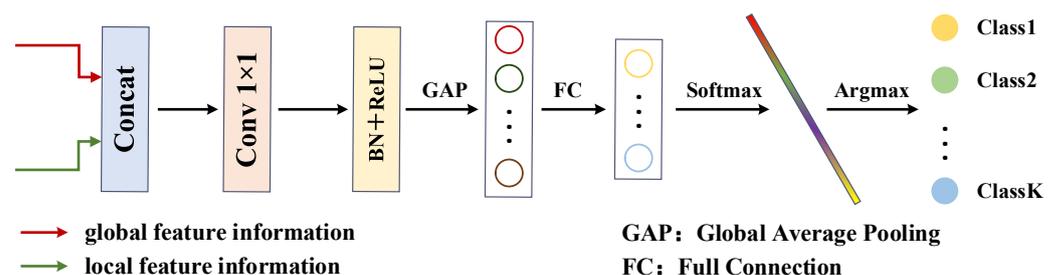


Figure 5. Feature fusion and classification.

4. Results

4.1. Data Description

To assess the classification performance of MADANet, we utilize four HSI datasets: the Indian Pines, University of Pavia, Salinas, and WHU-Hi-LongKou datasets [44,45].

Indian Pines dataset: The Indian Pines dataset stands as the inaugural set of hyperspectral image classification test data. It was generated in 1992 by the airborne visible/infrared imaging spectrometer (AVIRIS) by imaging a section of Indian Pines in Indiana, USA. Subsequently, this dataset was resized to 145×145 for the specific purpose of hyperspectral image classification testing. It encompasses 220 continuous spectral bands, spanning the electromagnetic spectrum from 0.4 to $2.5 \mu\text{m}$. The false color composite image and the ground truth are shown in Figure 6.

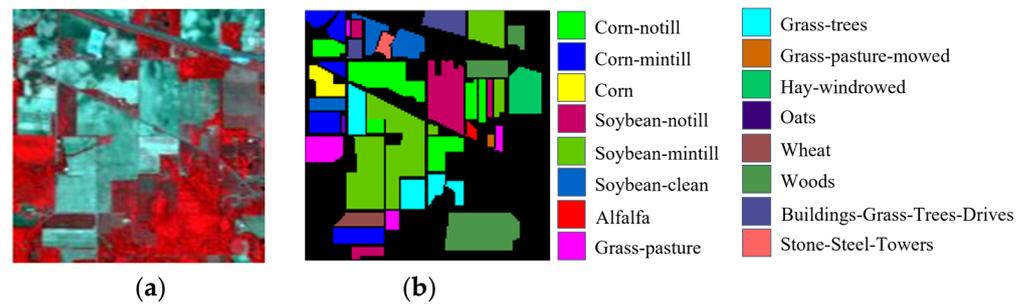


Figure 6. Indian Pines dataset. (a) False color composite and (b) ground truth.

University of Pavia dataset: The University of Pavia dataset, representing Pavia in northern Italy, consists of an image with dimensions of 610×340 pixels and a spatial resolution of 1.3 m. Before HSI classification, a conventional preprocessing method was applied, including the removal of 12 noisy channels. The false color composite image and the ground truth are shown in Figure 7.

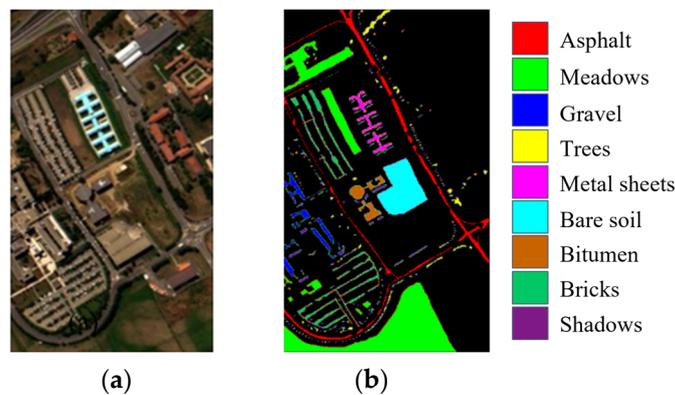


Figure 7. University of Pavia dataset. (a) False color composite and (b) ground truth.

Salinas dataset: This dataset boasts an impressive 3.7 m spatial resolution and comprises 224 spectral bands, covering an area of 512×217 pixels. Prior to HSI classification, standard preprocessing techniques were applied, including the removal of 20 channels affected by water absorption bands. The images encompass 16 distinct land cover categories, such as vegetables, bare soil, and vineyards. The false color composite image and the ground truth are shown in Figure 8.

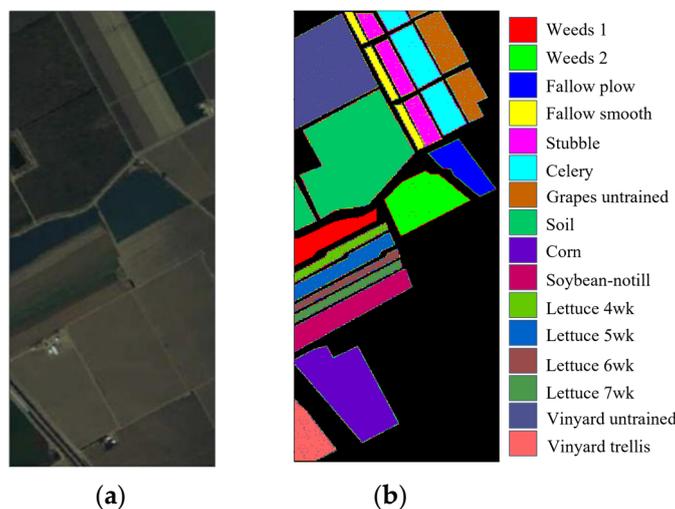


Figure 8. Salinas dataset. (a) False color composite and (b) ground truth.

