

Multiscale Fusion Network Based on Global Weighting for Hyperspectral Feature Selection

Jinjin Wang , Jiahang Liu , *Member, IEEE*, Jian Cui , Ji Luan, and Yangyu Fu

Abstract—Feature selection (FS) is an important way to achieve high-precision and efficient classification of hyperspectral remote sensing images. However, most existing FS methods use a fixed scale to extract features and the relationship between spatial and spectral dimensions is ignored. In fact, this correlation is useful for classification. In this article, a multiscale feature fusion network based on global weighting (MSFGW) is proposed in which a global weighting mechanism is explored to catch spatial-spectral information at multiple scales. First, the multiscale feature extraction module composed of group convolution and dilated convolution is utilized to extract the multiscale features. With the increase of the dilation rate, the module takes the spatial differences at varying scales. Second, a 3-D weighting mechanism is used to combine the spatial and spectral correlated information for reducing the interference of homologous and heterologous and boosting the feature discrimination ability. Then, multiscale weighted features are fused to integrate the internal information of all bands at different scales. Finally, the band reconstruction network is used to select representative bands according to their entropy. The experimental results with the state-of-the-art FS algorithms on four widely hyperspectral datasets demonstrate that the features selected by MSFGW have obvious advantages in classification with only a few training samples.

Index Terms—Attention mechanism, feature selection (FS), hyperspectral image (HSI), multiscale spectral features fusion.

I. INTRODUCTION

HYPERSPECTRAL imaging is one of the most important remote sensing detection methods because it achieves the effective integration of target spectral acquisition and spatial imaging. The full name of hyperspectral remote sensing is “hyperspectral resolution remote sensing,” which more intuitively reflects the ability of hyperspectral remote sensing to characterize spectral dimension details. Therefore, hyperspectral remote sensing is not only an important means of Earth observation but also an indispensable component of the spatial information network. In addition, it has also played an active role in many tasks, such as urban mapping [1], geological exploration [2], and military surveillance [3].

Manuscript received 13 December 2022; revised 4 February 2023; accepted 13 February 2023. Date of publication 6 March 2023; date of current version 30 March 2023. This work was supported in part by the Government of Jiangsu Province under Grant JSSCR2021501 and in part by National Major Project. (Corresponding author: Jiahang Liu.)

The authors are with the College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: jin.school@nuaa.edu.cn; jhliu@nuaa.edu.cn; cuijiannah@nuaa.edu.cn; luanji@nuaa.edu.cn; yangyufu@nuaa.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3251442

However, hyperspectral remote sensing images have a large number of bands, and there is a great correlation between the bands. This makes the analysis and processing based on hyperspectral images (HSIs) involve a lot of computation and a heavy computational burden. In practical classification applications, after the number of feature dimensions increases to a certain threshold, the performance of classification will deteriorate if the number of features continues to increase, which is called the “Hughes phenomenon” [4], [5]. Accordingly, feature extraction or band selection for dimension reduction is a good choice to overcome the above problems [6].

Feature extraction [7] achieves the purpose of dimensionality reduction by performing different forms of function mapping on the original features. Compared with feature extraction techniques, the features obtained by feature selection (FS) [8], [9] are subsets of the original set of bands, so the physical meaning of the original bands is preserved. Therefore, this article mainly discusses the related issues of band selection.

According to the different FS methods, the band selection is divided into rank-based, cluster-based, search-based, etc. [10]. The rank-based FS method usually quantifies and sorts all the bands according to a certain evaluation criterion and selects the high-priority bands according to the sorting index threshold. For example, maximum variance principal component analysis [11], sparse representation (SpaBS) [12], [13], and geometry-based BS (OPBS) [14]. The selection result is mainly influenced by the ranking criteria. The search-based FS method regards FS as the multiobjective optimization problem, which is essentially an optimization problem of a criterion function, e.g., multiobjective evolutionary algorithm [15], quantum search algorithm [16], and particle swarm optimization [17], [18]. The search process is usually time-consuming. The cluster-based FS method divides the spectrum into multiple clusters according to the task requirements from the perspective of considering the similarity between the bands. Their typical ones are sparse nonnegative matrix factorization clustering (SNMF) [19], affinity propagation clustering [20], [21], and (K-means) clustering [22], [23]. Nevertheless, the existing FS methods have problems, such as large computation, high similarity, and easily falling into local optimal solutions.

Recently, convolutional neural networks (CNNs) [24], [25] in deep learning have received extensive attention by transforming initial “low-level” feature representations into abstract “high-level” representations through multilayer convolutional network structures. This characteristic is suitable for solving

the complex problem of band FS in HSI data. Zhan et al. [26] proposed the algorithm to apply CNN to band selection first, which models the relationship between bands by a simple combination of convolution and pooling. BSNet-Conv [27] first introduced the attention mechanism into band selection, which uses the attention mechanism to consider the global information, and simulates the global nonlinear correlation between spectral bands instead of estimating each band independently. However, the method is weak in capturing long-range contextual information in both spatial and spectral directions. On this basis, Roy et al. [28] utilize a dual attention mechanism (DAM) to capture long-range nonlinear contextual information in spectral and spatial directions and achieves information weighting in both channel and spectral dimensions. Objects at different spatial scales have their specific spectral features, but most algorithms do not address the spatial scales of various objects. They use a uniform spatial scale to measure the feature information of all objects, which may affect the choice of representative bands.

Inspired by the above research, in this article, a multiscale fusion network based on global weighting (MSFGW) is proposed to solve the mentioned problem. Based on the assumption that all bands can be completely reconstructed from a subset of bands, the appropriate band subset is selected according to the contribution to the band reconstruction. First, MSFGW applies a multibranch module consisting of group convolutions and dilated convolutions to extract features with representation ability, fully considering the spatial structure and spatial correlation of the target object. Then, to enhance the impact of useful features and reduce the information dispersion, the 3-D convolution is used to enhance the information interaction between channel and spatial. Finally, the features at different scales are fused by integrating the information flow from different branches, revealing the inner connections of all bands. The proposed model achieves state-of-the-art on several challenging datasets, demonstrating the effectiveness and superiority of the method. The main contributions of this article are as follows.

- 1) The multibranch convolution module is employed to extract various spectral features on different spatial scales. A series of spectral features are extracted under the effect of dilation rate and grouped convolution. As the dilation rate increases, the obtained feature data cubes acquire the spectral feature information of the features at an incremental scale.
- 2) The information selection mechanism of human eye vision is simulated by using 3-D attention to realize spatial attention and channel attention together. Weights are applied among channel, spatial width, and spatial height to realize the combination of spatial and spectral information and reduce information loss.

The rest of this article is organized as follows. We first define the notations and review the basic concepts of group convolution and dilated convolution in Section II. Second, we introduce the proposed MSFGW for hyperspectral band selection in Section III. Next, in Section IV, we explain the experiments on four hyperspectral datasets and compare them with many existing FS methods. Finally, Section V concludes this article.

II. PRELIMINARY

A. Definition and Notations

For convenience, in this article, the 3-D HIS cube is represented as $I \in \mathbb{R}^{W \times H \times C}$, where W and H are the length and width of the band image, respectively, and C represents the number of bands. So, I can be regarded as the set containing C band images $B = \{B_1, B_2, \dots, B_C\}$. The target of the band selection is to select a subset D from the set B that meets the task requirements, where D consists of b bands and satisfies $D \subseteq B, D \in \mathbb{R}^{W \times H \times b}$. In addition, only considering point pixels will ignore the information related to the spatial arrangement of pixels in the scene. Therefore, in the data processing stage, a 3-D neighborhood block $P \in \mathbb{R}^{S \times S \times C}$ is extracted from the original image I . Taking the spatial position $x(i, j)$ as the center, where $i = 1, 2, \dots, W$ and $j = 1, 2, \dots, H$, its ground-truth label is determined by this pixel. For convenience, the input and output of the neural network are represented by tensors. For example, the input of the convolutional layer is represented as $X \in \mathbb{R}^{N \times M \times C}$, where $N \times M$ is the spatial size of the input feature map, and C is the number of channels.

B. Group Convolution

HSIs have the typical characteristics of a large amount of data, so the convolution processing based on HSIs is computationally intensive. Group convolution [29] can effectively alleviate this problem without affecting the results.

The inspiration for group convolution comes from Inception [30] and AlexNet [29] that separates the convolution of channel dimension and spatial dimension. The feature mappings obtained by different convolutional paths are less coupled with each other and the features of interest are different, so better results can be obtained. As the convolution can be split into multiple paths, the model can be trained on multiple GPUs in parallel. Moreover, the model parameters will decrease as the number of groups increases, so it has the characteristics of efficient training.

As shown in Fig. 1, Fig. 1(a) represents the standard convolution operation. Suppose the size of the input feature is $H_1 \times W_1 \times C_1$, where the size of the convolution kernel is $h_1 \times w_1 \times C_1$, and the number is C_2 . The final output is $H_2 \times W_2 \times C_2$. Then, the parameter quantity of the convolutional layer is $h_1 \times w_1 \times C_1 \times C_2$. Fig. 1(b) represents the group convolution operation. Assuming that the input feature map is divided into two groups, the input feature size of each group is $H_1 \times W_1 \times (C_1/2)$. The size of convolution kernels is $h_1 \times w_1 \times (C_1/2)$, and the number is $(C_2/2)$. The output feature map size of each group is $H_2 \times W_2 \times (C_2/2)$. The parameters of the two groups of convolutions are $h_1 \times w_1 \times (C_1/2) \times (C_2/2) \times 2 = (h_1 \times w_1 \times C_1 \times C_2)/2$. From the above example, it can be concluded that the parameter amount of the group convolution is $1/g$ of the regular convolution, where g is the number of groups (the number of groups in Fig. 1(b) is 2).

C. Dilated Convolution

In the deep convolution network, downsampling (such as pooling) is performed to increase the receptive field and reduce

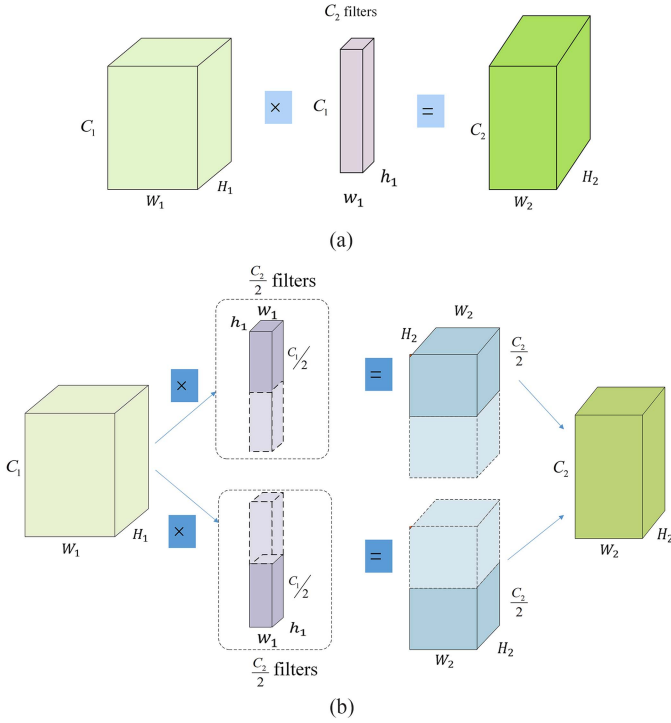


Fig. 1. Flowchart of different convolution methods. (a) Convolution. (b) Group convolution.

