IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 February 2023, accepted 2 March 2023, date of publication 6 March 2023, date of current version 13 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3253627

RESEARCH ARTICLE

Spatial Proximity Feature Selection With Residual Spatial–Spectral Attention Network for Hyperspectral Image Classification

XINSHENG ZHANG AND ZHAOHUI WANG, (Member, IEEE)
School of Computer Science and Technology, Hainan University, Haikou 570228, China

Corresponding author: Zhaohui Wang (william_hig@163.com)

This work was supported in part by the Framework of the Norwegian Research Council INTPARTInternational Partnerships for Excellent Education, Research and Innovation (INTPART) Project through the International Network for Image-Based Diagnosis (INID) under Grant 309857; and in part by the Hainan Key Research and Development Plan for Scientific and Technological Collaboration Projects through the Research on Medical Imaging Aided Diagnosis of Infant Brain Development Diseases under Grant GHYF2022015.

ABSTRACT Over the past few years, deep learning has been introduced to tackle hyperspectral image (HSI) classification and demonstrated good performance. In particular, the convolutional neural network (CNN) based methods have progressed. However, due to the high dimensionality of HSI and equal treatment of all bands, the performances of CNN based methods are hampered. The labels of land-covers often differ between edge and the center pixels in pixel-centered spatial information. These edge pixels may weaken the discrimination of spatial features and reduce classification accuracy. Motivated by the attention mechanism of the human visual system, the spatial proximity feature selection with residual spatial–spectral attention network is proposed in this article. It contains a residual spatial attention module, a residual spectral attention module, and a spatial proximity feature selection module. The residual spatial attention module aims to select the crucial spatial information, which assigns weights to different features by measuring the similarity between the surrounding elements and their central ones. The residual spectral attention module is designed for spectral bands which are selected from raw input data by emphasizing the valuable bands and suppressing the valueless. According to the spatial distribution of features, the spatial proximity feature selection module is used to filter features effectively. Experiments on three public data sets demonstrate that the proposed network outperforms the state-of-the-art methods in comparison.

INDEX TERMS Residual spatial attention module, residual spectral attention module, spatial proximity feature selection, hyperspectral image classification.

I. INTRODUCTION

Unlike traditional RGB images, hyperspectral images (HSIs) are composed of 1-D spectral signatures and 2-D spatial information. The spectral signatures encompass hundreds of continuous spectral bands. The spatial information contains the detailed spatial distribution of objects. Based on such abundant spectral bands, HSI has a wide range of applications, such as urban development [1], surveillance [2], environment management [3] and etc. Hyperspectral image classification (HSIC) primarily aims to classify each pixel as one of a given set of land-cover classes, such as land, forest, rivers, and etc.

Numerous methods have been applied during the past decades to obtain better classification accuracy. Early strategies mainly focused on the spectral signatures. Representative approaches include support vector machine (SVM) [4], extreme learning machine [5], and multinomial logistic regression [6]. These methods mainly use the spectrum of a pixel to determine its class. However, the accuracy is affected by the spectral variability which is brought by many factors such as incident illumination, atmospheric effects, unwanted

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo.

shade and shadow, natural spectrum variation, and instrument noises [7], [8], [9], [10], [11]. Therefore, the classification results of the spectral-based methods are unsatisfactory.

Many methods introduce spatial information into the classification process. Some works extract the spatial features via morphological operators [12], Gabor filters [13], and hypergraph structure [14] or Markov random fields [15]. These spatial features are combined with spectral features for classification. Others directly extract the joint spectral–spatial features using 3-D discrete wavelets [16], 3-D scattering wavelets [17], 3-D Gabor filters [18], and so on. The spectral–spatial methods attempt to utilize the neighboring pixels to complement the spectral signatures. Based on Tobler's First Law of Geography, the adjacent pixels are assumed to share the same land-cover label [19]. The spectral–spatial methods can extract spectral–spatial features from the adjacent pixels and have shown outstanding classification performances. Meanwhile, the influence of spectral variability on classification results is reduced effectively. Nevertheless, these methods extract features in a shallow manner, which is difficult to gain the performance substantially.

Recently, deep learning algorithms have made up for the defects of traditional methods in feature extraction, which has been achieving progress in computer vision tasks, including object detection [20], semantic segmentation [21], and image classification [22]. Furthermore, various deep learning models have been investigated, such as multilayer perceptron (MLP) [23], stacked autoencoder (SAE) [24], and convolutional neural network (CNN) [25]. Feng et al. [26] proposed a new adaptive spatial regularization edge SAE which used a super-pixel segmentation method to segment the image into multiple homogeneous regions. Shi and Pun [24] employed SAE to fully use the spatial features between each superpixel through local and nonlocal similarity measures. In [27], CNN was involved in deep spectral-spatial feature extraction and classification. In [28], considering the abundance of unlabeled samples, a novel self-supervised divide-and-conquer GAN is proposed.

Among deep learning algorithms, CNN shows its effects in feature extraction because the characteristics of local connections and shared weights reduce the number of network parameters [29]. According to the input of networks, existing CNN models can be grouped into two classes: spectral CNNs [30], [31], [32], [33] and spectral–spatial CNNs [34], [35], [36]. Spectral CNNs extract spectral features for each pixel only. For example, Hu et al. [30] designed a 1-D CNN model to extract features from the spectral information of each pixel. Because of the small numbers of training pixels, Hu's model is not very deep, which limits the ability of 1-D CNNs to represent features. A novel pixel-pair method [31] was proposed to regard the pixel learning problem as the pixel-pair counterpart, by which the number of training pixels is significantly increased. Wu and Saurabh [32] and Hao and Prasad [33] proposed the combination of 1-D CNN and RNN, which fed the spectral features learned by a 1-D CNN into an RNN to enhance the discriminative ability of the extracted features. However, the use of spectral information alone is challenging to recognize the classes of land covers. The spectral CNNs are lack of taking spatial features into account. Spectral–spatial features that are properly extracted can be more discriminative than the spectral features alone for HSI processing tasks.