WHU-Hi-LongKou dataset: The WHU-Hi-LongKou dataset is a subset of the WHU-Hi dataset, meticulously captured and compiled by the RSDIEA research team at Wuhan University within the region of Longkou in Hubei Province, China. This dataset is characterized by a spatial resolution of approximately 0.463 m, incorporating a total of 270 spectral bands and encompassing an area spanning 550×400 pixels. The imagery portrays an agricultural setting featuring six distinct crop types. The false color composite image and the ground truth are shown in Figure 9.

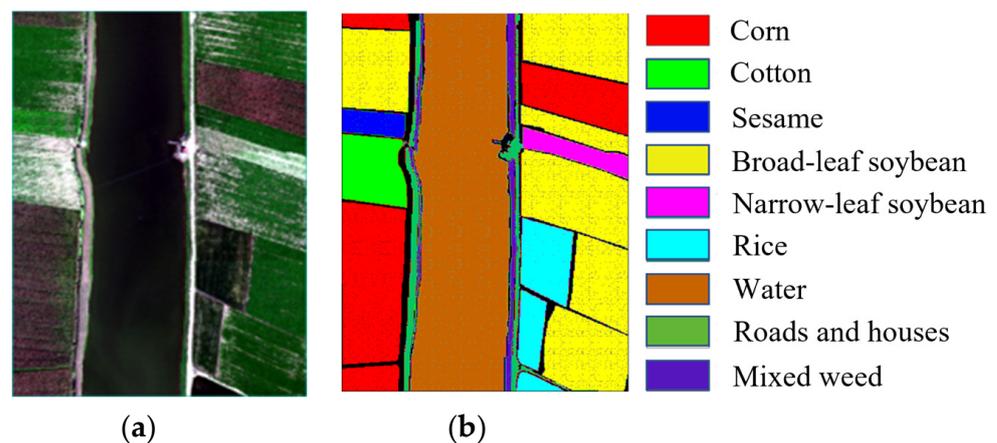


Figure 9. WHU-Hi-LongKou dataset. (a) False color composite and (b) ground truth.

4.2. Experimental Design

We used three metrics to compare the classification results with the actual conditions in order to evaluate the performance of the proposed method. The overall accuracy (OA) was used to evaluate the proportion of correctly classified pixels to the total test pixels, providing an overall performance overview. The average accuracy (AA) was used to calculate the accuracy of each category. In addition, these accuracy values were averaged to reflect the performance differences between different categories more comprehensively. The Kappa coefficient was used to determine the consistency between the model's predictions and the actual classification results. Furthermore, the evaluation took into account the computing resources, including the processing time of the Central Processing Unit (CPU), the computation time used by the Graphics Processing Unit (GPU), and the total number of model parameters used in the method.

Considering the large differences in sample size between the four datasets, we used a different proportion of training samples for each dataset. For the Indian Pines dataset, a partitioning strategy was employed wherein 10% of the labeled samples were designated for the training subset, an additional 10% were reserved for validation purposes, and the remaining 80% were utilized for testing. Conversely, for the Salinas and Pavia University datasets, a distinct distribution scheme was implemented, allocating 5% of the labeled samples for training, another 5% for validation, and a substantial 90% for the testing phase. For the WHU-Hi-LongKou dataset, 2% of the labeled samples in each category were used for training and validation, and the remaining samples were used for testing. In the training phase of the network, the Adam optimizer was used to update the parameters, where the initial learning rate was set to 0.0001, the batch size was set to 32, and the number of training epochs was set to 200.

All algorithms involved in this study were run in PyCharm Professional 2023.2 based on the PyTorch 1.12.1 deep learning framework. The CPU processor was an 12th Gen Intel®Core™ i5-12500 CPU @ 3.00 GHz, and an NVIDIA GeForce RTX 2080Ti GPU graphics card was used.

To validate MADANet's effectiveness, we compared it with six classification methods: two benchmark machine learning algorithms (SVM [46] and RF [47]) and four deep-learning approaches (DFFN [48], HybridSN [49], A²MFE [33], and LANet [50]).

DFFN is a 2D CNN network that fuses outputs from different hierarchical layers using a residual structure. HybridSN enhances its ability to focus on relevant features by incorporating an attention mechanism. A²MFE applies adaptive attention mechanisms at multiple scales to capture local and global contextual information effectively in hyperspectral images. Lastly, LANet leverages local attributes of hyperspectral images, offering a unique approach to feature extraction.

These models extract HSI features from diverse angles, representing the latest advancements in HSI classification. To ensure fairness, all methods used the same parameters from their original works. Deep learning methods shared spatial patch size and dimensions, while SVM and RF utilized serialized data.

4.3. Experimental Results

4.3.1. Experimental Results on the Indian Pines Dataset

The inaugural experiment was conducted utilizing the Indian Pines dataset. Table 1 shows the overall accuracy (OA), average accuracy (AA), and Kappa coefficient obtained by the different methods, as well as the classification accuracy of each class. From Table 1, we can see that compared with the traditional SVM and RF methods, the deep learning methods show better classification performance. However, both DFFN and LANet considered only the spectral–spatial features on a single scale, resulting in poor performance. MADANet combines global features with local features on multiple scales, making it significantly more effective for small-scale oats and past-mowed grass. In terms of OA, the MADANet proposed in this paper outperforms the other methods with 98.34% classification accuracy. In addition, the visualization results of all methods are presented in Figure 10, where the proposed MADANet shows satisfactory performance. The SVM and RF methods did not consider spatial information, so these classification maps contained considerable noise. DFFN, HybridSN, A²MFE, and LANet showed more misclassifications.