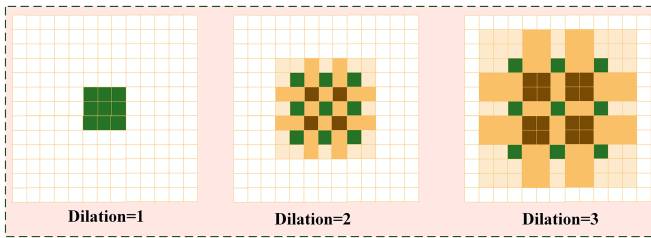


Fig. 2. Schematic diagram of convolution with different dilated rates.

the computation frequently. Although the receptive field can be increased in this way, the spatial resolution will be reduced, which will directly affect the subsequent application of HSIs. To expand the receptive field without losing resolution, dilated convolution [31], [32] is utilized to retain more feature map information. Assuming that a variable α is used as the expansion coefficient to measure the dilation convolution, the relationship between the dilated convolution kernel size and the original convolution is $K = k + (k - 1)(\alpha - 1)$, where k is the kernel size of the original convolution, and α is the dilation rate. The dilation rate can represent the degree of convolution kernel expansion. As illustrated, when $\alpha = 1, 2$, and 3 , the receptive field of the convolution kernel is shown in Fig. 2.

As shown in Fig. 2, although the size of the three convolutional kernels is the same, that is 3×3 , the receptive field observed by the model is different. In Fig. 2(a), when $\alpha = 1$, the size of the convolution kernel at this time is 3×3 , which is the same as the general convolution. In Fig. 2(b), when $\alpha = 2$, the size of the convolution kernel of the dilation convolution is 5 , here the

receptive field is 7×7 . Similarly, when $\alpha = 3$, the convolution kernel size will change to 7 , and the receptive field can grow to 11×11 . From this, dilated convolution can obtain the larger receptive field without increasing the cost of parameter operation.

III. PROPOSED NETWORK

This section mainly introduces the backbone structure and various components of the proposed band selection network, including the multiscale feature extraction part, the 3-D feature weighting, the feature fusion part, and the band reconstruction network. The main idea is to extract the target features in the multiscale spatial first. Second, different branches are weighted in both channel and spatial dimensions based on the consideration of the importance of cross-dimensional interactions between the different bands. Then, to obtain the complete information of all bands, the output features of each branch are fused in a summation manner. Finally, the fused features are applied to the band reconstruction, and the subset of bands that contribute most to the band reconstruction is selected as the final result.

A. Multiscale Feature Extraction

Most of the existing algorithms based on CNN model use the combination of convolution and pooling operations to learn features, and the feature scale extracted by convolution is single. However, in reality, the size and shape of objects in images are different, so the features extracted from a uniform size are not enough to meet the needs of complex situations. This requires different sizes of receptive fields to obtain contextual information. For images containing different objects or images with different resolutions, learning object features from different scales can more compactly understand the spatial structure of objects.

Inspired by the work of the Inception model [30] and others [33], [34], this article attempts to take target characteristics from different scales. The Inception model first attempts multibranch convolution with different kernel sizes, which extends the convolution operation between layers of the neural network, resulting in different sizes of perceptual fields. Similar to the Inception structure, this section designs the multibranch module composed of dilated convolution to extract features from different receptive fields. The module decomposes the feature extraction of each image patch into three different parts (branch A, branch B, and branch C) in a manner of increasing the dilation rate to describe the spatial characteristics of different scales.

Specifically, the multibranch feature extraction module is shown in Fig. 3. The multibranch module consists of three branches, and the characteristics of different scale are extracted from the input image cube in an increasing manner. For the input image cube $X \in R^{S \times S \times C}$, apply different operations separately: $F' : X \rightarrow U_1 \in R^{S \times S \times C}$, $F'' : X \rightarrow U_2 \in R^{S \times S \times C}$, $F''' : X \rightarrow U_3 \in R^{S \times S \times C}$, where F' , F'' , and F''' are composed of grouped convolution and dilated convolution. The kernel size of the three branches is 3×3 . However, the dilated rate of branch A is 1 , the dilated rate of branch B is 2 , and the branch C is 3 . From the content of Section II, we can see that the dilated rate has a direct relationship with the receptive field obtained by convolution. The larger the dilated rate, the larger the size of

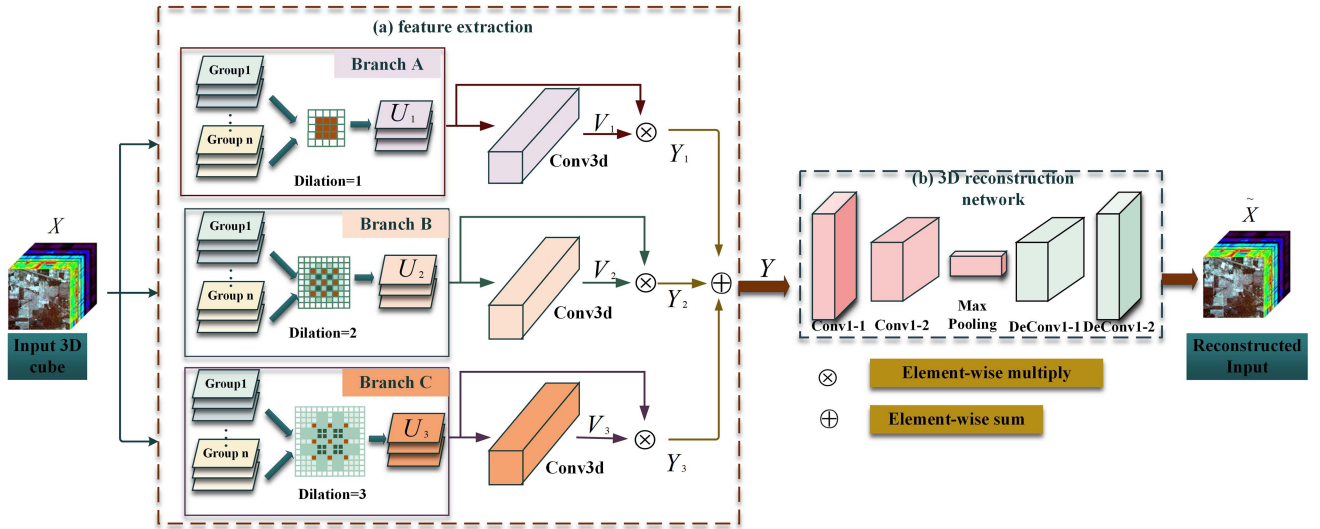


Fig. 3. Overview of the proposed multibranch network for band selection.

the receptive field obtained. In this article, different dilated rates are used to achieve the effect of multiscale feature extraction by multibranch modules. With the action of the dilated rate, the convolution results obtained for each branch are the same as using convolution kernel sizes of 3×3 , 5×5 , and 7×7 . The finally extracted features contain context information of multiple receptive fields without significantly increasing the number of parameters. In addition, to reduce the amount of calculation, group convolution is adopted. After inputting the image cube into different branches, the same grouping is performed in the channel dimension. Each group of images is convoluted separately and then spliced into a complete image. The outputs of the final multiscale feature extraction module are U_1 , U_2 , and U_3 , which contain the feature information on different scales for the next step. However, since the number of image bands in the UP dataset is 103, it cannot be grouped. So, in the experiment of this dataset, the operation of grouped convolution is not applied. The experiments are carried out in the way of ordinary convolution.

B. Feature Weighting and Fusion

The rich features extracted by the multibranch module describe the characteristics of the target at different scales, but the features are not all useful for the task inevitably. Therefore, it is important to measure the importance of features to filter out useful features. In prioritizing feature importance, correlation operations are usually performed from two dimensions, channel and spatial [35], [36]. But they often ignore the global interaction between spectral and spatial information of bands.

In contrast to previous weighting mechanisms, we adopt a 3-D convolution method, taking into account the interaction information of spatial and channel. The main difference between the 2-D convolution and 3-D convolution is the spatial dimension of the filter sliding. In 3-D convolution, 3-D filters can move in all three directions (the height, width, and channel) of all three directions. At each position, the multiplication and addition of elements will provide a value. Because the filter slides through a 3-D space,

the output is a 3-D data. The advantage of 3-D convolution is to describe the object relationship in the 3-D space. There is also the advantage of reducing information dispersion while capturing important features in three dimensions. The size of the convolution kernel is set to $(1, 3, 3)$, the stride is set to $(1, 1, 1)$, and the padding is set to $(1, 1, 0)$.

In three branches, the output of the multiscale module U_1 , U_2 , and U_3 is subjected to a 3-D convolution operation

$$V_t = \text{conv3d}(U_t, \Theta_b^t) \quad (1)$$

where Θ_b^t denotes the trainable parameters involved in the conv3d. V_t represents the output of the third branch, $t < 3$. The spatial and channel 3-D feature weights obtained by the weighting module can be used to measure the characterization capability of the features. In addition, considering the importance of the features from a global perspective, spatial and channel information fusion can be used to maximize the retention of the 3-D information of the features. To create an interaction between the original input and the weights, the output of the multiscale feature extraction module is multiplied by the weight matrix to improve important features and suppress unwanted features

$$Y_t = U_t \otimes V_t. \quad (2)$$

The information contained in Y_1 , Y_2 , and Y_3 from the three branches are different, so the single output cannot maximize the effect of contextual information. To make the final output contain features of different scales, the cross-channel connection is adopted for information fusion next. For the weighted features from each branch, the information is integrated by a summation function to achieve the complementarity of spectral features on different spatial scales

$$Y = Y_1 \oplus Y_2 \oplus Y_3. \quad (3)$$

The final output Y achieves the fusion of multiscale features, which is beneficial to uncover the real structure of all spectral bands.

C. Reconstruction Network

The evaluation of the selected bands is based on the assumption that the spectral bands can be reconstructed sparsely with a small number of informative bands. For the selected bands set, if the band reconstruction is performed better, it must contain more useful information. Therefore, to demonstrate the representativeness of the selected bands, a reconstruction network is applied to achieve the reconstruction from the weighted fused images to the original spectral bands [27]. The band with the highest entropy ranking in the final reconstruction band set is the selected band. For convenience, the band reconstruction network is defined as a function Φ with the multiscale weighted output Y as input

$$\tilde{X} = \Phi(Y, \Theta_c) \quad (4)$$

where Θ_c denotes the trainable parameters involved in the reconstruction network. The reconstruction network is a completely symmetric coding and decoding structure strategy in which the encoder mainly analyzes the object information, and then the decoder corresponds the parsed information into the final image form. The encoder is mainly composed of the convolution layer, pooling layer, and batch normalization layer, but the difference is that the decoder adopts deconvolution. The encoder classifies and analyzes the low-level local pixel values of the image to obtain higher level semantic information. The decoder upsampling the feature image containing high-order information and then convolves the upsampled image to restore the geometric shape of the object. To illustrate the effectiveness of the network, it is measured by mean-square error

$$L_1(\Theta_d, \Theta_e) = \frac{1}{2S_{\text{tra}}} \sum_{i=1}^{S_{\text{tra}}} \|x_i - \tilde{x}_i\|_1 \quad (5)$$

where $x \in X$, $\tilde{x} \in \tilde{X}$, $\tilde{X} \in R^{S \times S \times C}$, is the reconstructed output for the given input $X \in R^{S \times S \times C}$, and S_{tra} is the number of training samples. For quantitative analysis of selected subsets of bands, the entropy and mean spectral divergence (MSD) of the reconstructed bands were calculated. According to Shannon's entropy theorem, entropy is related to the image information contained in the band. The larger the MSD value, the less redundancy between the selected bands [37], [38]

$$H(c_i) = - \sum_h p(h) \log(p(h)) \text{ s.t. } \sum_h p(h) = 1 \quad (6)$$

where h is the gray level of histogram bins in a band consisting of $S \times S$ pixels and $p(h)$ is the probability that h occurs

$$\text{MSD} = \frac{2}{b(b-1)} \sum_{i=1}^b \sum_{j=1}^b D_{\text{SKL}}(C_i \| C_j) \quad (7)$$

where D_{SKL} is the symmetrical Kullback–Leibler divergence, which measures the dissimilarity between C_i and C_j . Specifically, D_{SKL} is defined as follows:

$$D_{\text{SKL}}(C_i \| C_j) = D_{\text{KL}}(C_i \| C_j) + D_{\text{KL}}(C_j \| C_i). \quad (8)$$

And $D_{\text{KL}}(C_i \| C_j)$ is calculated from gray-level histogram bins.