Unlike spectral CNNs, spectral–spatial CNNs extract spectral and spatial features simultaneously. Both spectral signatures and spatial information contribute clues for HSIC. In some recent researches, 3-D CNNs have been used to extract spatial-spectral features [37], [38], [39], [40]. Because the training time of the spectral-spatial residual network (SSRN) [41] was too long, Ahmad et al. [37] designed a fast dense spectral-spatial convolutional network. In [42], a residual conv–deconv network was used to extract spectral-spatial features from unlabeled HSI cubes. For the problems of limited training samples and unbalanced classes, Chen et al. [43] combined the virtual sample enhancement technology with a CNN to effectively extract the spectral and spatial information. However, the number of parameters in 3-D CNN increased considerably, which will lead to the overfitting and significantly increase the computational consumption [44]. Many works have attempted to design the dual-branch networks to alleviate this problem. One branch focuses on spectral feature extraction, and the other focuses on spatial feature extraction. For examples, Zhao and Du et al. [45] combined a local discrimination embedding algorithm with a CNN. The local discrimination was used to extract the spectral information from images and the CNN was used to extract the spatial information continuously. Yue et al. [46] combined a deep convolutional neural network and logical regression by generating the spectral and spatial feature maps. Zhang et al. [47] learned the spatial–spectral context-sensitivity using CNNs in different regions and enhanced the recognition ability of network by combining various distinguishable appearance factors. Mei et al. [48] proposed a five-layer neural network to integrate contextual information and spectral information. Xu et al. [49] proposed a spectral-spatial unified network which contains a CNN and a long short-term memory model based on band grouping. The parallel two-branch frameworks were proposed in references [50] and [51], which effectively extract the spatial-spectral features and dramatically reduces the number of parameters. Roy et al. [52] proposed a hybrid spectral CNN which effectively reduces the complexity of the model. Nevertheless, the above-mentioned methods utilize all possible information in data regardless of whether the information is helpful or not. Some useless information may be considered, which will result in a waste of computing resources.

Recently, the attention mechanism has been popularly employed in language modeling [53], [54] and computer vision tasks [55], [56], [57]. Its success mainly depends on the reasonable assumption that human vision only focuses on particular parts of the visual space when and where needed

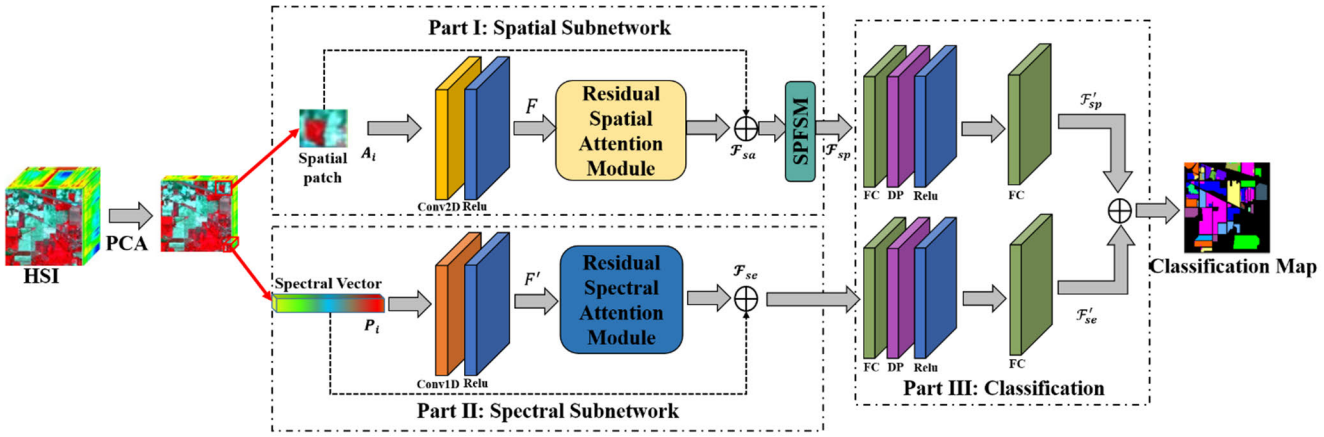

FIGURE 1. Architecture of the proposed SPFS-RSSAN. The Indian Pines database is utilized as an example to illustrate the architecture.

[58]. Common attention mechanisms, such as SENets [56] and bottleneck attention module [59], use the global max and average pooling to summarize the information carried by every feature map. Yu et al. [60] developed a feedback attention module that assembles the spectral-space features in a compact connection. To reduce the loss in band selection, Feng et al. [61] took advantage of the band-independent convolution and hard thresholding to integrate band selection, feature extraction and classification. In [62], the SpectralFormer was proposed to mine and represent the sequence properties of spectral features. Roy et al. [63] introduced a multi-modal fusion transformer that achieves superior performance by using multi-modal information as external classification markers.

In this paper, a spatial proximity feature selection with residual spatial–spectral attention network (SPFS-RSSAN) is proposed. Our goal is to enhance the representation capacity using attention mechanism, by which the SPFS-RSSAN will focus on more discriminative spectral bands and spatial positions while suppressing the unnecessary. Three core modules are designed: the residual spatial attention module (RSaAM), the residual spectral attention module (RSeAM), and the spatial proximity feature selection module (SPFSM). The RSaAM is used to distinguish the importance of each surrounding pixel to the central pixel classification by which the contribution of pixels with the same label as the central pixel to classification is enhanced. RSeAM is used to assign weight to each spectral feature, which can be interpreted as a band selector. The modified Minkowski distance is applied to avoid overfitting and find pixels with the same label as the central target. The effectiveness of SPFS-RSSAN has been validated on three public data sets. Experimental studies demonstrate that the proposal can achieve better classification results.

The main contributions of this work are as follows.

1) An RSaAM is designed to exploit the spatial feature correlation between the center pixel and its surroundings, which improves the spatial feature representation related to the center pixel specifically. In this module, the 2-D convolution is used to extract the spatial features. The sigmoid activation function is then in charge of generating the proper spatial weights.

2) An RSeAM is designed to generate a spectral weight vector that reflects the importance of different spectral bands. This attention module exploits the pooling layers and sigmoid function for producing a series of recalibrated spectral information, which can effectively improve the classification results.

3) An SPFSM is proposed to reduce the impact of useless information. It consists of a proximity selection layer and a tanh-derivative activation function. Tanh-derivative function conforms the operation logic of the SPFSM to improve the classification accuracy.

The rest of the paper is organized as follows. Section II describes the proposed method in detail. The experimental results are given in Section III. Finally, the paper is concluded in Section IV.

II. METHODOLOGY

A. OVERVIEW OF THE PROPOSAL

As shown in Fig. 1, the network consists of the spatial subnetwork, the spectral subnetwork, and the classification part. For the spectral subnetwork, considering that each pixel can be represented as a continuous spectral curve having a rich spectrum character. The RSeAM is used to focus on the inter-band relationship of features. The spatial features are regarded as the complements in spatial dimensions. The spatial subnetwork improves the representation of interests and focuses on the inter-spatial relationships of features by exploiting the RSaAM and the SPFSM. Then, the features extracted by the spectral and spatial subnetworks are fed to the fully connected layers to learn the high-level joint spatial-spectral features and acquire a prediction by the softmax function.