Table 1. Comparison of the different methods in terms of class accuracy, OA, AA, and Kappa coefficient for the Indian Pines dataset.

Class	SVM	RF	DFFN	HybridSN	A ² MFE	LANet	MADANet
Alfalfa	83.16	77.15	92.76	98.46	97.98	98.37	98.37
Corn-N	66.73	70.24	89.36	96.40	95.32	91.64	96.98
Corn-M	76.25	71.16	93.31	96.64	94.57	93.25	96.32
Corn	80.64	76.69	94.47	95.05	91.34	90.51	95.57
Grass-P	85.89	84.83	97.54	97.63	95.21	95.26	96.59
Grass-T	92.41	89.27	98.11	98.54	94.72	98.17	98.65
Grass-P-M	52.32	39.41	99.06	99.21	92.87	94.38	99.25
Hay-windrowed	59.24	63.17	99.70	98.24	97.31	98.21	98.24
Oats	62.24	60.17	96.02	96.72	96.35	96.93	97.44
Soybean-notill	82.73	74.21	95.85	93.28	92.41	92.74	98.38
Soybean-mintill	84.47	80.41	95.32	94.51	97.51	96.36	99.43
Soybean-clean	66.11	64.04	84.78	99.29	97.45	93.14	97.65
Wheat	89.68	88.26	98.77	98.72	96.89	98.59	99.19
Woods	84.87	86.52	95.36	98.70	95.76	98.71	98.12
Buildings-G-T-D	82.59	81.78	97.23	98.56	98.61	97.47	97.32
Stone-S-T	86.57	84.43	92.03	97.21	96.25	96.44	98.26
OA (%)	84.12	77.88	96.95	97.35	97.78	95.06	98.34
AA (%)	82.76	76.14	95.65	97.92	97.24	96.23	98.12
Kappa	0.8234	0.7606	0.9523	0.9613	0.9632	0.9542	0.9703

The best results for each indicator are highlighted in bold.

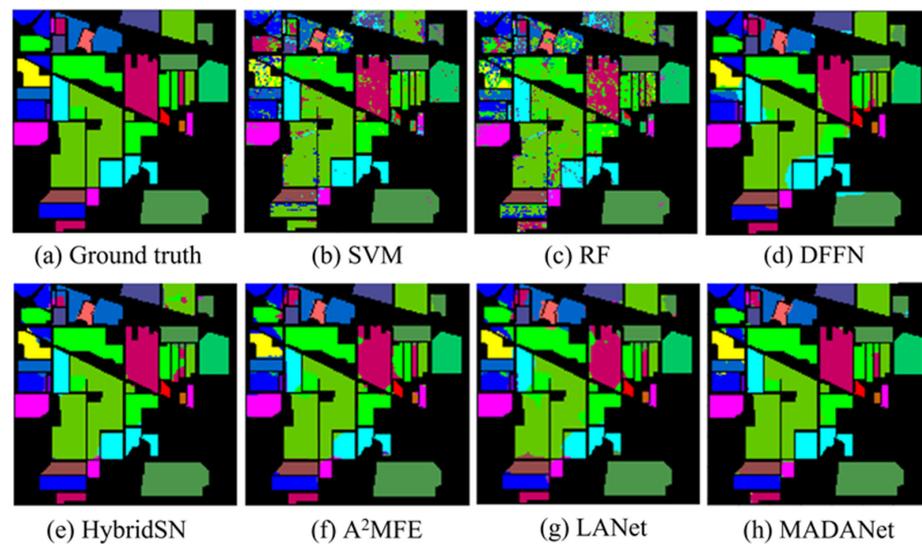


Figure 10. Classification maps produced by the different methods for the Indian Pines image. (a) Ground truth, (b) SVM, (c) RF, (d) DFFN, (e) HybridSN, (f) A²MFE, (g) LANet, and (h) MADANet.

4.3.2. Experimental Results on the University of Pavia Dataset

The second experiment was performed on the University of Pavia dataset. Table 2 and Figure 11 show the classification accuracies and visualization results for all methods, respectively. The spectral-based SVM and RF classification methods contain significant noise. However, the classification maps produced by HybridSN and the other deep learning methods gave good visual results. In addition, we observed that MADANet more accurately classified linear geo-objects, such as asphalt, metal sheets, and bricks, with 99.04%, 99.42%, and 99.37% classification accuracy, respectively.

4.3.3. Experimental Results on the Salinas Dataset

The third experiment was performed on the Salinas dataset. The classification accuracies and visualization results for all methods are shown in Table 3 and Figure 12, respectively. MADANet provided better classification results than other methods by capturing and fusing multiscale features with an overall accuracy of 99.17%. In the Salinas dataset, there are two classes with strong spectral similarity, namely, vineyard untrained and vineyard trellis. MADANet achieved 99.34% and 98.49% classification accuracy for these two classes, respectively, which were much higher than the accuracies of the comparison methods, which indicates that our method can extract discriminative spectral–spatial features.

Table 2. Comparison of the different methods in terms of class accuracy, OA, AA, and Kappa coefficient for The University of Pavia image.