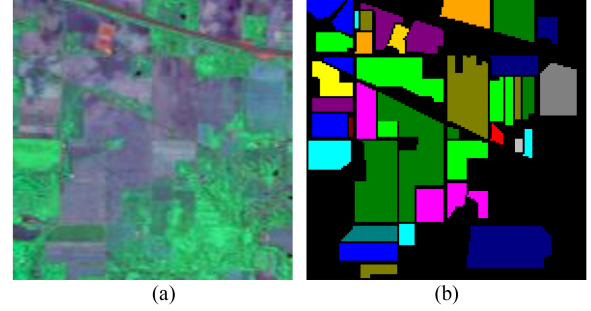


Fig. 4. Indian Pines dataset. (a) False-color image of the Indian Pines data. (b) Ground-truth map of the Indian Pine data.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

To verify the effectiveness of the proposed method, we validate our experimental results on four widely used datasets. In addition to qualitative analysis, three popular quantitative analysis standards, overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa), are also used as experimental evaluation indicators. To illustrate the advancement of the proposed algorithm, it is compared with the state-of-the-art band selection methods, such as LvaHAI [39], EGCSR_BS [40], IBRA-GSS [41], and NGNMF-E2DSSA [42]. Furthermore, comparative experiments of all bands are added to intuitively analyze the performance. According to the relevant contents of the article presented in [43], we uniformly extract 3-D patches of size $9 \times 9 \times C$ for training, where band C of IP, UP, and SA datasets is set to 200, 103, and 204. FDSSC [44] is used as the classifier to verify the classification performance of all band selection algorithms.

The entire framework is implemented in PyTorch with CUDA 10.1. For all datasets, various FS methods are independently computed ten times, with a learning rate of 0.0001, num epochs are 200, and batch size is 32.

A. Hyperspectral Datasets

In this section, we use four well-known HSI datasets (ie., Indian Pines, University of Pavia, Salinas Scene and WHU-Hi-LongKou) to demonstrate the classification performance of the proposed method.

Indian pine is the earliest test data for HSI classification. In 1992, the airborne visible infrared imaging spectrometer (AVIRIS) imaged an Indiana pine tree in Indiana, USA. The ground object is imaged for 220 consecutive wavebands. However, since the 104–108, 150–163, and 220 bands cannot be reflected by water, the remaining 200 bands after the removal of these 20 bands are generally taken as research objects. The data size is 145×145 , including 16 classes of ground objects, as shown in Fig. 4 and Table I. The spatial resolution of the image formed by the spectral imager is about 20 m, so it is easy to generate mixed pixels, which makes classification difficult

The data of Pavia University are a part of the hyperspectral data of Pavia, Italy, imaged by the German airborne reflective optical spectral imager in 2003. The spectral imager continuously imaged 115 wavebands and the spatial resolution of the resulting image was 1.3 m. Among them, 12 bands are deleted due to the

TABLE I
NUMBER OF SAMPLES AVAILABLE IN THE INDIAN PINES DATASET


























No.	color	class	sample
1		Alfalfa	46
2		Corn-notill	1428
3		Corn-mintill	830
4		Corn	237
5		Grass-pasture	483
6		Grass-tress	730
7		Grass-pasture-mowed	28
8		Hay-windrowed	478
9		Oats	20
10		Soybean-notill	972
11		Soybean-mintill	2455
12		Soybean-clean	593
13		Wheat	205
14		Woods	1265
15		Buildings-grass-trees-drives	386
16		Stone-steel-towers	93
Total			10249

TABLE II
NUMBER OF SAMPLES AVAILABLE IN THE PAVIA UNIVERSITY DATASET

No.	color	class	sample
1		Asphalt	6631
2		Meadows	18649
3		Gravel	2099
4		Trees	3064
5		Painted metal sheets	1345
6		Bare soil	5029
7		Bitumen	1330
8		Self-modulating bricks	3682
9		Shadows	947
Total			42776

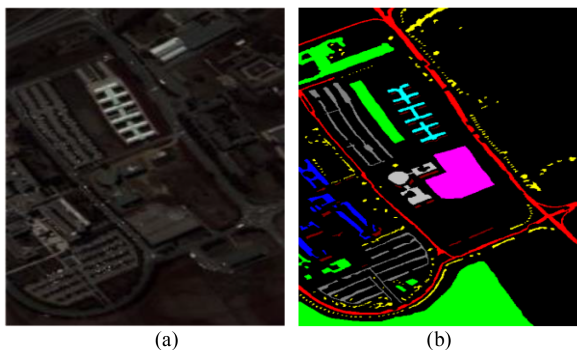


Fig. 5. Pavia University dataset. (a) False-color image of the Pavia University data. (b) Ground-truth map of the Pavia University data.

influence of noise, so the image formed by the remaining 103 spectral bands is generally used, and the data size is 610×340. The information on nine classes of the main ground objects is shown in Fig. 5 and Table II.

Salinas data were taken by AVIRIS imaging spectrometer, which imaged Salinas Valley in California, USA. Unlike the Indian pine dataset, its spatial resolution is 3.7 m. The image initially has 224 bands. Generally, 204 bands remain after removing 108–112, 154–167, and the 224th band that cannot be reflected by water. The size of the image is 512×217, which is divided into 16 classes, as shown in Fig. 6 and Table III.

The WHU-Hi-LongKou dataset was collected in July 2018 in Longkou Town, Hubei Province, China. This dataset is equipped with an 8 mm focal length head-wall nano-ultrahigh specification imaging sensor on the DJI Matrice 600 Pro (DJI M600 Pro)

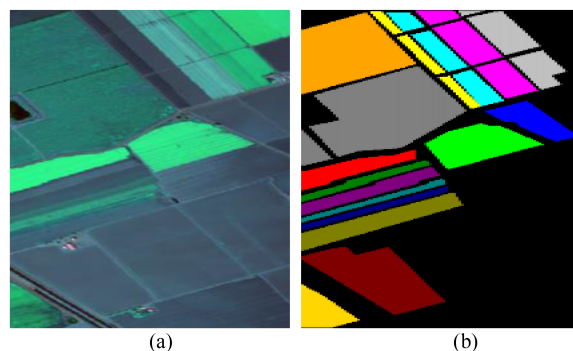


















Fig. 6. Salinas dataset. (a) False-color image of the Salinas data. (b) Ground-truth map of the Salinas data.

TABLE III
NUMBER OF SAMPLES AVAILABLE IN THE SALINAS DATASET

No.	color	class	sample
1		Brocoli_green_weeds_1	2009
2		Brocoli_green_weeds_2	3726
3		Fallow	1976
4		Fallow_rough_plow	1394
5		Fallow_smooth	2678
6		Stubble	3959
7		Celery	3579
8		Graps_untrained	11271
9		Soil_vinyard_develop	6203
10		Corn_cenesced_green_weed	3278
11		Lettuce_romaine_4wk	1068
12		Lettuce_romaine_5wk	1927
13		Lettuce_romaine_6wk	916
14		Lettuce_romaine_7wk	1070
15		Vinyard_untrained	7268
16		Vinyard_vertical_trellis	1807
Total			54129

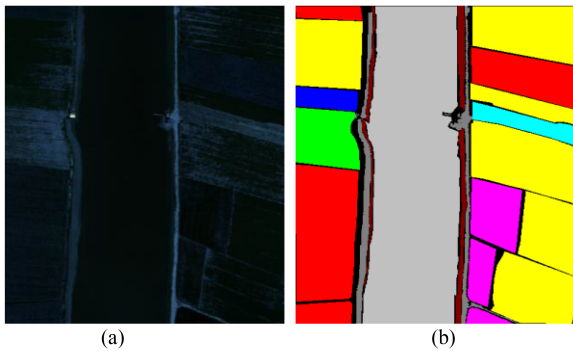











Fig. 7. WHU-Hi-LongKou dataset. (a) False-color image of the WHU-Hi-LongKou data. (b) Ground-truth map of the WHU-Hi-LongKou data.

TABLE IV
NUMBER OF SAMPLES AVAILABLE IN THE WHU-HI-LONGKOU DATASET

No.	color	class	sample
1		Corn	34511
2		Cotton	8374
3		Sesame	3031
4		Broad-leaf soybean	63212
5		Narrow-leaf soybean	4151
6		Rice	11854
7		Water	67056
8		Roads and houses	7124
9		Mixed weed	5229
Total			204542

drone platform. It mainly shoots for agricultural scenes, including nine crops: such as corn, cotton, sesame, etc.

The image size is 550×400 pixels with 270 bands, and the spatial resolution of the UAV hyperspectral imagery is about 0.463 m. The dataset is shown in Fig. 7 and Table IV.

B. Results on Indian Pines Dataset

To prove the effectiveness of the proposed algorithm, we conducted two experiments using various FS methods. First, hyperspectral classification is performed using band subsets of different sizes, ranging from 5 to 30. Second, training samples of different sizes are used for classification, ranging from 1% to 25%. To ensure the reliability of the experimental results, the training and test sets are randomly selected and each method was run ten times independently.

1) *Classification Performance With Different Numbers of Selected Bands*: Fig. 8(a)–(c) shows the average comparison results of OA(%), AA(%), and kappa, respectively. It can be seen that MSFGW achieves the best OA(%), AA(%), and Kappa. When greater than 15, the classification accuracy of LvaHAI [39] is higher than that of BSNet-Conv [27]. The results show that the graph learning algorithm has advantages in mining HSI band clustering structure over using spectral information alone. From the trend of the results, compared with other methods, the algorithm based on deep learning is powerful. In addition,

the experimental results can verify the aforementioned Hughes phenomenon that the classification accuracy does not always increase with the increase of the number of bands. For example, both BSNet-Conv [27] and NGNMF-E2DSSA [42] have a clear downward trend when the number of bands exceeds 15. This phenomenon occurs earlier in SpaBS [12], and when the band is greater than 10, there is a downward trend. However, MSFGW occurs the phenomenon later than the other algorithms, which means that the algorithm in this article is more robust and efficient. In the interval of 20–25, the classification accuracy is the best, which is consistent with the results of virtual dimension (VD) analysis [45], [46] evaluated using the false alarm probability $pF = 10^{-5}$.

2) *Classification Performance With Different Proportions of Training Samples*: Fig. 8(d)–(f) shows the classification performance under different percentages of training samples. In the experiment, we fixed the band subset to 25 and changed the training size from 1% to 25% in 5% intervals. The results showed that MSFGW significantly outperformed other FS methods in terms of OA(%), AA(%), and Kappa. The classification accuracy keeps increasing as the number of training samples increases.