Let the HSI cube be denoted by $I \in \mathbb{R}^{H\times W\times B}$, where I, H, W, and B present the original input, the length, the

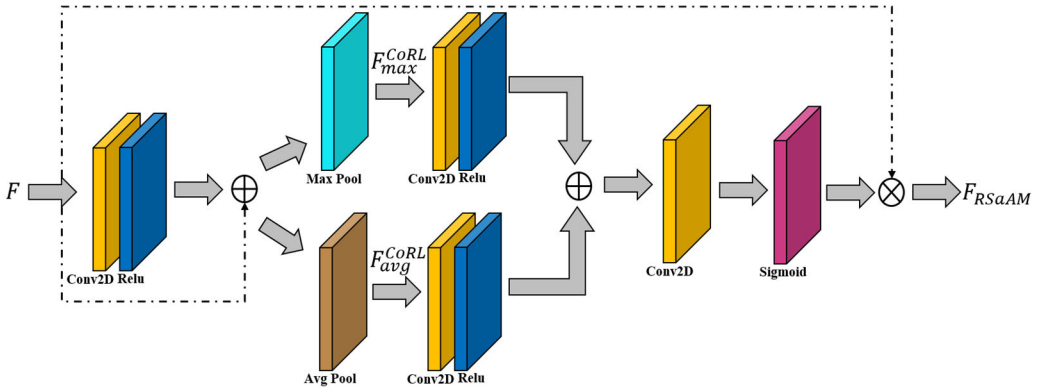

FIGURE 2. Residual spatial attention module. It utilizes both max-pooling outputs and average-pooling outputs pooled along the spectral axis and stacked horizontally as the input of a convolution layer.

width, and the number of bands, respectively. The label of every HSI pixel $\mathrm{P} = \{p_1, p_2, \cdots, p_N\}$ forms a one-hot vector $\mathrm{Y} = \{y_1, y_2, \cdots, y_N \in \mathbb{R}^{1\times1\times C}\}$, where N is the number of labeled pixels and C is the number of categories. To reduce the number of spectral bands and maintain the same spatial dimension, the principal component analysis (PCA) was applied to compress the original input. The reduced number of bands is denoted as b. Only spectral bands are reduced to preserve spatial information, which is very important for recognizing land-cover. After dimension reduction, for a pixel $p_i (i = \{1, 2, \cdots, N\})$, a spatial patch $A_i \in \mathbb{R}^{H\times W\times b}$ centered at p_i is taken as the input of the spatial subnetwork. It passes through a 2-D convolution, a Rectified Linear Unit (ReLU) layer, the RSaAM, and an element-wise addition to produce the spatial features. The spatial subnetwork can be formulated as follows:

$$\mathcal{F} = Co_{2D}RL\,(A_i) \tag{1-a}$$
$$\mathcal{F}_{sa} = A_i \oplus RSaAM\,(F) \tag{1-b}$$

where F_{sa} denotes the features extracted by the spatial subnetwork. "$\oplus$" denotes the element-wise addition. The $RSaAM\,(\cdot)$ represents the feature processed by the residual spatial attention module. $Co_{2D}RL$ represents the 2-D convolutional and ReLU layers. The classification accuracy decreasing with the increasing of convolution layers stems from the fact that the representation capacity of CNNs is too high compared with the relatively small number of training samples under the same regularization settings. However, this decreasing-accuracy issue can be alleviated by adding the shortcut connections between every other layer to build the residual blocks. The element-wise addition is the core step in the shortcut.

The spectrum of p_i is taken as the input of the spectral subnetwork. The 1-D convolutional and ReLU layers are used to extract spectral features. The spectral subnetwork can be formulated as follows:

$$F' = Co_{1D}RL\,(p_i) \tag{2-a}$$
$$\mathcal{F}_{se} = p_i \oplus RSeAM\left(F'\right) \tag{2-b}$$

where F_{se} denotes features extracted by the spectral subnetwork. $Co_{1D}RL$ represents the 1-D convolutional and ReLU layers. $RSeAM\,(\cdot)$ represents the feature processed by the residual spectral attention module. The spatial and spectral features extracted are fed into the classification part to determine the category.

After the network is built, its parameters are initialized by the normalization and regularized with the L2 weight decay penalty. The network is trained in an end-to-end manner. During the training process, the backpropagation algorithm is used to update the parameters of the network.

B. RESIDUAL SPATIAL ATTENTION MODULE

The RSaAM aims to enhance the spatial information from the neighboring pixels with the same label as the center pixel and suppress those with a different class label. Thus, the ideal output of the RSaAM should be a matrix with the same height and weight as the input patch F, where the value of the pixel in this location with the same label as the center is equal to 1. Otherwise, it is 0. Fig. 2 shows the operations of the proposed RSaAM. Given a feature map $\mathrm{F} \in \mathbb{R}^{h\times w\times b}$, where $h \times w$ denotes the spatial size and b denotes the number of channels. Two pooling operations are used to aggregate the channel information of a feature map, thereby generating two maps: F_{avg} and F_{max}, which can be expressed as follows:

$$F_{avg} = Avg(F) = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W} F_{i,j}}{H \times W} \tag{3-a}$$
$$F_{max} = Max(F) = \max\,(F) \tag{3-b}$$

where $F_{i,j}$ is the position (i, j) of the input F. The $\max(F)$ denotes the maximum value of F. The 2-D convolutional and ReLU layers are used to limit the model complexity and aid for generalization.

Firstly, the input features are extracted by:

$$F^{CoRl} = CoRl\,(F) \tag{4-a}$$
$$f = F \oplus F^{CoRl} \tag{4-b}$$

where $CoRl(\cdot)$ denotes that the features are processed with the convolutional and ReLU layers, respectively. f is processed

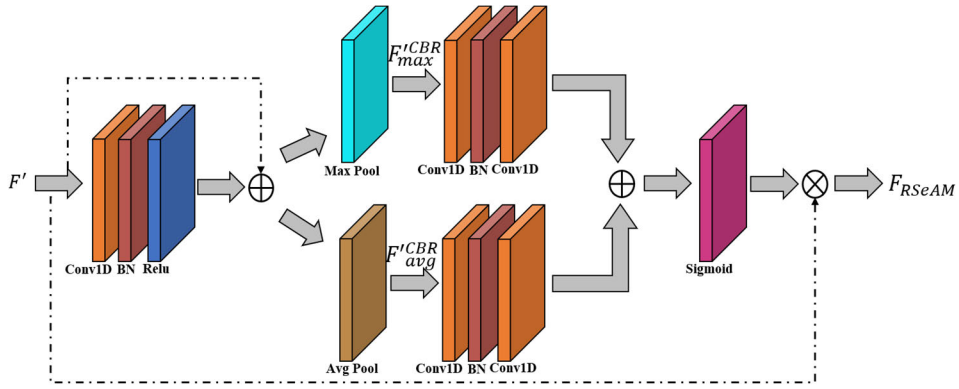

FIGURE 3. Residual spectral attention module. This module utilizes both max-pooling outputs and average-pooling outputs with a two-layer weight-shared bottleneck network.

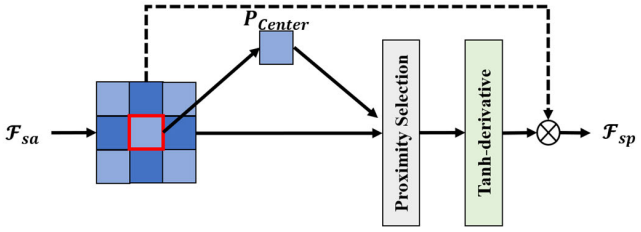

FIGURE 4. SPFSM. It contains a proximity selection layer and a tanh-derivative active function.