Class	SVM	RF	DFFN	HybridSN	A ² MFE	LANet	MADANet
Asphalt	94.96	94.83	95.30	98.90	98.89	94.70	99.04
Meadows	88.14	90.28	91.42	99.21	98.72	98.89	99.32
Gravel	24.62	88.54	96.96	97.27	96.57	97.42	98.24
Trees	92.63	92.86	93.81	93.64	97.32	93.07	100.00
Metal Sheets	91.74	91.79	98.63	99.60	99.71	98.43	99.42
Bare Soil	42.68	67.39	89.25	98.89	97.56	95.56	99.55
Bitumen	23.12	63.56	96.53	85.33	96.76	97.64	97.56
Bricks	77.91	81.27	91.35	98.21	94.66	93.51	99.37
Shadows	66.25	75.79	96.69	95.27	96.45	90.37	100.00
OA (%)	80.43	87.85	94.31	98.06	98.43	95.92	99.13
AA (%)	79.63	86.23	93.88	96.49	97.21	94.14	98.93
Kappa	0.7821	0.8612	0.9472	0.9642	0.9646	0.9456	0.9812

The best results for each indicator are highlighted in bold.

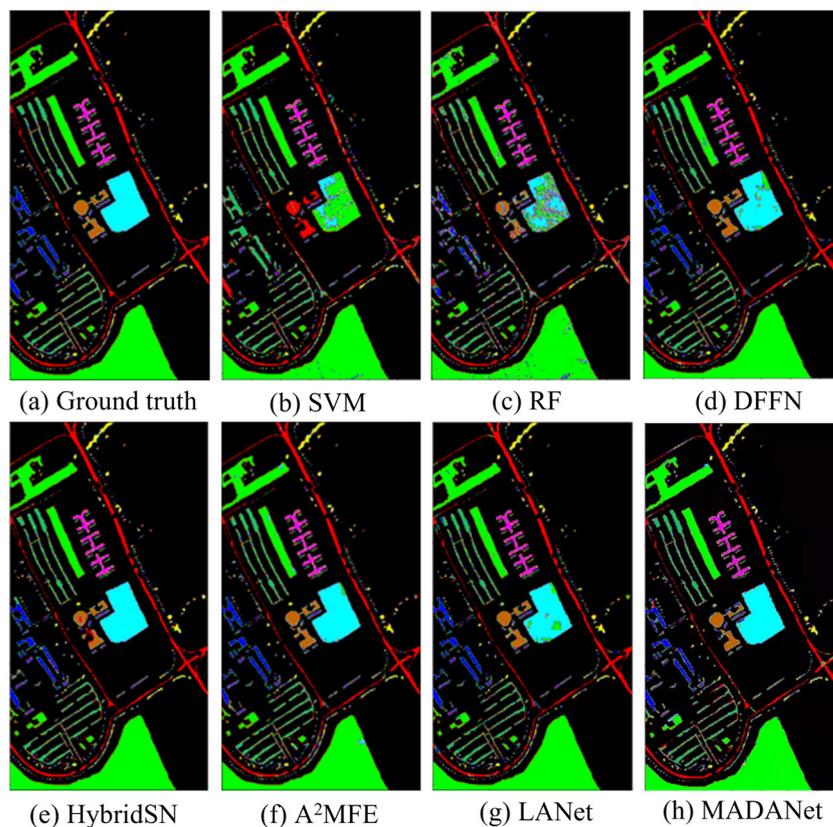


Figure 11. Classification maps produced by the different methods for the University of Pavia image. (a) Ground truth, (b) SVM, (c) RF, (d) DFFN, (e) HybridSN, (f) A²MFE, (g) LANet, and (h) MADANet.

Table 3. Comparison of the different methods in terms of class accuracy, OA, AA, and Kappa coefficient for the Salinas image.

Class	SVM	RF	DFFN	HybridSN	A ² MFE	LANet	MADANet
weeds_1	98.47	97.24	98.23	99.30	99.46	98.32	99.64
weeds_2	97.15	98.37	98.43	99.96	97.28	98.64	98.95
Fallow	92.31	87.36	95.32	98.16	97.99	95.09	98.90
Fallow-P	92.27	88.42	95.27	98.05	95.67	98.01	97.58
Fallow-S	96.11	97.13	96.23	98.42	98.85	98.73	99.02
Stubble	92.12	92.22	94.58	98.67	98.31	98.86	98.01
Celery	89.28	89.53	96.17	98.96	95.76	97.04	99.34
Grapes	90.84	90.78	91.13	92.70	97.82	96.49	97.77
Soil	77.56	62.26	97.65	98.53	97.56	93.11	98.23
Corn	96.45	93.65	93.61	95.44	95.23	94.27	95.91
Lettuce_4wk	76.16	79.67	93.78	98.08	97.45	97.19	98.86
Lettuce_5wk	87.02	92.34	91.98	99.33	98.34	98.42	99.23
Lettuce_6wk	85.16	91.36	95.12	99.02	99.21	98.58	100.00
Lettuce_7wk	82.46	85.69	90.97	98.21	98.12	98.07	99.34
Vineyard_U	72.39	68.22	87.71	92.29	94.55	92.46	98.49
Vineyard_T	89.43	88.76	91.59	80.43	93.23	90.11	98.36
OA (%)	87.53	90.45	95.69	97.05	98.56	96.67	99.17
AA (%)	88.92	91.23	96.72	97.52	98.72	97.12	99.12
Kappa	0.8689	0.8916	0.9536	0.9672	0.9735	0.9639	0.9834

The best results for each indicator are highlighted in bold.