In Table V, the detailed classification performance of different methods is presented by selecting the best 25 bands and using 1% of the training samples. In the end, MSFGW achieves the highest OA (87.36%), AA (88.32%), and Kappa (0.85). Among the 16 classes, 8 classes win, and the individual classes exceeded the suboptimal results by 10% or 20%. The suboptimal algorithm wins three classes, but most of them were similar to the results of MSFGW, and the advantage are not obvious. According to the analysis, the poor results obtained in class 4 are due to the Corn class being similar to the Corn-min-till class and the Grass-pasture-mowed class, thus making it easy to produce incorrect classification results. The overall analysis of MSFGW has no obvious disadvantages and has a strong ability to classify various targets with a small number of training samples. Fig. 9 shows the classification results of different FS methods. The shape of various objects is well preserved, and the internal smoothness is higher. This proves the powerful capability of the features extracted by MSFGW and the effectiveness of HSI classification.

C. Results on Pavia University Data

To prove the applicability of the model, we did the same experiment on the Pavia University dataset in this experiment.

1) *Classification Performance With Different Numbers of Selected Bands*: We show the obtained OA(%), AA(%), and Kappa in Fig. 10(a)–(c), respectively. As can be seen, the proposed FS network significantly outperforms other FS methods. In particular, after the number of selected bands is greater than 10, the classification accuracy and average accuracy results are even superior to all bands, and the best can obtain about 2% improvements. The fact that the accuracy increases first and then decreases with the increase of the number of bands confirms the Hughes phenomenon exactly. MSFGW produces a larger improvement in results compared with BSNet-Conv [27], which also indicates that the combination of spatial and

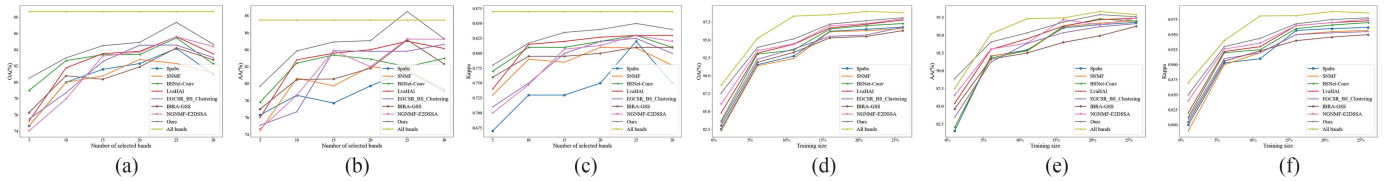


Fig. 8. Performance comparison of different methods with (a)–(c) different band subset sizes and (d)–(e) different training sample sizes on the Indian Pines dataset. (a) OA. (b) AA. (c) Kappa. (d) OA. (e) AA. (f) Kappa.

TABLE V
CLASSIFICATION PERFORMANCE OF DIFFERENT METHODS USING 25 BANDS ON THE INDIAN PINES DATASET

No.	SpaBS	SNMF	BSNet-Conv	LvaHAI	EGCSR_BS_Clustering	IBRA-GSS	NGNMF-E2DSSA	OURS
1	64.57 ± 27.12	66.78 ± 26.26	81.63 ± 19.08	72.75 ± 9.90	66.53 ± 16.21	62.14 ± 18.12	84.48 ± 17.10	86.93 ± 25.82
2	76.96 ± 10.55	75.37 ± 8.92	81.18 ± 10.69	77.06 ± 11.74	81.03 ± 7.85	81.01 ± 9.97	77.08 ± 7.64	81.71 ± 11.88
3	76.95 ± 14.97	76.05 ± 15.91	82.68 ± 14.86	82.58 ± 13.26	83.63 ± 10.99	80.19 ± 15.91	86.72 ± 7.28	87.25 ± 16.25
4	73.32 ± 11.19	74.63 ± 11.63	64.37 ± 10.05	79.80 ± 10.03	70.10 ± 9.79	76.96 ± 9.51	78.40 ± 11.41	67.55 ± 26.68
5	97.90 ± 7.28	98.86 ± 2.10	90.01 ± 1.94	91.99 ± 4.65	95.13 ± 4.35	94.95 ± 4.66	92.73 ± 14.09	96.99 ± 5.11
6	91.00 ± 3.89	95.34 ± 6.98	94.11 ± 3.55	95.21 ± 3.15	91.51 ± 3.02	91.92 ± 6.03	94.84 ± 3.52	96.83 ± 5.04
7	55.44 ± 17.32	68.84 ± 22.46	73.07 ± 14.52	76.74 ± 23.15	66.73 ± 21.09	78.73 ± 23.32	67.91 ± 24.49	76.20 ± 17.28
8	98.02 ± 0.29	98.85 ± 2.63	99.35 ± 1.49	99.86 ± 0.25	99.43 ± 0.89	99.48 ± 0.81	99.50 ± 1.35	99.70 ± 3.83
9	62.28 ± 17.62	44.05 ± 21.53	45.75 ± 23.58	68.79 ± 16.20	84.77 ± 7.33	87.56 ± 18.79	58.69 ± 26.70	88.94 ± 30.95
10	81.33 ± 5.47	80.68 ± 8.93	82.16 ± 4.77	83.01 ± 8.61	81.18 ± 6.16	86.4 ± 5.30	80.36 ± 9.30	82.42 ± 6.34
11	81.14 ± 4.79	79.37 ± 8.78	83.14 ± 6.58	83.29 ± 4.70	83.26 ± 5.72	81.25 ± 4.24	84.75 ± 5.23	88.00 ± 5.54
12	78.49 ± 10.87	75.98 ± 8.66	77.54 ± 9.22	78.02 ± 10.72	76.53 ± 7.28	74.19 ± 10.78	86.52 ± 7.84	87.91 ± 7.97
13	99.10 ± 0.57	98.43 ± 1.27	96.26 ± 1.55	99.68 ± 3.39	97.69 ± 3.56	98.91 ± 2.07	97.84 ± 2.26	99.08 ± 3.66
14	92.53 ± 3.15	94.08 ± 4.83	91.80 ± 2.74	93.47 ± 3.01	93.38 ± 3.58	94.27 ± 3.06	95.39 ± 3.03	94.01 ± 2.32
15	80.21 ± 11.02	81.59 ± 13.96	76.44 ± 6.80	84.12 ± 11.38	79.18 ± 16.66	80.81 ± 14.19	84.39 ± 14.89	84.44 ± 5.22
16	95.28 ± 3.81	95.58 ± 3.90	92.58 ± 3.47	93.63 ± 3.62	90.98 ± 5.02	93.67 ± 1.96	94.30 ± 4.29	95.10 ± 4.43
OA(%)	84.65 ± 5.06	82.56 ± 6.09	85.45 ± 4.64	85.59 ± 5.24	84.57 ± 3.70	84.22 ± 4.87	85.59 ± 3.72	87.36 ± 6.70
AA(%)	81.52 ± 5.91	81.53 ± 6.09	82.00 ± 5.04	85.00 ± 4.36	83.82 ± 3.97	85.15 ± 4.00	85.25 ± 5.50	88.32 ± 7.72
Kappa	0.82 ± 0.05	0.80 ± 0.07	0.83 ± 0.05	0.83 ± 0.05	0.82 ± 0.04	0.81 ± 0.05	0.83 ± 0.04	0.85 ± 0.07

The bold entities best classification result of each class.

spectral information better characterizes the internal structure of hyperspectral responsibility. The results of LvaHAI [39] are similar to BSNet-Conv [27], but the optimal results appear earlier as the number of choices increases. The other algorithms have poorer classification results with large fluctuations on this dataset, so the algorithms are not as robust. From the curves of changes, it can be derived that MSFGW achieves the best classification performance when the band subset is around 15. Similarly, according to the VD analysis [45], [46], by setting the false alarm probability $pF = 10^{-5}$, the optimal subset size for the University of Pavia dataset is 13. It is consistent with the experimental results in Fig. 10.

2) *Classification Performance With Different Proportions of Training Samples:* Fig. 10(d)–(f) shows the classification performance using different numbers of training samples. Compared with other FS methods, MSFGW achieves the best classification performance. Specifically, MSFGW and all bands show comparable performance. Under the same training samples, such as 10% and 15%, the results are similar to all bands overall. In Table VI, we compare the detailed classification performance by setting the band subset size to 15 and the training

size to 1%. Finally, under the premise of limited samples, our network achieves OA(%) of 90.23%, AA(%) of 91.36%, and Kappa of 0.86. Meanwhile, OA(%) for all bands is 89.02%, AA(%) is 90.2%, and Kappa is 0.88. Especially, our algorithm wins four classes and the suboptimal algorithm wins one class in the classification of nine classes. It can be seen from the classification results that MSFGW has obvious advantages, and the classification results of the nine classes are far superior to other algorithms. At the same time, although the accuracy of other classes is not the best, the gap is relatively small, which proves the effectiveness of the algorithm. Fig. 11 shows the classification results of different FS methods.

D. Results on Salinas Dataset

Similarly, we also verified the algorithm on the Salinas dataset. Similar to the Indian Pines and Pavia University dataset, it uses different band subsets and training samples of different sizes for experiments.

1) *Classification Performance With Different Numbers of Selected Band:* We show the obtained OA(%), AA(%), and kappa

TABLE VI
CLASSIFICATION PERFORMANCE OF DIFFERENT METHODS USING 15 BANDS ON THE PAVIA UNIVERSITY DATASET

No.	SpaBS	SNMF	BSNet-Conv	LvaHAI	EGCSR_BS_Clustering	IBRA-GSS	NGNMF-E2DSSA	OURS
1	94.11 ± 5.08	92.92 ± 3.66	92.40 ± 3.04	93.96 ± 2.45	93.54 ± 2.60	90.11 ± 6.21	95.01 ± 3.03	94.59 ± 3.94
2	88.73 ± 3.29	88.15 ± 2.46	88.45 ± 3.84	89.42 ± 2.27	87.61 ± 3.02	86.16 ± 2.92	88.94 ± 2.52	91.33 ± 1.06
3	71.44 ± 16.81	82.58 ± 8.03	69.79 ± 18.74	68.59 ± 15.74	78.66 ± 16.72	79.82 ± 13.12	75.04 ± 19.16	87.48 ± 8.22
4	93.16 ± 0.67	97.58 ± 7.39	95.78 ± 1.37	97.30 ± 7.56	99.89 ± 0.11	98.97 ± 2.21	98.60 ± 2.44	94.57 ± 0.20
5	97.02 ± 0.49	99.58 ± 3.52	99.10 ± 3.17	99.58 ± 2.47	98.29 ± 2.63	99.88 ± 0.11	99.24 ± 0.94	98.62 ± 0.08
6	83.20 ± 15.69	73.00 ± 9.66	75.85 ± 11.32	81.85 ± 6.71	77.26 ± 16.26	82.06 ± 12.23	75.40 ± 13.00	87.06 ± 4.67
7	74.04 ± 21.67	60.51 ± 11.98	81.43 ± 15.08	86.87 ± 9.40	78.94 ± 11.36	71.52 ± 11.71	76.64 ± 16.64	85.82 ± 12.66
8	80.41 ± 10.86	81.02 ± 10.62	78.37 ± 10.06	79.61 ± 9.74	76.52 ± 12.35	78.20 ± 6.89	78.86 ± 8.23	84.52 ± 5.11
9	98.37 ± 1.82	97.99 ± 1.90	99.45 ± 1.39	98.29 ± 2.75	97.75 ± 3.31	98.35 ± 3.96	98.83 ± 1.98	98.30 ± 0.03
OA(%)	86.02 ± 3.01	85.95 ± 2.11	86.22 ± 2.76	87.50 ± 1.96	86.44 ± 4.97	85.66 ± 2.58	86.33 ± 2.86	90.23 ± 1.46
AA(%)	86.72 ± 2.47	85.93 ± 2.63	86.74 ± 3.27	88.39 ± 2.78	87.61 ± 4.87	87.23 ± 2.87	87.40 ± 2.77	91.36 ± 1.91
Kappa	0.81 ± 0.03	0.81 ± 0.02	0.81 ± 0.03	0.83 ± 0.02	0.81 ± 0.06	0.80 ± 0.03	0.81 ± 0.03	0.86 ± 0.01

The bold entities best classification result of each class.