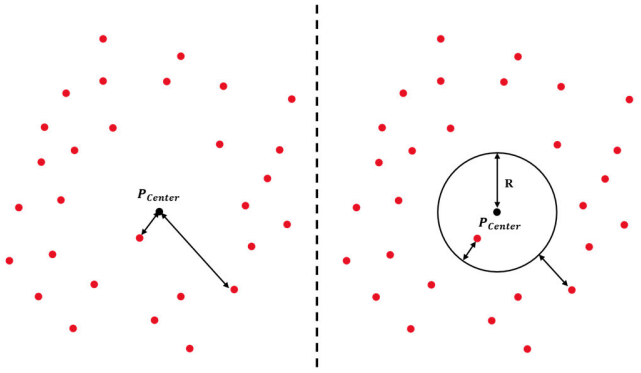

FIGURE 5. Comparison of the original distance (left) and the distance with slack variable (right). Where R is a slack variable.

with the global max pooling and global average pooling layers, respectively. The results are F_{max} and F_{avg}:

$$F_{max} = Max\,(f) \tag{5-a}$$
$$F_{avg} = Avg\,(f) \tag{5-b}$$

Then, the convolutional layer and ReLU activation function are used for feature extraction and activation, which are represented as F_{max}^{CoRL} and F_{avg}^{CoRL}, respectively:

$$F_{max}^{CoRL} = ReLU\,(Con\,(F_{max})) \tag{6-a}$$
$$F_{avg}^{CoRL} = ReLu(Con(F_{avg})) \tag{6-b}$$

The two outputs are concatenated horizontally as the input of a new convolutional layer followed by a

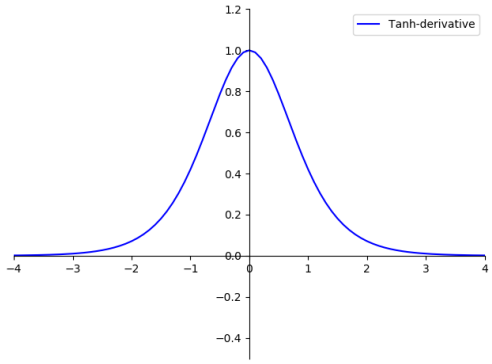

FIGURE 6. Tanh-derivative activation function.

sigmoid activation function:

$$F_{RSaAM} = F \otimes Sig(CoRl\left(F_{max}^{CoRL}\right) \oplus CoRl(F_{avg}^{CoRL})) \tag{7}$$

where $Sig(\cdot)$ is the sigmoid activation function. "$\otimes$" denotes the element-wise multiplication.

C. RESIDUAL SPECTRAL ATTENTION MODULE

The RSeAM aims to increase the weight of those spectral information that are helpful to represent features. The input spectral feature vectors $F' \in \mathbb{R}^{l \times c}$, where the length of the spectral feature vectors and the number of channels are represented by l and c, respectively. The process of the RSeAM is shown in Fig. 8:

$$F^{'CBR} = CBR\left(F'\right) \tag{8-a}$$
$$f' = F' \oplus F^{'CBR} \tag{8-b}$$
$$F_{max}^{'CBR} = Max\left(f'\right) \tag{8-c}$$
$$F_{avg}^{'CBR} = Avg\left(f'\right) \tag{8-d}$$
$$F_{RSeAM} = F' \otimes Sig(CRC\left(F_{max}^{'CBR}\right) \oplus CRC(F_{avg}^{'CBR})) \tag{8-e}$$

where $CBR(\cdot)$ denotes that the features are processed with convolution, Batch Normalization (BN), and ReLU activation function respectively. $CRC(\cdot)$ denotes that convolution,

ReLU activation function, and convolution are combined in sequence.

D. SPATIAL PROXIMITY FEATURE SELECTION MODULE

The module is designed to capture the spatial areas relevant to the center pixel. The architecture is shown in Fig. 4. It is a lightweight module which consists of two parts: the proximity selection and the tanh-derivative activation function.

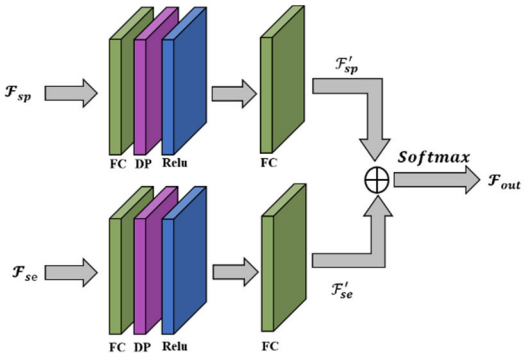

FIGURE 7. Classification part.

The Minkowski distance is used to analyze the possibility that the central pixel and its adjacent pixels belong to the same category. As shown in Fig. 5, considering the influence of environmental factors such as atmospheric transmittance and steam, a Minkowski distance with slack variable is proposed. When the distance between a pixel and the center pixel is smaller, they are more likely to belong to the same category. The extent of the contribution of a pixel to classification is inversely proportional to the distance between itself and the central pixel. The tanh-derivative activation function is used to allocate a proper weight for each pixel. The functional graph is shown in Fig. 6, and the formula is as follows:

$$TanDe\,(x) = 1 - \left(\frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}\right)^{2} \tag{9}$$

Specifically, the process of SPFSM is as follows:

$$S = MD\left(P_{center}, P_{neighbor}\right) - R \tag{10-a}$$

$$F_{sp} = TanDe\,(S) \otimes F_{sa} \tag{10-b}$$

where $MD(\cdot)$ denotes the Minkowski distance, and R is slack variable.

Table 1 displays the parameters of the layers of the proposal.

E. CLASSIFICATION PART

As shown in Fig. 7, this part receives and adds the inputs from the two subnetworks. The features are fed to the FC, dropout, and ReLU layers. The number of neurons in the last FC layer is equal to the number of classes, and the value of each neuron can be regarded as a class-specific response. The process can be expressed as follows:

$$\mathcal{F}'_{sp} = FC(FDR(F_{sp})) \tag{11-a}$$

TABLE 1. Detailed architecture of the proposed SPFS-RSSAN model.