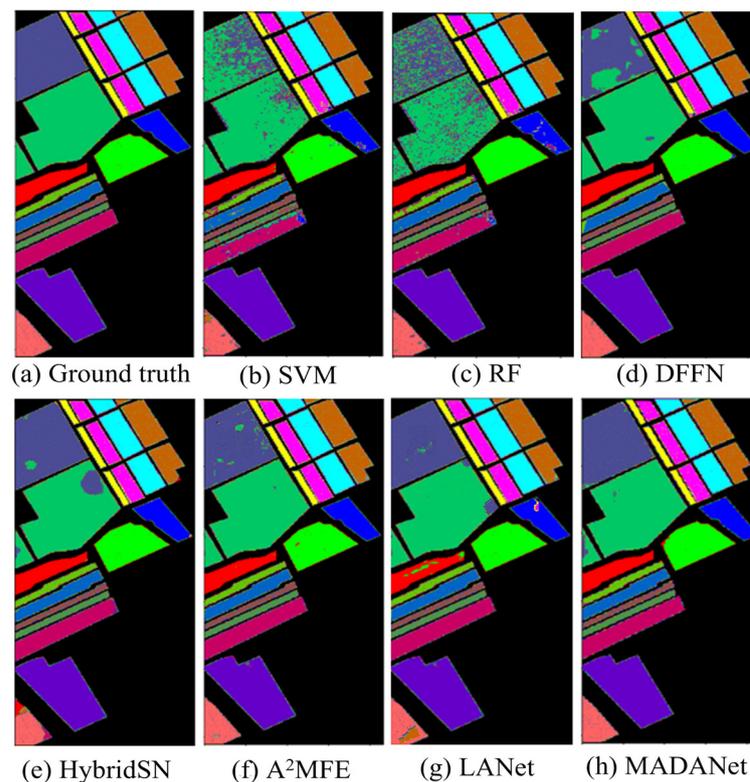


Figure 12. Classification maps produced by the different methods for the Salinas image. (a) Ground truth, (b) SVM, (c) RF, (d) DFFN, (e) HybridSN, (f) A²MFE, (g) LANet, and (h) MADANet.

4.3.4. Experimental Results on the WHU-Hi-LongKou Dataset

The fourth experiment was performed on the WHU-Hi-LongKou dataset. The classification accuracies and visualization results of all methods are shown in Table 4 and Figure 13, respectively. In Table 4, MADANet achieved an OA of 99.08%, showing a significant improvement in classification accuracy compared to SVM, RF, and LANet. A²MFE and MADANet showed better performance in two categories, roads and houses and mixed weed, as shown in the two longitudinal striped areas in Figure 13. MADANet achieved a much higher accuracy in the sesame category than the other models.

Table 4. Comparison of the different methods in terms of class accuracy, OA, AA, and Kappa coefficient for The WHU-Hi-LongKou dataset.

Class	SVM	RF	DFFN	HybridSN	A ² MFE	LANet	MADANet
Corn	93.56	92.85	99.74	99.23	99.72	98.47	99.89
Cotton	83.21	82.98	99.09	73.67	98.56	88.46	98.72
Sesame	81.52	81.30	89.25	56.72	92.34	81.13	98.83
Broad-leaf-S	76.39	74.21	98.03	95.64	98.87	96.39	99.56
Narrow-leaf-S	63.33	60.75	75.57	56.49	98.97	85.48	99.26
Rice	91.24	89.72	98.10	92.54	99.37	88.52	99.13
Water	89.76	96.61	95.80	98.91	99.53	99.76	99.64
Roads and houses	75.63	76.82	75.98	82.23	96.69	82.95	96.34
Mixed weed	76.63	72.32	71.56	92.51	94.67	81.53	94.23
OA (%)	91.05	88.72	97.80	96.08	98.46	95.77	99.08
AA (%)	88.79	86.56	92.92	84.85	97.96	89.19	98.72
Kappa	0.8972	0.8725	0.9711	0.9413	0.9832	0.9446	0.9894

The best results for each indicator are highlighted in bold.

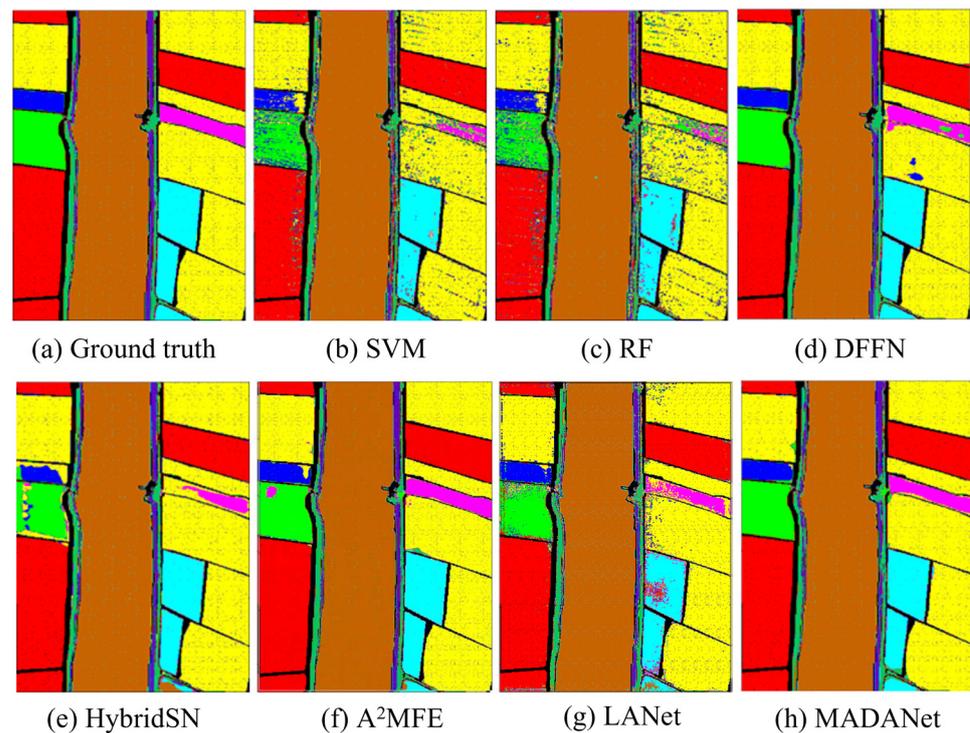


Figure 13. Classification maps produced by the different methods for the WHU-Hi-LongKou image. (a) Ground truth, (b) SVM, (c) RF, (d) DFFN, (e) HybridSN, (f) A²MFE, (g) LANet, and (h) MADANet.