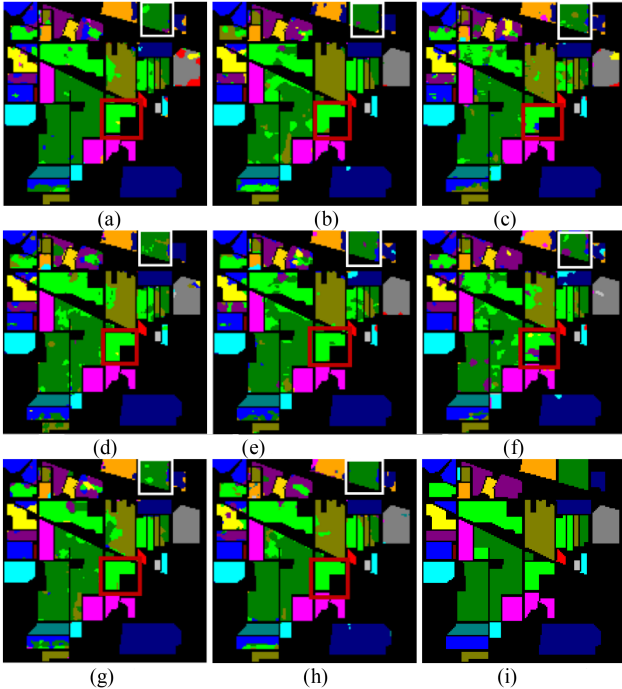


Fig. 9. Performance comparison of different FS methods on Indian Pines Data. (a) SPaBS. (b) SNMF. (c) BSNet-Conv. (d) LvaHAI. (e) EGCSR_BS_Clustering. (f) IBRA-GSS. (g) NGNMF-E2DSSA. (h) MSFGW. (i) Ground truth.

in Fig. 12(a)–(c), respectively. From the results, the proposed FS network can achieve better classification performance than SpaBS [12], EGCSR_BS_Clustering [40], IBRA-GSS [41], and other networks. Moreover, the performance of the MSFGW is not affected by the number of band subsets. The results of BSNet-Conv [27] and LvaHAI [39] continue to outperform other algorithms except for MSFGW, so deep learning algorithms have obvious advantages in datasets of different resolutions. The result of SNMF [19] is the worst, which has a large gap with various algorithms, indicating that it cannot adapt to such complex situations as hyperspectral. It also proves the powerful

capability of the proposed MSFGW. When the band set size is greater than 20, the classification accuracy of most FS methods does not increase anymore, which corresponds to the estimation of VD analysis [45], [46].

2) *Classification Performance With Different Proportions of Training Samples:* Fig. 12(d)–(f) shows the classification performance under different percentages of training samples. In this experiment, we fixed the band set to 20 and changed the training size from 1% to 25% in 5% interval. From the results, even if the training sample is 1%, by comparing EGCSR_BS_Clustering [40], IBRA-GSS [41], and BSNet-Conv [27], our MSFGW obtains at least 2% improvements. With the increase of training samples, the classification accuracy is also growing, and the gap between MSFGW and the results of all bands gradually decreases. In Table VII, we show the detailed classification performance of different methods by selecting the best 20 bands and using 1% of the training samples. The Salinas dataset contains 16 classes, MSFGW wins eight classes and the suboptimal algorithm wins three classes. In particular, class 6 achieves 100% of the classification results, although the training samples are limited. For classes 7, 10, 13, etc., the classification results are close to the best ones, although they are not optimal. The final classification accuracy of MSFGW reached the best OA (90.63%), AA (94.86%), and kappa (0.92). Fig. 13 shows the classification results of different FS methods.

E. Results on WHU_Hi_LongKou Dataset

To prove the effectiveness of the proposed algorithm, in addition to the above classic datasets, the newer dataset WHU_Hi_LongKou is also tested. To be fair, the same verification method as the above experiment is used, that is, the classification performance of different bands and the classification performance of different proportions of training samples.

1) *Classification Performance With Different Numbers of Selected Band:* Fig. 14(a)–(c) shows the comparison results of OA(%), AA(%), and kappa, respectively. It can be seen that for difficult experimental data, our MSFGW still achieves the best OA(%), AA(%), and Kappa. When the number of band

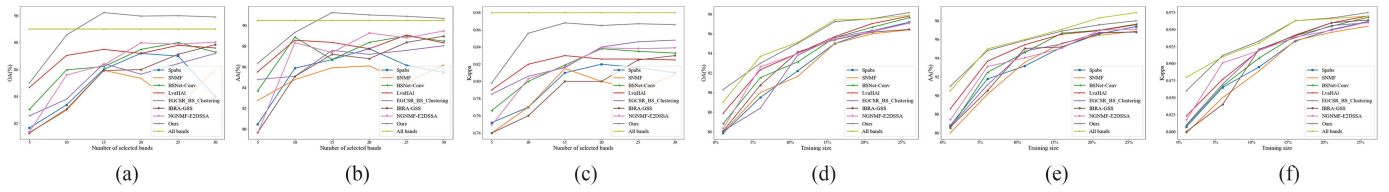


Fig. 10. Performance comparison of different methods with (a)–(c) different band subset sizes and (d)–(e) different training sample sizes on the Pavia University dataset. (a) OA. (b) AA. (c) Kappa. (d) OA. (e) AA. (f) Kappa.

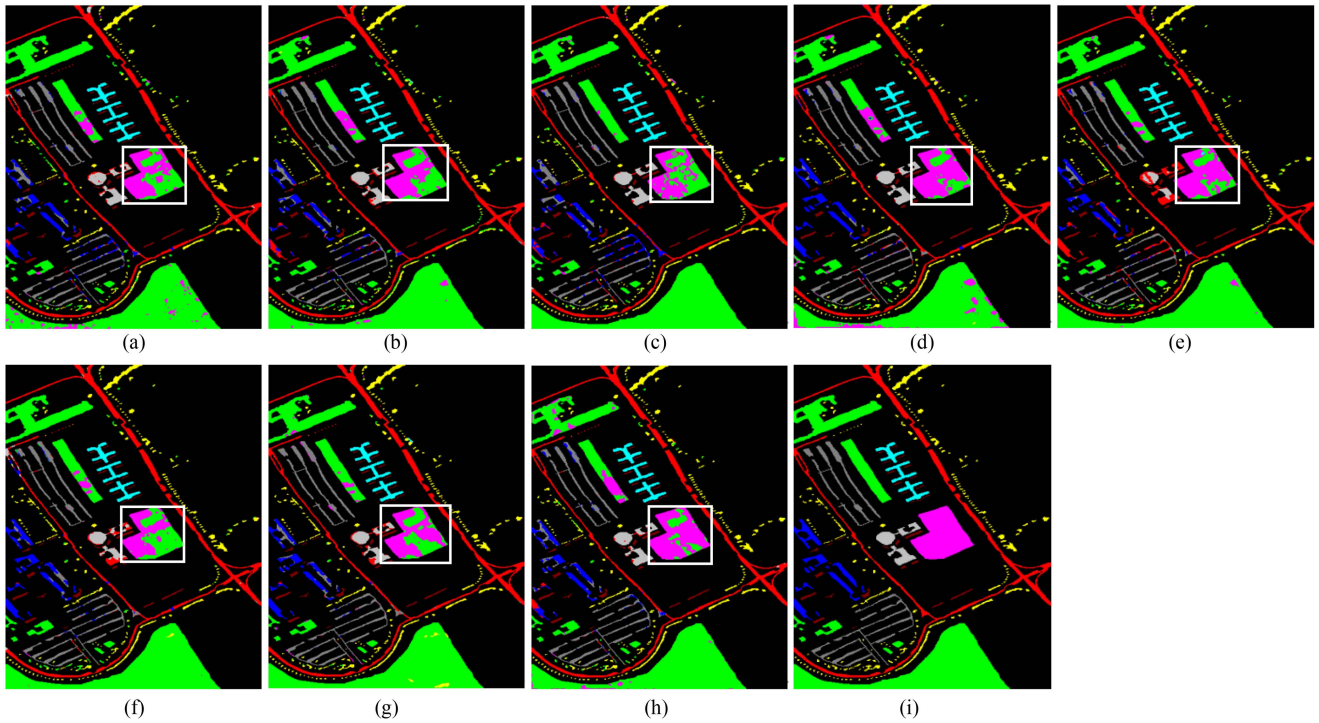


Fig. 11. Performance comparison of different FS methods on Pavia University Dataset. (a) SPABS. (b) SNMF. (c) BSNet-Conv. (d) LvaHAI. (e) EGCSR_BS_Clustering. (f) IBRA-GSS. (g) NGNMF-E2DSSA. (h) MSFGW. (i) Ground truth.

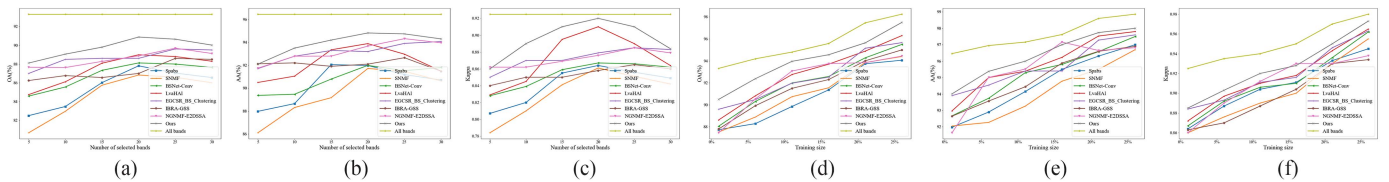


Fig. 12. Performance comparison of different methods with (a)–(c) different band subset sizes and (d)–(e) different training sample sizes on the Salinas Dataset. (a) OA. (b) AA. (c) Kappa. (d) OA. (e) AA. (f) Kappa.

subsets is greater than 15, MSFGW achieves better results than all bands, proving that with the increase of band number, the classification accuracy does not always increase, which is consistent with the content described in the first section. During the whole experiment, even when using the minimum number of bands of 5, MSFGW achieved the best results except for all bands, and it was close to the result of all bands. When the number of bands is around 20, the classification accuracy of most algorithms is optimal. Among them, EGCSR_BS_Clustering [40] and IBRA-GSS [41] perform better than other algorithms

and can achieve results comparable to all bands. Overall, deep learning has more advantages than traditional algorithms in band selection. When the number of bands continues to increase, the accuracy tends to decrease that just verifies the necessity of band selection.