Module		Layer Name	Kernel Size	Filters	Connected to	Configuration
		Input	-	-	-	-
Spatial Subnetwork		Conv2D_1	3×3	32	Input	strides 1, ReLU
	RSaM	Conv2D_2	3×3	64	Conv2D_1	strides 1, ReLU
		Add_1	-	-	Conv2D_1, Conv2D_2	-
		MP_1	2×2	-	Add_1	strides 2
		Conv2D_3	3×3	128	MP_1	strides 1, ReLU
		AP_1	2×2	-	Add_1	strides 2
		Conv2D_4	3×3	128	AP_1	strides 1, ReLU
		Add_2	-	-	Conv2D_3, Conv2D_4	-
		Conv2D_5	3×3	64	Add_2	strides 1
		Sigmoid_1	-	-	Conv2D_5	-
		Multiply_1	-	-	Conv2D_1, Sigmoid_1	-
		Add_3	-	-	Input, Multiply_1	-
	SPFSM	Proximity Selection	-	-	Add_3	-
		Tanh-derivate	-	-	Proximity Selection	-
		Multiply_2	-	-	Add_3, Tanh-derivate	-
Spectral Subnetwork	RSeAM	Conv1D_1	2	32	Input	strides 1, ReLU
		Conv1D_2	2	64	Conv1D_1	strides 1, ReLU, BN
		Add_3	-	-	Conv1D_1, Conv1D_2	-
		MP_2	2×2	-	Add_3	strides 2
		Conv1D_3	2	128	MP_2	strides 1, BN
		Conv1D_4	2	64	Conv1D_3	strides 1
		AP_2	2×2	-	Add_3	strides 2
		Conv1D_5	2	128	AP_2	strides 1, BN
		Conv1D_6	2	64	Conv1D_5	strides 1
		Add_4	-	-	Conv1D_5, Conv1D_6	-
		Sigmoid_2	-	-	Add_4	-
		Multiply_3	-	-	Conv1D_1, Sigmoid_2	-
		Add_5	-	-	Input, Multiply_3	-
Classification Part		FC_1	-	-	Multiply_2	dropout, ReLU
		FC_2	-	-	FC_1	-
		FC_3	-	-	Add_5	dropout, ReLU
		FC_4	-	-	FC_3	-
		Add_6	-	-	FC_2, FC_4	-

$$\mathcal{F}'_{se} = FC(FDR(F_{se})) \tag{11-b}$$

$$F_{out} = Sof\,(F'_{sa}\mathcal{F}_{se}) \tag{11-c}$$

where $Sof\,(\cdot)$ denotes the softmax function. $FDR(\cdot)$ is the combination of the FC, dropout, and ReLU layers. F_{sp} and F_{se} are the outputs of the spatial and spectral subnetwork, respectively.

III. EXPERIMENTS AND DISCUSSIONS

Three public data sets are employed for experiments. The factors that may influence the performance of the proposal are analyzed. The proposal is compared with the state-of-the-arts based on deep learning methods. Finally, the results are discussed.

A. DATA SETS

The data sets of Indian Pines (IN), Loukia (LK), and Botswana (BW) are shown in Fig. 8. IN is one of the most classical data sets. LK and BW are two newly published data sets. Data set IN includes 16 vegetation classes and has 145 × 145 pixels with a resolution of 20 m by pixel. The 20 bands corrupted by water absorption effects have been discarded. The remaining 200 bands are adopted for analysis

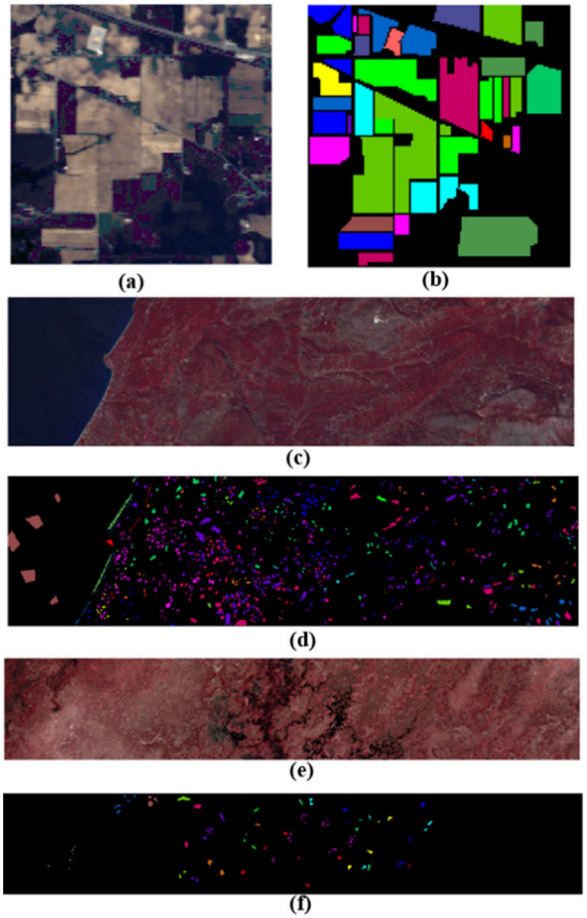

FIGURE 8. IN, LK and BW data sets: (a), (c) and (e) False-color images; (b), (d) and (f) Ground-truth maps.

and range from 400 to 2500 nm. Details of these classes and the number of training and testing samples in each category are shown in Table 2.

Data set LK includes 14 vegetation classes and has 250 × 1376 pixels with a resolution of 30 m by pixel. Two hundred twenty bands are used in experiments. Details of the land-cover types and the number of training and testing samples in each class are listed in Table 3.

Data set BW consists of 1476 × 256 pixels and 242 spectral bands ranging from 0.4 to 2.5 μm. The spatial and spectral resolutions are 30 m/p and 10 nm. By removing the uncalibrated and noisy bands that cover water absorption features, 145 bands remain. Details of the land-cover types and the number of training and testing samples in each class are listed in Table 4.

B. EXPERIMENTAL CONFIGURATION

The experiments are implemented on a computer with Intel(R) Xeon(R) Gold 5218 CPU at 2.30 GHz with 64 GB RAM and an NVIDIA GeForce RTX 3090 graphical processing unit (GPU) with 24 GB RAM. The software environment is the system of Ubuntu 14.04 ultimate 64-bit with deep learning frameworks of Pytorch.

To quantify the classification performance of the proposal, the overall accuracy (OA), average accuracy (AA), and kappa coefficient (Kappa). The higher scores they get, the superior the performance of the model. The values of OA, AA, and Kappa can be calculated as follows:

$$OA = sum(diag(M)/sum(M)) \tag{12-a}$$

$$AA = mean(diag\,(M)\,/sum(M,2)) \tag{12-b}$$

$$Kappa = \frac{OA - (sum(M,1)sum(M,2))/sum(M,2)^2}{1 - (sum(M,1)sum(M,2))/sum(M,2)^2} \tag{12-c}$$

where M represents the error matrix of classification results, $diag(M)$ is a vector of diagonal elements of M, $sum(\cdot)$ is the sum of all elements, $sum(\cdot,1)$ is the vector of the sum of elements in each column, $sum(\cdot,2)$ is the vector of the sum of elements in each row, $mean(\cdot)$ is the mean of all elements and represents the elementwise division.

TABLE 2. Amounts of training and test data on the IN data set.