4.3.5. Parameter Analysis

To evaluate the effect of the patch size on the performance of MADANet, we set the image patch size to 15, 19, 23, 27, 31, and 35. As illustrated in Figure 14, a discernible upward trend in OA is evident as the patch size increases from 15 to 27. When the patch size is 27, the OA reaches a peak. Henceforth, the patch size is established at 27 within the context of this research.

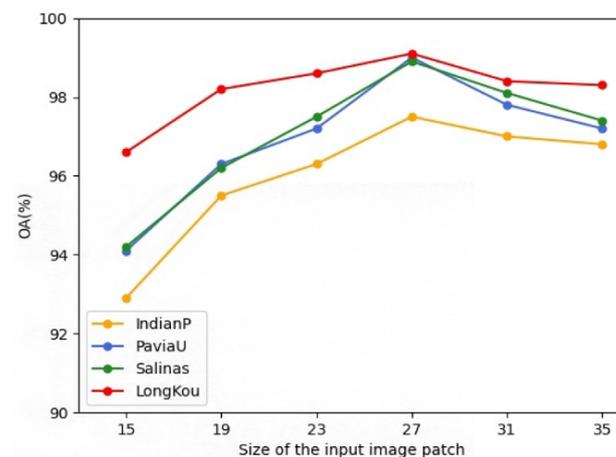


Figure 14. Classification performance of different input patch sizes on four datasets.

4.3.6. Ablation Experiments

To ascertain the effectiveness of the proposed DA unit and MA unit, four specific ablation experiments were designed. As shown in Table 5, the DA unit improved the OA by 1.97–3.23%, and the MA unit improved the OA by 1.43–2.88% on the four datasets. After combining the DA unit and the MA unit, the proposed MADANet improved the OA by 3.45–4.75% compared to the baseline. The results of the ablation experiments

validate the effectiveness and complementarity of the two proposed modules in the HSI classification task.

Table 5. Effect of DA unit and MA unit on four HSI datasets.

Datasets	DA Unit	MA Unit	OA (%)	AA (%)	Kappa
Indian Pines			93.59	94.37	0.9312
	✓		96.74	96.88	0.9577
		✓	95.33	95.50	0.9498
	✓	✓	98.34	98.12	0.9703
University of Pavia			95.45	95.12	0.9429
	✓		97.42	96.53	0.9587
		✓	98.31	97.97	0.9724
	✓	✓	99.13	98.93	0.9812
Salinas			95.72	96.21	0.9531
	✓		97.77	97.84	0.9706
		✓	97.15	97.16	0.9655
	✓	✓	99.17	99.12	0.9834
WHU-Hi-LongKou			95.41	96.67	0.9501
	✓		98.64	98.55	0.9827
		✓	97.45	97.32	0.9745
	✓	✓	99.08	98.72	0.9894

✓ indicates that the module is included; the best results for each indicator are highlighted in bold.

4.3.7. Evaluation of Model Complexity

We conducted a comparative analysis of various methods on the Indian Pines dataset, with a specific focus on the number of parameters and test time. As indicated in Table 6, the deep learning-based methods require more test time than SVM and RF. HybridSN, which involves 3D convolution, has the most parameters and the longest test time. Contextual aggregation-based LANet also requires more computations and test time. The test times for the models A²MFE, DFFN, and MADANet are relatively short, but the number of parameters for A²MFE is more than twice that of the other two models. MADANet is slightly inferior to DFFN in terms of the number of parameters and test time, but its classification accuracy is much higher.

Table 6. Test time and parameters of different methods.

Method	Test Time—CPU (s)	Test Time—GPU (s)	Parameters (M)
SVM	1.59	1.44	-
RF	0.24	0.21	-
DFFN	11.05	2.34	0.14
HybridSN	150.18	4.93	12.59
A ² MFE	14.23	2.78	0.35
LANet	109.23	4.46	2.05
MADANet	12.02	2.52	0.16

5. Discussion

In this paper, we design a simple and efficient model for the fine classification of hyperspectral images. The main idea is to find features that are visible in multiple receptive fields. To achieve this, our model uses Hadamard products to aggregate multiscale features. The experimental results demonstrate that this design not only enhances the classification accuracy but also significantly reduces the inference time. Furthermore, ablation experiments show that the DA unit, which combines position attention and channel attention, complements the MA units.

Pixels should not be sensitive to their positions in classification tasks (for example, whether maize appears in the upper left or lower right corner, it should be recognized

as maize). However, relying solely on spectral information for hyperspectral classification is quite challenging in hyperspectral tasks due to the presence of “the same object with different spectrums” and “different objects with the same spectrum” phenomena in hyperspectral images. Under such circumstances, pixels must remain sensitive to the spatial location of features to incorporate spatial context information for achieving higher accuracy in hyperspectral image classification. In our DA unit, we used spectrum attention to extract spectral information that is more important for category labels. In addition, we used spatial attention in the DA unit and multiscale attention in the MA unit to extract pixel-rich contextual features. Specifically, we can extract and merge multiscale features using our methods, allowing each pixel to obtain rich contextual information from its neighboring regions. In the meantime, we introduced the Hadamard product in the MA unit for nonlinear feature fusion, enhancing the model’s ability to represent surface features and thus enhancing the accuracy of hyperspectral image classification tasks with a certain number of model parameters.