2) *Classification Performance With Different Proportions of Training Samples:* Fig. 14(d)–(f) shows the classification performance under different percentages of training samples. In the experiments, we fixed the subset of bands to 20 and changed the training size from 1% to 25% in 5% interval. The

TABLE VII
CLASSIFICATION PERFORMANCE OF DIFFERENT METHODS USING 20 BANDS ON THE SALINAS DATASET

No.	SpaBS	SNMF	BSNet-Conv	LvaHAI	EGCSR_BS_Clustering	IBRA-GSS	NGNMF-E2DSSA	OURS
1	86.28 ± 2.93	97.58 ± 5.46	100.0 ± 0	94.56 ± 0.32	96.82 ± 3.81	96.76 ± 8.88	98.67 ± 2.17	89.46 ± 20.24
2	94.63 ± 9.68	99.55 ± 0.56	92.56 ± 2.08	99.64 ± 9.02	95.66 ± 6.77	97.41 ± 5.46	95.35 ± 6.88	98.99 ± 0.98
3	86.30 ± 1.06	89.25 ± 7.37	86.08 ± 12.94	94.30 ± 12.53	87.64 ± 12.97	94.12 ± 3.32	95.29 ± 3.97	94.63 ± 5.56
4	96.33 ± 2.42	96.78 ± 3.46	97.14 ± 2.81	97.24 ± 1.81	97.52 ± 1.18	97.93 ± 1.20	96.11 ± 1.96	98.25 ± 1.67
5	93.60 ± 7.32	91.46 ± 6.75	95.37 ± 10.04	96.25 ± 9.46	96.41 ± 1.48	96.55 ± 2.89	95.08 ± 6.73	97.13 ± 3.75
6	99.98 ± 2.25	100.00 ± 0.	99.97 ± 0.02	99.97 ± 0.07	99.98 ± 0.05	99.97 ± 0.05	99.99 ± 7.27	100.0 ± 0
7	98.19 ± 1.66	99.15 ± 1.1	98.53 ± 1.79	98.86 ± 0.93	99.80 ± 0.35	98.46 ± 2.20	97.58 ± 2.82	97.76 ± 4.44
8	79.35 ± 5.41	80.73 ± 6.35	78.77 ± 6.33	76.76 ± 6.93	79.58 ± 7.25	74.95 ± 7.92	79.47 ± 8.05	85.60 ± 6.84
9	97.86 ± 1.34	98.91 ± 0.49	96.97 ± 5.24	99.43 ± 5.45	99.29 ± 0.38	97.94 ± 1.89	98.77 ± 1.11	99.75 ± 0.15
10	91.54 ± 4.45	85.23 ± 7.02	90.24 ± 4.39	95.32 ± 7.75	93.59 ± 6.34	93.38 ± 3.99	94.94 ± 3.33	94.50 ± 2.38
11	77.30 ± 1.42	67.51 ± 10.1	74.03 ± 24.42	87.23 ± 33.24	93.82 ± 5.71	88.36 ± 11.69	95.26 ± 2.94	94.12 ± 5.23
12	94.42 ± 8.73	94.46 ± 5.03	96.01 ± 1.23	97.09 ± 1.96	94.12 ± 8.96	86.55 ± 11.5	94.41 ± 3.63	96.13 ± 3.52
13	98.57 ± 1.82	91.55 ± 6.44	94.56 ± 2.12	97.26 ± 28.43	96.22 ± 3.49	92.13 ± 4.21	96.10 ± 2.59	96.88 ± 3.09
14	97.05 ± 1.70	95.68 ± 3.92	97.14 ± 4.04	97.78 ± 12.99	93.19 ± 10.52	96.42 ± 4.83	97.58 ± 2.7	98.45 ± 2.37
15	76.40 ± 1.03	64.10 ± 7.88	75.70 ± 10.28	70.04 ± 3.80	69.84 ± 9.07	68.74 ± 10.97	74.68 ± 1.32	76.49 ± 14.00
16	98.68 ± 2.19	99.50 ± 0.48	98.45 ± 0.20	99.51 ± 0.48	99.22 ± 0.80	93.93 ± 18.05	89.47 ± 1.4	99.61 ± 0.51
OA(%)	88.00 ± 2.08	86.53 ± 1.86	88.12 ± 3.13	88.99 ± 3.71	88.61 ± 2.37	87.00 ± 3.60	88.87 ± 3.05	90.63 ± 3.99
AA(%)	91.66 ± 2.90	91.72 ± 1.74	91.97 ± 3.01	93.89 ± 5.66	93.29 ± 2.31	92.10 ± 2.53	93.67 ± 1.6	94.86 ± 1.95
Kappa	0.86 ± 0.02	0.86 ± 0.02	0.86 ± 0.03	0.91 ± 0.04	0.87 ± 0.02	0.85 ± 0.04	0.87 ± 0.03	0.92 ± 0.04

The bold entities best classification result of each class.

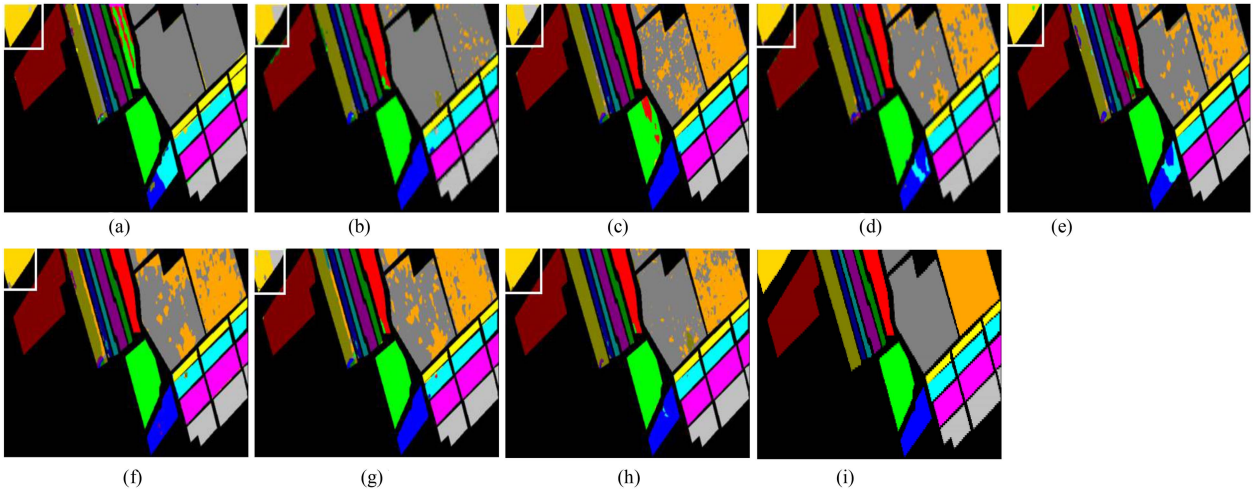


Fig. 13. Performance comparison of different FS methods on Salinas dataset. (a) SPABS. (b) SNMF. (c) BSNet-Conv. (d) LvaHAI. (e) EGCSR_BS_Clustering. (f) IBRA-GSS. (g) NGNMF-E2DSSA. (h) MSFGW. (i) Ground truth.

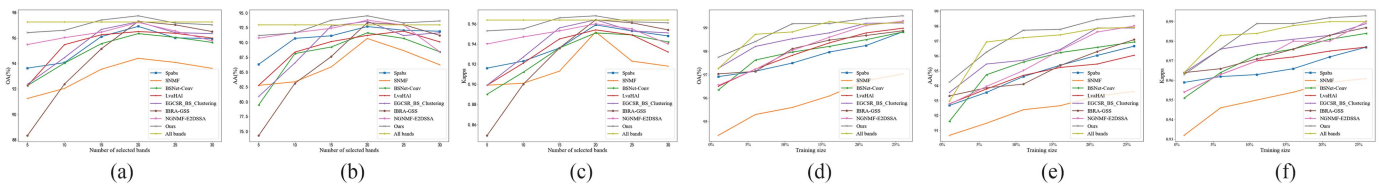


Fig. 14. Performance comparison of different methods with (a)–(c) different band subset sizes and (d)–(e) different training sample sizes on the WHU_Hi_LongKou Dataset.

TABLE VIII
CLASSIFICATION PERFORMANCE OF DIFFERENT METHODS USING 20 BANDS ON THE WHU_HI_LONGKOU DATASET

No.	SpaBS	SNMF	BSNet-Conv	LvaHAI	EGCSR_BS_Clustering	IBRA-GSS	NGNMF-E2DSSA	OURS
1	98.20 ± 1.42	98.76 ± 0.88	98.34 ± 1.04	99.09 ± 0.40	99.22 ± 0.52	99.29 ± 0.55	98.20 ± 2.46	99.47 ± 0.30
2	97.35 ± 2.61	93.36 ± 8.94	90.62 ± 9.24	90.06 ± 13.04	91.54 ± 7.02	89.16 ± 12.44	90.97 ± 10.83	94.42 ± 3.75
3	89.45 ± 6.77	69.52 ± 7.09	87.26 ± 8.61	82.20 ± 9.46	88.61 ± 12.19	90.99 ± 4.80	90.05 ± 16.60	91.03 ± 9.31
4	96.53 ± 0.84	95.71 ± 1.65	95.93 ± 1.83	96.61 ± 1.86	97.31 ± 1.70	97.02 ± 1.95	97.48 ± 1.84	96.76 ± 1.35
5	81.53 ± 17.54	76.43 ± 13.12	85.50 ± 11.29	79.71 ± 13.80	85.22 ± 10.75	87.82 ± 7.96	88.97 ± 4.41	85.52 ± 10.49
6	96.91 ± 7.15	98.88 ± 0.98	98.40 ± 2.06	98.52 ± 1.00	98.58 ± 1.16	98.99 ± 0.83	97.55 ± 1.74	99.02 ± 1.70
7	99.80 ± 0.11	99.73 ± 0.31	99.84 ± 0.24	99.86 ± 0.11	99.85 ± 0.10	99.81 ± 0.16	99.86 ± 0.14	99.89 ± 0.21
8	89.02 ± 5.09	94.05 ± 2.51	83.57 ± 8.50	86.35 ± 11.09	86.60 ± 9.68	84.90 ± 11.99	88.53 ± 6.84	91.17 ± 5.36
9	85.89 ± 7.65	89.57 ± 5.36	85.15 ± 10.05	88.30 ± 6.09	94.11 ± 4.58	93.60 ± 5.06	93.95 ± 3.52	93.13 ± 5.93
OA(%)	96.92 ± 1.45	94.41 ± 1.07	96.35 ± 0.74	96.51 ± 1.31	97.28 ± 0.88	97.04 ± 1.17	97.23 ± 1.25	97.75 ± 0.90
AA(%)	92.75 ± 2.40	90.67 ± 2.20	91.62 ± 1.66	91.19 ± 2.67	93.45 ± 1.97	93.50 ± 1.66	93.95 ± 2.25	94.49 ± 2.37
Kappa	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.96 ± 0.01	0.96 ± 0.01	0.96 ± 0.01	0.96 ± 0.01

The bold entities best classification result of each class.