NO.	Land-cover type	Training	Test
1	Alfalfa	9	37
2	Corn-notill	285	1143
3	Corn-mintill	166	664
4	Corn	47	190
5	Grass-pasture	97	386
6	Grass-trees	146	584
7	Grass-pasture-mowed	6	22
8	Hay-windrowed	96	382
9	Oats	4	16
10	Soybean-notill	194	778
11	Soybean-mintill	491	1964
12	Soybean-clean	118	475
13	Wheat	41	164
14	Woods	253	1012
15	Building-Grass-Trees-Drives	77	309
16	Stone-Steel-Towers	19	74
	Total	2049	8200

C. PARAMETERS SETTING

For PCA-based dimensionality reduction, the number of the preserved principal components is set to 30. The network is trained for 100 epochs with a batch size of 128 and a learning rate of 0.001. In particular, the optimization is performed by the diffGrad optimizer [64] which can control the learning rate based on the optimization phase.

1) EFFECT OF THE NUMBER OF CONVOLUTIONAL LAYERS

Fig. 9 shows the effect of the number of convolutional layers on the OA of the proposed network. The number of convolutional layers is calculated within each subnetwork, excluding those in the attention modules. Deeper networks generally have more powerful feature representation ability,

TABLE 3. Amounts of training and test data on the LK data set.

NO.	Land-cover type	Training	Test
1	Dense urban fabric	58	230
2	Mineral extraction sites	13	54
3	Non-irrigated arable land	108	434
4	Fruit trees	16	63
5	Olive groves	280	1121
6	Broad-leaved forest	45	178
7	Coniferous forest	100	400
8	Mixed forest	214	858
9	Dense sclerophyllous vegetation	759	3034
10	Sparce sclerophyllous vegetation	561	2242
11	Sparsely vegetated areas	81	323
12	Rocks and sand	97	390
13	Water	279	1114
14	Coastal water	90	361
	Total	2701	38072

TABLE 4. Amounts of training and test data on the BW data set.

NO.	Land-cover type	Training	Test
1	Water	54	216
2	Hippo grass	20	81
3	Floodplain grasses1	50	201
4	Floodplain grasses2	43	172
5	Reeds	51	202
6	Riparian	54	215
7	Firescar	52	207
8	Island interior	41	162
9	Acacia woodland	63	251
10	Acacia shrublands	50	198
11	Acacia grasslands	61	244
12	Short mopane	36	145
13	Mixed mopane	54	214
14	Exposed soils	19	76
	Total	650	2598

but too deep networks may cause gradient instability and network degradation. The highest accuracy can be obtained by selecting "3-3" convolution layers as shown in Fig. 9. "3-3" represents that the SPFS-RSSAN combined with 3 convolutional layers in the spatial and spectral subnetworks, respectively. Therefore, the number of convolutional layers in the following experiments is set to "3-3" for all data sets.

2) EFFECT OF THE INPUT PATCH SIZE

The effect of patch size is investigated as shown in Fig. 10. It can be seen that the OA shows an upward trend, while getting slower decrease after the size of 25. Generally, the larger patches contain more spatial information conducive to classification. However, when the patch size is too large, it may include negative information. On the other hand, a large patch will increase the computational load. Therefore, the patch size is set to 25 for all the data sets.

3) EFFECT OF THE TRAINING SAMPLE PROPORTION

In this session, the performance of the proposal with different proportions of training samples is investigated. For each data set, 1%, 2%, 5%, 10%, 15%, 20%, 25% and 30% of samples are randomly selected from each of land-cover categories as the training set. The experimental results are shown in Fig. 11. The OAs increase as the proportions of training samples on three data sets increasing. When the proportions of training samples of three data sets are more than 20%, separately, the OAs will keep in high level.

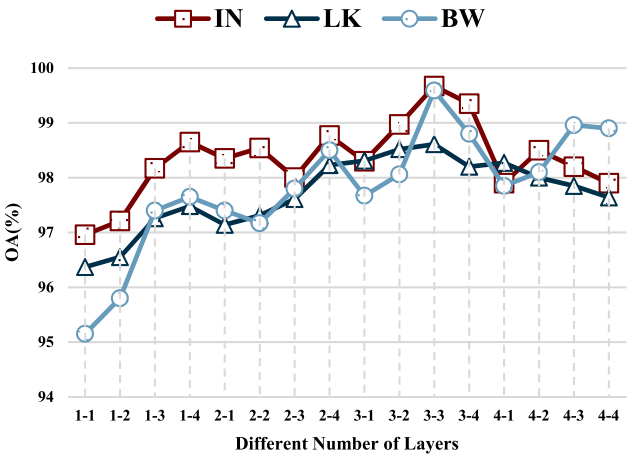

FIGURE 9. OAs of SPFS-RSSAN with several convolutional layers on the three datasets. "1-1" represents the SPFS-RSSAN model combined with one convolutional layer in the spatial subnetwork and one convolutional layer in the spectral subnetwork. The output of each layer is activated by the ReLU activation function.

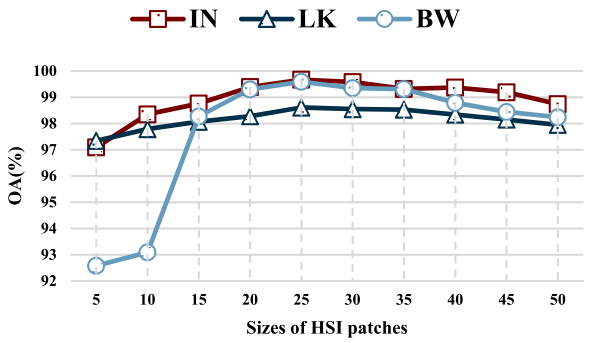

FIGURE 10. OAs of different sizes of HSI patches on three data sets.

D. ABLATION STUDY

In this section, a series of experiments are used to analyze the contribution of the proposed RSaAM, RSeAM and SPFSM.

To explore the correlations between the three modules and the impacts of them on the classification performance, eight schemes with different combinations of three modules are implemented on three data sets. The OAs of these schemes on the test sets are presented in Table 5. The spectral-spatial residual network (SSResNet) is the basic network. The last scheme is a complete SPFS-RSSAN model. The numbers

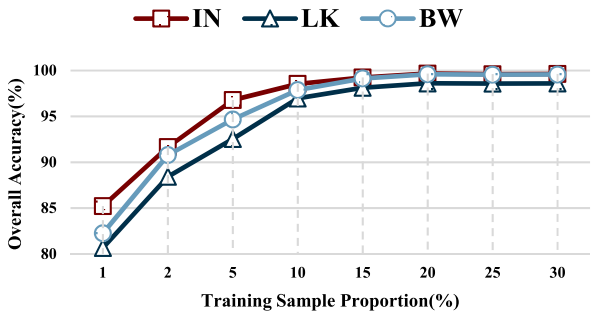

FIGURE 11. OAs (%) of different proportions of training samples on three data sets.

reported in bold-type denote the best results for each data set. Compared with SSResNet, the OAs of other schemes are improved, which shows the effectiveness of the proposed modules. Different from single module, higher OAs can be obtained when multiple modules are combined. Scheme 8 is a complete proposal. It receives the best classification accuracy on three data sets compared with the other combinations. Compared Scheme 6 and Scheme 7 with Scheme 8, the model using both RSaAM and SPFSM obtains better results than those models with separate module which shows the necessity of these two modules at the same time.