However, our proposed model also has some shortcomings. For example, the ground object classes in the remote sensing images are prespecified, and it is likely that those classes not seen in the training phase will be classified into some known class during inference. In addition, the classification accuracy of certain ground object classes that are spectrally similar in noisy images, such as corn, corn-notill, and corn-mintill, needs to be further improved.

In recent years, pretraining steps and large models have become popular. Now, combining the world knowledge embedded in large models to improve the inference of deep learning models on unseen classes is the next important research direction. Furthermore, if a deep learning model can both predict what class a sample belongs to and actively explain how it made this prediction, it will be easier to determine why the model makes mistakes, and thus, the robustness and generalization of the deep learning model can be further improved.

6. Conclusions

In this paper, a lightweight hyperspectral image classification model using multiscale feature aggregation and a dual attention mechanism is proposed. The MA unit concatenates multiscale spectral and spatial features through the Hadamard product to address the hyperspectral neighborhood pixel effect. The DA unit can capture contextual features to compensate for the lack of a convolutional field of view. Experiments on three crop scenes and one urban scene show that the method achieves good performance with low computational consumption, outperforming the comparison methods.

Author Contributions: Conceptualization, B.C.; methodology, B.C. and J.W.; software, J.W.; validation, B.C., X.S. and J.H.; investigation, B.C.; resources, J.W.; writing—original draft preparation, J.W.; writing—review and editing, B.C. and J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 42276185) and the Shandong Province Natural Science Foundation of China (Grant Nos. ZR2020MD096 and ZR2020MD099).

Data Availability Statement: The codes used in this study are publicly available on GitHub for the purpose of reproducibility. You can find them at the following link: <https://github.com/wjxcs/MADANet>. We encourage readers and researchers to use these codes and look forward to your feedback and suggestions. We believe that open and shared data will help advance science. Thank you for your attention and support to our work.

Acknowledgments: Thank you to the authors who provided experimental datasets and comparative detectors.

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

References

1. Tan, Y.; Lu, L.; Bruzzone, L.; Guan, R.; Chang, Z.; Yang, C. Hyperspectral Band Selection for Lithologic Discrimination and Geological Mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 471–486. [[CrossRef](#)]
2. Oehlschläger, J.; Schmidhalter, U.; Noack, P.O. UAV-Based Hyperspectral Sensing for Yield Prediction in Winter Barley. In Proceedings of the 2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 23–26 September 2018; pp. 1–4. [[CrossRef](#)]
3. Alias, M.S.; Adnan, A.M.I.; Jugah, K.; Ishaq, I.; Fizree, Z.A. Detection of Basal Stem Rot (BSR) disease at oil palm plantation using hyperspectral imaging. In Proceedings of the 2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Gainesville, FL, USA, 26–28 June 2013; pp. 1–4. [[CrossRef](#)]
4. Wan, Y.; Hu, X.; Zhong, Y.; Ma, A.; Wei, L.; Zhang, L. Tailings Reservoir Disaster and Environmental Monitoring Using the UAV-ground Hyperspectral Joint Observation and Processing: A Case of Study in Xinjiang, the Belt and Road. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 9713–9716. [[CrossRef](#)]
5. Odagawa, S.; Takeda, T.; Yamano, H.; Matsunaga, T. Bottom-type classification in coral reef area using hyperspectral bottom index imagery. In Proceedings of the 2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Tokyo, Japan, 2–5 June 2015; pp. 1–4. [[CrossRef](#)]
6. Taşkın, G.; Kaya, H.; Bruzzone, L. Feature Selection Based on High Dimensional Model Representation for Hyperspectral Images. *IEEE Trans. Image Process.* **2017**, *26*, 2918–2928. [[CrossRef](#)] [[PubMed](#)]
7. Jia, S.; Yuan, Y.; Li, N.; Liao, J.; Huang, Q.; Jia, X.; Xu, M. A Multiscale Superpixel-Level Group Clustering Framework for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5523418. [[CrossRef](#)]
8. Zhu, Q.; Wang, Y.; Wang, F.; Song, M.; Chang, C.-I. Hyperspectral Band Selection Based on Improved Affinity Propagation. In Proceedings of the 2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 24–26 March 2021; pp. 1–4. [[CrossRef](#)]
9. Manzanarez, S.D.; Manian, V. Supervised Spatially Coherent Nonlinear Dimensionality Reduction for Hyperspectral Image Classification. In Proceedings of the 2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 24–26 March 2021; pp. 1–5. [[CrossRef](#)]
10. Gong, J.; Liu, W.; Pei, M.; Wu, C.; Guo, L. ResNet10: A lightweight residual network for remote sensing image classification. In Proceedings of the 2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Changsha, China, 15–16 January 2022; pp. 975–978. [[CrossRef](#)]
11. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Twenty-Sixth Annual Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Curran Associates Inc.: Red Hook, NY, USA, 2012.
12. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9. [[CrossRef](#)]
13. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
15. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2017; pp. 4700–4708.
16. Alipourfard, T.; Arefi, H.; Mahmoudi, S. A Novel Deep Learning Framework by Combination of Subspace-Based Feature Extraction and Convolutional Neural Networks for Hyperspectral Images Classification. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4780–4783. [[CrossRef](#)]
17. Li, K.; Ma, Z.; Xu, L.; Chen, Y.; Ma, Y.; Wu, W.; Wang, F.; Liu, Z. Depthwise Separable ResNet in the MAP Framework for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5500305. [[CrossRef](#)]
18. Paoletti, M.E.; Haut, J.M.; Fernandez-Beltran, R.; Plaza, J.; Plaza, A.J.; Pla, F. Deep Pyramidal Residual Networks for Spectral–Spatial Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 740–754. [[CrossRef](#)]
19. Liu, X.; Meng, Y.; Fu, M. Classification Research Based on Residual Network for Hyperspectral Image. In Proceedings of the 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 19–21 July 2019; pp. 911–915. [[CrossRef](#)]
20. Wang, Y.; Li, K.; Xu, L.; Wei, Q.; Wang, F.; Chen, Y. A Depthwise Separable Fully Convolutional ResNet With ConvCRF for Semisupervised Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4621–4632. [[CrossRef](#)]
21. Li, A.; Shang, Z. A new Spectral-Spatial Pseudo-3D Dense Network for Hyperspectral Image Classification. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–7. [[CrossRef](#)]
22. Cui, B.; Li, X.; Wu, J.; Ren, G.; Lu, Y. Tiny-Scene Embedding Network for Coastal Wetland Mapping Using Zhuhai-1 Hyperspectral Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 2504105. [[CrossRef](#)]
23. Zhang, C.; Li, G.; Du, S. Multi-Scale Dense Networks for Hyperspectral Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9201–9222. [[CrossRef](#)]