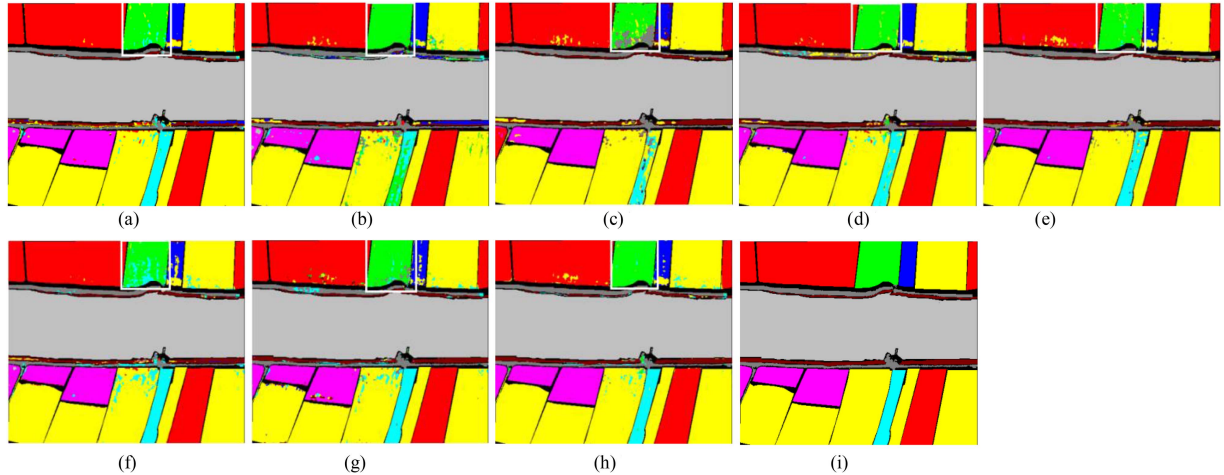


Fig. 15. Performance comparison of different methods on WHU_HI_LongKou dataset. (a) SPABS. (b) SNMF. (c) BSNet-Conv. (d) LvaHAI. (e) EGCSR_BS_Clustering. (f) IBRA-GSS. (g) NGNMF-E2DSSA. (h) MSFGW. (i) Ground truth.

results demonstrate the strong capability of the bands selected by MSFGW for hyperspectral classification. When the number of training samples is greater than 5%, the results of MSFGW in terms of OA(%), AA(%), and Kappa surpass all bands, achieving the best classification accuracy. As the number of training samples increases, the experimental accuracy keeps increasing. In Table VIII, the detailed classification performance of different methods is obtained by selecting the best 20 bands and using 1% of the training samples. MSFGW achieves the best results of OA (97.75%), AA (94.49%), and Kappa (0.96). Compared with all bands OA (97.27%), AA (92.97%), and Kappa (0.96), this signifies that the complete band set contains many noisy bands, which will damage the classification performance. Out of nine classes, MSFGW wins four classes and the suboptimal algorithm wins two classes. Compared with the MSFGW results, the advantage of the suboptimal algorithm is not prominent, and the two classes it wins are similar to the results of MSFGW. The results of other classes are close to MSFGW. It is proved that the band selected by MSFGW also has strong classification ability for the WHU_HI_LongKou dataset with similar spectra between

classes. Fig. 15 shows the classification results of different methods. The proposed MSFGW can generate more uniform and smoother classification maps while preserving edges. From the details in the white boxes, it is clear that MSFGW has excellent performance on irregular shapes and small ground objects.

F. Ablation Experiment

In this section, the effectiveness of 3-D attention is analyzed by comparison with the attention mechanisms of other structures. Specifically, we adopt PAM [35], CAM [47], DAM [36], and no attention mechanism (NAM) in contrast to weight multibranch extracted features. Compare the classification results with the proposed algorithm. For fairness, 1% of the training samples are used for all datasets, while the number of selected bands varies among datasets. The fix number of bands for the Salinas dataset and WHU_HI_LongKou dataset is 20, the fixed number of bands for the Pavia University dataset is 15, and for the Indian Pines dataset is 25. In both cases, the down arrow (\downarrow) indicates that MSFGW performed significantly better, and

TABLE IX
CLASSIFICATION PERFORMANCE GAIN (%) USING DIFFERENT ATTENTION
METHODS ON FOUR DATASETS

Data sets	Evaluation criteria	PAM	CAM	DAM	NAM
INDIAN PINES	OA(%)	↓4.19	↓3.11	↓2.29	↓6.05
	AA(%)	↓6.73	↓2.73	↓3.68	↓5.29
	Kappa	↓0.05	↓0.02	↓0.02	↓0.08
PAVIA UNIVERSITY	OA(%)	↓3.1	↓2.4	↓1.94	↓5.45
	AA(%)	↓2.8	↓2.57	↑0.27	↓6.08
	Kappa	↓0.04	↓0.03	↓0.02	↓0.07
SALINAS	OA(%)	↓2.06	↓1.39	↓0.65	↓3.31
	AA(%)	↓1.92	↓1.19	↑0.76	↓3.55
	Kappa	↓0.03	↓0.02	↓0.01	↓0.04
WHU_Hi_LongKou	OA(%)	↓3.29	↓2.24	↓0.82	↓3.38
	AA(%)	↓7.94	↓1.88	↓1.52	↓11.6
	Kappa	↓0.04	↓0.02	↓0.01	↓0.05

The bold entities indicate result of the comparison algorithm is better than algorithm proposed.

TABLE X
CLASSIFICATION PERFORMANCE GAIN (%) USING SINGLE-BRANCH
ARCHITECTURE ON FOUR DATASETS

Data sets	Evaluation criteria	Branch A	Branch B	Branch C
INDIAN PINES	OA(%)	↓4.14	↓2.97	↓5.05
	AA(%)	↓3.77	↓2.84	↓2.71
	Kappa	↓0.03	↓0.02	↓0.04
PAVIA UNIVERSITY	OA(%)	↓5.16	↓4.58	↓3.51
	AA(%)	↓5.29	↓4.32	↓4.83
	Kappa	↓0.08	↓0.06	↓0.04
SALINAS	OA(%)	↓2.72	↓1.69	↓3.19
	AA(%)	↓2.54	↓0.1	↑1.82
	Kappa	↓0.04	↓0.02	↓0.04
WHU_Hi_LongKou	OA(%)	↓1.07	↓0.66	↓0.44
	AA(%)	↓1.74	↓1.59	↓1.17
	Kappa	↓0.01	↓0.007	↓0.005

The bold entities indicate result of the comparison algorithm is better than algorithm proposed.

the up arrow (↑) indicates that the comparison method performed significantly better than the proposed MSFGW.

Table IX presents the classification results of the algorithm using different mechanisms as the weighting mechanism. Compared with the method without adding the attention mechanism, the classification results of other algorithms are improved to a certain extent, which proves the effectiveness of the attention mechanism in FS. The results of the DAM [36] can achieve better classification performance than PAM [35] or CAM [47], indicating that it has a stronger advantage than channel information in the case of sufficient spatial information and channel information. With 3-D attention, FS is performed in both spatial and spectral dimensions to reduce information loss, so it is better than the attention mechanism of parallel structure. It can be seen from the table that the proposed algorithm achieves the best OA (%) results on all four datasets.

Table X presents the classification results of each branch as a feature extraction module. It can be seen from the table that extracting features using a single-size convolution is not sufficient to characterize the complex case of HSIs. Compared with the multibranch fusion structure of the proposed algorithm, the classification results for single-branch structures are worse.

In addition, it also shows that different scales have various effects on images of different resolutions, so the multibranch structure of the proposed algorithm is more adaptable to different datasets and more robust.

G. Discussion

Aiming at the problem of information redundancy and curse of dimensionality in HSIs, this article proposes an MSFGW for FS of HSIs. From the experimental results, it can be seen that the multiscale feature extraction module composed of multibranch convolution proposed by MSFGW can effectively utilize the advantages of different scales and the complementarities between different scales. For example, in the Indian Pines dataset, compared with the other existing models, MSFGW achieved the best results for the class 9 of a few samples and the class 2 of more samples. In addition, compared with single-channel attention [27], MSFGW can effectively extract and integrate spectral and spatial information in HSIs using 3-D convolution. In the four datasets, in terms of OA, the proposed MSFGW in this article increased by 2%, 4%, 2.5%, and 1.4%, respectively. This shows that the features extracted by MSFGW are more representative and have better classification performance for HSIs.

In the comparative experiments on the classification of the four datasets, the method of this article has achieved outstanding results for the categories of urban buildings, plants, and roads. However, due to the influence of various factors, there are still some categories, such as the fourth category in Table I and the first category in Table III, which are lacking in performance and have a gap with the optimal results. Such as Brocoli_green_weeds_1 and Brocoli_green_weeds_2 in the Salinas dataset, as well as Corn and Corn-min-till in the Indian Pines dataset, etc., are difficult to distinguish in terms of semantic features. The difficulty of distinguishing similar categories in semantic description also interferes with the feature extraction of deep CNN models. Therefore, it also has a certain impact on the classification performance of the overall model.

At the same time, due to the multiscale feature network itself, the network model has efficient multiscale feature perception capabilities. In this complex situation, our method significantly improves the classification ability of datasets with different resolutions and irregular shapes compared with other methods. Hyperspectral datasets usually contain a large number of homogeneous regions, the use of spatial features has a greater effect, and spectral features can characterize the spectral characteristics of ground objects. Therefore, using a model that jointly extracts spatial and spectral features can yield more robust hyperspectral visual features. In the next step of research, more attention will be paid to improve the classification performance of the model.

V. CONCLUSION

This article presents a new FS method MSFGW for hyperspectral remote sensing images to explore the spatial scale problem of various ground objects. Considering the interaction between channels and spectral information in different bands from a global perspective, the proposed MSFGW adopts the multiscale dilated convolution module to extract the spectral

features of objects at different spatial scales. Then, the complementary information from multiscale spectral features combines into a consistent map. Finally, the bands are selected by the contribution to the band reconstruction task. The experimental results on four public hyperspectral datasets validate the better performance of the proposed MSFGW method than the other state-of-the-art comparison methods, and indicate the effectiveness of the proposed MSFGW in FS for HSIs.