E. COMPARISON WITH STATE-OF-THE-ART METHODS

To evaluate the performance of the proposed method for HSIC, the proposal is compared with other existing methods, such as SVM [65], CNN [66], a deep feature fusion network (DFFN) [67], a CNN with mapping layers (MCNN) [68],

TABLE 5. OAs (%) with different module embedding.

Scheme		IN	LK	BW
1	SSResNet	95.88	95.27	95.91
2	SSResNet+RSeAM	96.27	96.49	96.33
3	SSResNet+RSaAM	96.52	96.34	96.48
4	SSResNet+SPFSM	96.77	96.82	96.83
5	SSResNet+RSaAM +SPFSM	97.59	96.96	98.33
6	SSResNet+RSeAM +SPFSM	98.29	97.36	97.69
7	SSResNet+RSaAM +RSeAM	97.13	96.83	97.44
8	SSResNet+RSaAM +RSeAM+SPFSM	**99.74**	**98.95**	**99.72**

SSRN [41], fast dense spectral-spatial convolution (FDSSC) [69], localized spectral features and multiscale spatial features network (LSMSC) [70], adaptive spectral–spatial multiscale network (ASSMN) [71], double-branch multi-attention mechanism network (DBMA) [72], double-branch dual-attention mechanism network (DBDA) [73], cross-attention spectral–spatial network (CASSN) [74], and attention-based adaptive spectral–spatial kernel improved residual network (A^2S^2KResNet) [75].Specifically, CNN is a traditional spectral–spatial network. SSRN and FDSSC are based on ResNet and DenseNet, respectively. LSMSC and ASSMN are the spectral–spatial multiscale networks. DBMA, DBDA, CASSN and A^2S^2KResNet are the spectral–spatial attention networks.

1) QUANTITATIVE EVALUATION

The quantitative metric comparisons of different methods are shown in Tables 6-8. The proposed method achieved higher classification accuracy than other methods. For example, in Table 6 the proposed method achieved the highest accuracy of 99.74%, which exceeds SVM, CNN, DFFN, MCNN, SSRN, FDSSC, LSMSC, ASSMN, DBMA, DBDA, CASSN and A^2S^2K ResNet by 18.07%, 2.33%, 4.36%, 8.03%, 0.55%, 0.16%, 3.03%, 1.44%, 3.19%, 1.44%, 1.22% and 0.96%, respectively. In Tables 7 and 8, the proposal also achieved the highest accuracy in OA, AA, and Kappa.

Compared with the SVM, the deep learning-based methods perform better. This is due to the fact that deep learning combines hierarchical feature learning with classifier learning, so the deep learning-based methods can learn more discriminative and abstract high-level features. It can be seen from the tables that the accuracy for CNN is lower than those of other deep learning-based methods since it only uses a weak 2-D CNN to extract the spatial features. Compared with CNN, the accuracy of SPFS-RSSAN is improved because it uses the spectral and spatial residual blocks to learn spectral and spatial features consecutively. Especially, the proposal outperforms the 3D-CNN-based model of SSRN which shows that the spectral feature is essential for HSIC. FDSSC uses densely connected structures to learn features deeply, obtaining better results than SSRN. LSMSC fuses localized spectral features.

ASSMN employs a multiscale strategy in spectral and spatial simultaneously. However, FDSSC, LSMSC, and ASSMN have different performances on all datasets. For example, FDSSC achieved good results on the IN and BW data sets, but its accuracy is low on the LK data sets. Among DBMA, DBDA, CASSN and A^2S^2KResNet that use attention modules, A^2S^2KResNet achieved the best results. The proposed model uses the RSaAM and RSeAM to strengthen these valuable features and weaken those useless or harmless information, so the proposed model achieved better accuracy than A^2S^2KResNet. Overall, the proposed SPFS-RSSAN model achieves better performance on all these three data sets.

In Tables 6-8, the last two rows record the training time and testing time of different models. Overall, the SPFS-RSSAN achieves the best balance between the computation time and the classification performance.

2) QUALITATIVE EVALUATION

Figs. 12–14 visualize the false-color images of the original HSI, their corresponding ground- truth maps, and the best

TABLE 6. Classification results of different methods for labeled pixels of the IN data set.

Class	SVM	CNN	DFFN	MCNN	SSRN	FDSSC	LSMSC	ASSMN	DBMA	DBDA	CASSN	A2S2K ResNet	SPFS-RSSAN
1	96.78	**100.00**	**100.00**	91.67	97.82	**100.00**	**100.00**	99.23	**100.00**	98.30	97.56	97.37	**100.00**
2	78.74	97.27	88.49	93.40	99.17	98.94	92.51	96.48	93.23	96.20	97.75	98.41	**99.88**
3	82.26	98.00	97.12	94.92	99.53	**100.00**	98.25	98.68	92.37	95.30	98.31	98.24	99.80
4	99.03	92.81	**100.00**	96.67	97.79	99.30	**100.00**	99.71	93.90	94.20	98.12	96.08	**100.00**
5	93.75	99.25	**100.00**	95.76	99.24	99.31	94.35	98.88	98.16	98.20	98.66	99.75	99.31
6	85.96	99.52	97.18	92.64	99.51	**100.00**	97.55	99.97	99.24	98.70	99.62	99.08	**100.00**
7	40.00	97.58	92.59	**100.00**	98.70	88.89	**100.00**	98.75	**100.00**	96.40	97.60	**100.00**	93.33
8	91.80	99.00	99.78	**100.00**	99.85	**100.00**	**100.00**	**100.00**	99.30	**100.00**	**100.00**	**100.00**	**100.00**
9	0.00	96.95	**100.00**	20.00	98.50	**100.00**	**100.00**	**100.00**	88.89	92.20	93.89	61.54	**100.00**
10	86.00	95.38	89.87	92.23	98.74	**100.00**	93.13	98.20	93.59	97.80	98.20	99.03	99.49
11	70.94	97.72	99.33	88.25	99.30	99.46	98.27	97.29	98.14	99.60	98.76	98.45	**99.86**
12	74.73	97.13	98.50	95.42	98.43	**99.15**	97.20	99.23	94.18	97.20	96.00	97.68	**99.72**
13	99.04	99.65	96.02	**100.00**	**100.00**	**100.00**	**100.00**	99.52	98.91	**100.00**	99.51	**100.00**	**100.00**
14	94.29	97.95	93.22	97.37	99.31	**100.00**	98.03	99.45	99.74	99.70	99.82	99.90	**100.00**
15	85.11	92.30	96.99	**100.00**	99.20	**100.00**	**100.00**	**100.00**	97.41	98.80	98.65	99.84	98.25
16	96.78	**100.00**	94.38	95.83	97.82	98.25	100.00	100.00	98.80	95.80	90.60	**100.00**	98.18
OA(%)	81.67	97.41	95.38	91.71	99.19	99.58	96.71	98.30	96.55	98.30	98.52	98.78	**99.74**
AA(%)	79.84	97.39	96.47	90.88	98.93	98.96	98.08	99.09	96.62	97.40	98.32	96.59	**99.24**
Kappa ×100	78.76	97.05	94.74	91.81	99.07	99.53	96.11	97.03	96.07	98.00	97.69	98.61	**99.70**
Training(s)	**126.72**	157.88	401.25	242.51	153.67	258.79	201.32	875.25	173.54	162.54	371.99	1657.9	287.62
Test(s)	**1.44**	1.51	3.79	2.42	2.71	3.88	4.56	35.52	5.25	4.12	4.23	14.93	2.79

TABLE 7. Classification results of different methods for labeled pixels of the LK data set.