24. Dong, W.; Qu, J.; Zhang, T.; Li, Y.; Du, Q. Context-Aware Guided Attention Based Cross-Feedback Dense Network for Hyperspectral Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5530814. [[CrossRef](#)]
25. Meng, Z.; Jiao, L.; Liang, M.; Zhao, F. A Lightweight Spectral-Spatial Convolution Module for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5505105. [[CrossRef](#)]
26. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
27. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
28. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. *arXiv* **2019**, arXiv:1905.02244.
29. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
30. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; Lecture Notes in Computer Science. Volume 11218.
31. Yuan, Z.; Lu, C. Research on Image Classification of Lightweight Convolutional Neural Network. In Proceedings of the 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Nanchang, China, 26–28 March 2021; pp. 498–501. [[CrossRef](#)]
32. Zhang, H.; Yu, H.; Xu, Z.; Zheng, K.; Gao, L. A Novel Classification Framework for Hyperspectral Image Classification Based on Multi-Scale Dense Network. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 2238–2241. [[CrossRef](#)]
33. Yang, J.; Du, B.; Wu, C.; Zhang, L. Automatically Adjustable Multi-Scale Feature Extraction Framework for Hyperspectral Image Classification. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 3649–3652. [[CrossRef](#)]
34. Guan, L.; Han, Y.; Zhang, P. Hyperspectral image classification based on improved multi-scale residual network structure. In Proceedings of the 2021 2nd International Symposium on Computer Engineering and Intelligent Communications (ISCEIC), Nanjing, China, 6–8 August 2021; pp. 377–382. [[CrossRef](#)]
35. Lin, M.; Jing, W.; Di, D.; Chen, G.; Song, H. Multi-Scale U-Shape MLP for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6006105. [[CrossRef](#)]
36. Shang, R.; Chang, H.; Zhang, W.; Feng, J.; Li, Y.; Jiao, L. Hyperspectral Image Classification Based on Multi-scale Cross-branch Response and Second-Order Channel Attention. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5532016. [[CrossRef](#)]
37. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the CVPR, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
38. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the ECCV, Munich, Germany, 8–14 September 2018; pp. 3–19.
39. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.
40. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 7794–7803.
41. Ren, Y.; Li, X.; Yang, X.; Xu, H. Development of a Dual-Attention U-Net Model for Sea Ice and Open Water Classification on SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4010205. [[CrossRef](#)]
42. Wang, R.; Chen, Z.; Zhang, S.; Li, W. Dual-Attention Generative Adversarial Networks for Fault Diagnosis Under the Class-Imbalanced Conditions. *IEEE Sens. J.* **2022**, *22*, 1474–1485. [[CrossRef](#)]
43. Liu, M.; Hu, Q.; Wang, C.; Tian, T.; Chen, W. Daff-Net: Dual Attention Feature Fusion Network for Aircraft Detection in Remote Sensing Images. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 4196–4199. [[CrossRef](#)]
44. Zhong, Y.; Hu, X.; Luo, C.; Wang, X.; Zhao, J.; Zhang, L. WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sens. Environ.* **2020**, *250*, 112012. [[CrossRef](#)]
45. Zhong, Y.; Wang, X.; Xu, Y.; Wang, S.; Jia, T.; Hu, X.; Zhao, J.; Wei, L.; Zhang, L. Mini-UAV-borne hyperspectral remote sensing: From observation and processing to applications. *IEEE Geosci. Remote Sens. Mag.* **2018**, *6*, 46–62. [[CrossRef](#)]
46. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [[CrossRef](#)]
47. Zhang, Y.; Cao, G.; Li, X.; Wang, B. Cascaded random forest for hyperspectral image classification. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2018**, *11*, 1082–1094. [[CrossRef](#)]

48. Song, W.; Li, S.; Fang, L.; Lu, T. Hyperspectral image classification with deep feature fusion network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3173–3184. [[CrossRef](#)]
49. Roy, S.K.; Krishna, G.; Dubey, S.R.; Chaudhuri, B.B. HybridSN: Exploring 3D-2D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 277–281. [[CrossRef](#)]
50. Ding, L.; Tang, H.; Bruzzone, L. LANet: Local Attention Embedding to Improve the Semantic Segmentation of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 426–435. [[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.