REFERENCES

- [1] L. Liu, N. C. Coops, N. W. Aven, and Y. Pang, "Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data," *Remote Sens. Environ.*, vol. 200, pp. 170–182, Oct. 2017, doi: [10.1016/j.rse.2017.08.010](https://doi.org/10.1016/j.rse.2017.08.010).
- [2] Y. Tan, L. Lu, L. Bruzzone, R. Guan, Z. Chang, and C. Yang, "Hyperspectral band selection for lithologic discrimination and geological mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 471–486, Jan. 2020, doi: [10.1109/JSTARS.2020.2964000](https://doi.org/10.1109/JSTARS.2020.2964000).
- [3] M. Shimoni, R. Haelterman, and C. Perneel, "Hyperspectral imaging for military and security applications: Combining myriad processing and sensing techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 101–117, Jun. 2019, doi: [10.1109/MGRS.2019.2902525](https://doi.org/10.1109/MGRS.2019.2902525).
- [4] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, Jun. 2019, doi: [10.1109/MGRS.2019.2911100](https://doi.org/10.1109/MGRS.2019.2911100).
- [5] J. P. Rivera-Caicedo, J. Verrelst, J. Muñoz-Marí, G. Camps-Valls, and J. Moreno, "Hyperspectral dimensionality reduction for biophysical variable statistical retrieval," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 88–101, Oct. 2017, doi: [10.1016/j.isprsjprs.2017.08.012](https://doi.org/10.1016/j.isprsjprs.2017.08.012).
- [6] F. Luo, L. Zhang, B. Du, and L. Zhang, "Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5336–5353, Aug. 2020, doi: [10.1109/TGRS.2020.2963848](https://doi.org/10.1109/TGRS.2020.2963848).
- [7] X. Jiang, X. Song, Y. Zhang, J. Jiang, J. Gao, and Z. Cai, "Laplacian regularized spatial-aware collaborative graph for discriminant analysis of hyperspectral imagery," *Remote Sens.*, vol. 11, no. 1, 2019, Art. no. 29, doi: [10.3390/rs11010029](https://doi.org/10.3390/rs11010029).
- [8] W. Sun, L. Zhang, B. Du, W. Li, and Y. M. Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2784–2797, Jun. 2015, doi: [10.1109/JSTARS.2015.2417156](https://doi.org/10.1109/JSTARS.2015.2417156).
- [9] Y. Yuan, J. Lin, and Q. Wang, "Dual-clustering-based hyperspectral band selection by contextual analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1431–1445, Mar. 2016, doi: [10.1109/TGRS.2015.2480866](https://doi.org/10.1109/TGRS.2015.2480866).
- [10] W. Sun, L. Tian, Y. Xu, D. Zhang, and Q. Du, "Fast and robust self-representation method for hyperspectral band selection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 5087–5098, Nov. 2017, doi: [10.1109/JSTARS.2017.2737400](https://doi.org/10.1109/JSTARS.2017.2737400).
- [11] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999, doi: [10.1109/36.803411](https://doi.org/10.1109/36.803411).
- [12] K. Sun, X. Geng, and L. Ji, "A new sparsity-based band selection method for target detection of hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 329–333, Feb. 2015, doi: [10.1109/LGRS.2014.2337957](https://doi.org/10.1109/LGRS.2014.2337957).
- [13] Y. Yuan, G. Zhu, and Q. Wang, "Hyperspectral band selection by multitask sparsity pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 631–644, Feb. 2015, doi: [10.1109/TGRS.2014.2326655](https://doi.org/10.1109/TGRS.2014.2326655).
- [14] W. Zhang, X. Li, Y. Dou, and L. Zhao, "A geometry-based band selection approach for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4318–4333, Aug. 2018, doi: [10.1109/TGRS.2018.2811046](https://doi.org/10.1109/TGRS.2018.2811046).
- [15] M. Zhang, "Hyperspectral feature selection and extraction based on evolutionary optimization learning," Xidian Univ., Xi'an, China, 2018.
- [16] L. Yu, Y. Han, and L. Mu, "Improved quantum evolutionary particle swarm optimization for band selection of hyperspectral image," *Remote Sens. Lett.*, vol. 11, no. 9, pp. 866–875, Sep. 2020, doi: [10.1080/2150704X.2020.1782501](https://doi.org/10.1080/2150704X.2020.1782501).
- [17] H. Su, Q. Du, G. Chen, and P. Du, "Optimized hyperspectral band selection using particle swarm optimization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2659–2670, Jun. 2014, doi: [10.1109/JSTARS.2014.2312539](https://doi.org/10.1109/JSTARS.2014.2312539).
- [18] D. Xiaohui, H. Li, L. Yong, Y. Ji, and Z. Shuqing, "Comparison of swarm intelligence algorithms for optimized band selection of hyperspectral remote sensing image," *Open Geosci.*, vol. 12, no. 1, pp. 425–442, Jul. 2020, doi: [10.1515/geo-2020-0155](https://doi.org/10.1515/geo-2020-0155).
- [19] W. Sun, W. Li, J. Li, and Y. M. Lai, "Band selection using sparse nonnegative matrix factorization with the thresholded Earth's mover distance for hyperspectral imagery classification," *Earth Sci. Inform.*, vol. 8, pp. 907–918, Jan. 2015.
- [20] Q. Zhu, Y. Wang, F. Wang, M. Song, and C.-I. Chang, "Hyperspectral band selection based on improved affinity propagation," in *Proc. 11th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens.*, 2021, pp. 1–4, doi: [10.1109/WHISPERS52202.2021.9484004](https://doi.org/10.1109/WHISPERS52202.2021.9484004).
- [21] H. Su, Y. Sheng, P. Du, and K. Liu, "Adaptive affinity propagation with spectral angle mapper for semi-supervised hyperspectral band selection," *Appl. Opt.*, vol. 51, no. 14, pp. 2656–2663, May 2012, doi: [10.1364/AO.51.002656](https://doi.org/10.1364/AO.51.002656).
- [22] D. Varade, A. K. Maurya, and O. Dikshit, "Unsupervised band selection of hyperspectral data based on mutual information derived from weighted cluster entropy for snow classification," *Geocarto Int.*, vol. 36, no. 15, pp. 1709–1731, Sep. 2021, doi: [10.1080/10106049.2019.1665717](https://doi.org/10.1080/10106049.2019.1665717).
- [23] K. R. Kalidindi, P. S. Varma G., and R. Davuluri, "Featured clustering and ranking-based bad cluster removal for hyperspectral band selection and classification using ensemble of binary SVM classifiers," *Int. J. Inf. Technol. Project Manage.*, vol. 12, no. 4, pp. 61–78, 2021.
- [24] J. He, C. Zhang, L. Zhou, and Y. He, "Simultaneous determination of five micro-components in chrysanthemum morifolium (Hangbaiju) using near-infrared hyperspectral imaging coupled with deep learning with wavelength selection," *Infrared Phys. Technol.*, vol. 116, Aug. 2021, Art. no. 103802, doi: [10.1016/j.infrared.2021.103802](https://doi.org/10.1016/j.infrared.2021.103802).
- [25] J. Feng et al., "Convolutional neural network based on bandwise-independent convolution and hard thresholding for hyperspectral band selection," *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4414–4428, Sep. 2021, doi: [10.1109/TCYB.2020.3000725](https://doi.org/10.1109/TCYB.2020.3000725).
- [26] Y. Zhan et al., "A new hyperspectral band selection approach based on convolutional neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 3660–3663, doi: [10.1109/IGARSS.2017.8127792](https://doi.org/10.1109/IGARSS.2017.8127792).
- [27] Y. Cai, X. Liu, and Z. Cai, "BS-Nets: An end-to-end framework for band selection of hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1969–1984, Mar. 2020, doi: [10.1109/TGRS.2019.2951433](https://doi.org/10.1109/TGRS.2019.2951433).
- [28] S. K. Roy, S. Das, T. Song, and B. Chanda, "DARecNet-BS: Unsupervised dual-attention reconstruction network for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 12, pp. 2152–2156, Dec. 2021, doi: [10.1109/LGRS.2020.3013235](https://doi.org/10.1109/LGRS.2020.3013235).
- [29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017, doi: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [30] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9, doi: [10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594).
- [31] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," Jun. 7, 2016, Accessed: Sep. 14, 2022. [Online]. Available: <http://arxiv.org/abs/1412.7062>
- [32] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: [10.1109/TPAMI.2017.2699184](https://doi.org/10.1109/TPAMI.2017.2699184).
- [33] R. P. K. Poudel, S. Liwicki, and R. Cipolla, "Fast-SCNN: Fast semantic segmentation network," Feb. 12, 2019, Accessed: Sep. 14, 2022. [Online]. Available: <http://arxiv.org/abs/1902.04502>
- [34] H. Ren, M. Pan, Y. Li, X. Zhou, and J. Luo, "ST-SiameseNet: Spatio-temporal Siamese networks for human mobility signature identification," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 1306–1315, doi: [10.1145/3394486.3403183](https://doi.org/10.1145/3394486.3403183).
- [35] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19, doi: [10.1007/978-3-030-01234-2_1](https://doi.org/10.1007/978-3-030-01234-2_1).

- [36] J. Fu et al., "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3141–3149.
- [37] M. Gong, M. Zhang, and Y. Yuan, "Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 544–557, Jan. 2016, doi: [10.1109/TGRS.2015.2461653](https://doi.org/10.1109/TGRS.2015.2461653).
- [38] P. Groves and P. Bajcsy, "Methodology for hyperspectral band and classification model selection," in *Proc. IEEE Workshop Adv. Techn. Anal. Remotely Sensed Data*, 2003, pp. 120–128, doi: [10.1109/WARSD.2003.1295183](https://doi.org/10.1109/WARSD.2003.1295183).
- [39] X. Wei, L. Cai, B. Liao, and T. Lu, "Local-view-assisted discriminative band selection with hypergraph autolearning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 2042–2055, Mar. 2020, doi: [10.1109/TGRS.2019.2952383](https://doi.org/10.1109/TGRS.2019.2952383).
- [40] Y. Cai, Z. Zhang, X. Liu, and Z. Cai, "Efficient graph convolutional self-representation for band selection of hyperspectral image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4869–4880, Aug. 2020.
- [41] G. Morales, J. W. Sheppard, R. D. Logan, and J. A. Shaw, "Hyperspectral dimensionality reduction based on inter-band redundancy analysis and greedy spectral selection," *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3649.
- [42] H. Fu et al., "A novel band selection and spatial noise reduction method for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5535713.
- [43] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.
- [44] W. Wang, S. Dou, Z. Jiang, and L. Sun, "A fast dense spectral-spatial convolution network framework for hyperspectral images classification," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1068.
- [45] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004, doi: [10.1109/TGRS.2003.819189](https://doi.org/10.1109/TGRS.2003.819189).
- [46] C. Yu, L.-C. Lee, C.-I. Chang, B. Xue, M. Song, and J. Chen, "Band-specified virtual dimensionality for band selection: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2822–2832, May 2018, doi: [10.1109/TGRS.2017.2784372](https://doi.org/10.1109/TGRS.2017.2784372).
- [47] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141, doi: [10.1109/CVPR.2018.00745](https://doi.org/10.1109/CVPR.2018.00745).



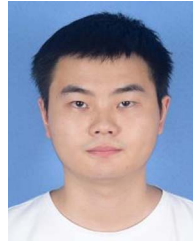
Jinjin Wang received the B.S. degree in communication engineering and the M.S. degree in information and communication engineering from the Zhongyuan University of Technology, Zhengzhou, China, in 2017 and 2020, respectively. She is currently working toward the Ph.D. degree in optical engineering with the Nanjing University of Aeronautics and Astronautics, Nanjing, China.

Her research interests include image processing, hyperspectral dimensionality reduction, and deep learning.



Jiahang Liu (Member, IEEE) received the B.S. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2000, the M.S. degree in tectonics from the Institute of Geology, China Earthquake Administration, Beijing, China, in 2003, and the Ph.D. degree in pattern recognition and intelligent system from Shanghai Jiao Tong University, Shanghai, China, in 2011.

He has been a Full Professor with the Nanjing University of Aeronautics and Astronautics, Nanjing, China. His research interests include remote sensing, image processing, computer vision, and digital twin.



Jian Cui received the B.S. degree in communication engineering and the M.S. degree in information and communication engineering from the Zhongyuan University of Technology, Zhengzhou, China, in 2017 and 2020, respectively. He is currently working toward the Ph.D. degree in optical engineering with the Nanjing University of Aeronautics and Astronautics, Nanjing, China.

His research interests include instance segmentation, object detection, edge detection, etc.



Ji Luan received the B.S. degree in information engineering in 2021 from the School of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing, China, where she is currently working toward the master's degree in electronic information engineering.

Her research interests include hyperspectral image processing.



Yangyu Fu is master degree in optical engineering majoring in optoelectronic information science and engineering with the School of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing, China.

She has published two papers entitled "Teaching Simulation Based on Time Coherence virtual lab" and "Research on Forest Management System Based on Cellular Automata." Her current research focuses on data processing in interferometric imaging spectrometers.