Class	SVM	CNN	DFFN	MCNN	SSRN	FDSSC	LSMSC	ASSMN	DBMA	DBDA	CASSN	A^2S^2K ResNet	SPFS-RSSAN
1	94.60	83.22	**100.00**	99.50	99.79	93.26	99.92	95.91	95.64	97.67	99.42	86.79	99.42
2	68.64	**100.00**	69.14	98.87	98.59	**100.00**	99.15	**100.00**	91.30	98.14	**100.00**	**100.00**	**100.00**
3	81.11	94.22	**100.00**	96.29	99.09	96.53	97.20	93.22	98.99	**100.00**	99.07	92.64	99.39
4	65.45	22.22	**100.00**	79.78	98.54	88.89	99.46	80.95	82.23	76.92	90.91	56.25	95.65
5	89.10	90.46	87.04	78.42	92.05	97.41	85.52	96.80	**100.00**	97.75	98.58	93.39	99.29
6	69.28	67.33	99.07	90.63	93.82	89.06	93.55	77.06	95.58	92.80	95.52	66.88	93.85
7	80.07	83.10	**100.00**	**100.00**	98.36	93.71	99.83	93.66	**100.00**	94.77	96.22	82.55	99.00
8	89.36	93.02	**100.00**	95.03	97.83	96.48	99.77	95.85	95.63	97.40	98.16	87.78	98.91
9	55.53	88.33	82.07	90.85	93.22	94.02	97.54	91.96	96.50	96.20	96.98	90.48	**98.07**
10	81.69	84.32	99.50	**100.00**	99.35	93.57	**100.00**	90.22	98.79	95.27	96.71	91.13	97.62
11	92.48	93.72	97.13	98.04	99.82	97.52	97.16	97.05	99.67	97.93	98.33	97.79	99.59
12	90.91	93.24	**100.00**	96.16	98.51	98.27	97.02	98.26	**100.00**	**100.00**	98.98	97.50	**100.00**
13	88.59	**100.00**	90.70	93.70	98.43	**100.00**	99.58	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
14	**100.00**	**100.00**	92.11	**100.00**	99.65	99.63	99.69	**100.00**	**100.00**	98.98	**100.00**	**100.00**	**100.00**
OA(%)	80.90	89.66	95.79	95.22	97.26	95.53	97.18	93.88	96.24	96.86	97.78	91.80	**98.95**
AA(%)	81.92	85.23	94.05	94.09	97.60	95.60	97.53	93.64	96.74	95.99	97.78	88.80	**98.72**
Kappa ×100	79.30	87.64	94.41	94.42	97.02	94.70	96.94	92.71	95.93	96.28	97.36	90.24	**98.89**
Training(s)	**143.54**	183.65	546.23	356.79	289.87	272.89	332.97	1124.65	225.38	234.17	508.48	1131.61	392.44
Test(s)	**1.79**	2.03	4.56	3.65	3.79	4.78	6.72	38.54	7.21	4.87	5.11	17.11	3.11

classification results of different methods on three data sets. On three data sets, the classification maps of SVM, CNN, DFFN, MCNN, and DBMA have some dot noises in some classes. DBDA, CASSN and A^2S^2KResNet generated more

TABLE 8. Classification results of different methods for labeled pixels of the BW data set.

Class	SVM	CNN	DFFN	MCNN	SSRN	FDSSC	LSMSC	ASSMN	DBMA	DBDA	CASSN	A^2S^2K ResNet	**SPFS-RSSAN**
1	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	96.31	**100.00**	99.42	98.85	99.77	98.18	**100.00**	**100.00**
2	90.12	**100.00**	96.30	98.77	**100.00**	**100.00**	81.32	**100.00**	99.39	**100.00**	**100.00**	**100.00**	**100.00**
3	96.50	96.20	96.00	**100.00**	**100.00**	**100.00**	97.78	97.99	99.50	98.53	**100.00**	**100.00**	**100.00**
4	**100.00**	97.84	97.09	**100.00**	**100.00**	98.57	**100.00**	90.91	97.45	99.14	97.45	**100.00**	99.42
5	77.31	84.77	87.50	94.91	97.69	94.01	86.01	98.87	97.44	95.63	**97.86**	97.13	97.62
6	59.53	86.94	87.91	80.00	98.60	94.79	88.43	89.96	97.37	97.62	96.88	95.92	**98.58**
7	**100.00**	99.65	99.52	**100.00**	**100.00**	**100.00**	**100.00**	95.71	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
8	85.80	99.77	99.38	**100.00**	98.15	99.07	**100.00**	98.37	**100.00**	99.38	**100.00**	**100.00**	**100.00**
9	92.75	86.72	94.42	**100.00**	**100.00**	**100.00**	94.66	97.26	**100.00**	**100.00**	99.60	98.28	**100.00**
10	97.99	98.99	**100.00**	91.96	96.98	**100.00**	99.55	96.41	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
11	97.95	99.13	97.95	**100.00**	91.39	**100.00**	98.91	98.32	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
12	98.62	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	99.31	96.03	98.14	99.66	**100.00**	**100.00**
13	**100.00**	98.57	95.35	**100.00**	98.14	**100.00**	90.50	**100.00**	97.13	98.22	99.54	**100.00**	99.91
14	97.37	**100.00**	88.16	97.37	96.05	**100.00**	94.12	**100.00**	99.34	99.34	**100.00**	97.30	**100.00**
OA(%)	94.03	95.35	96.15	97.97	98.53	98.61	95.01	97.70	98.77	98.94	99.15	99.20	**99.72**
AA(%)	92.66	96.33	95.68	97.36	98.63	98.77	95.12	97.32	98.75	98.98	99.23	99.19	**99.77**
Kappa ×100	91.78	95.04	95.45	96.95	98.12	98.50	95.18	97.27	98.67	98.85	99.08	99.13	**99.69**
Training(s)	**104.92**	132.42	361.72	223.65	148.63	237.11	185.19	756.25	150.23	147.23	342.87	778.01	273.59
Test(s)	**1.28**	1.32	3.42	2.32	2.35	3.22	4.15	30.12	4.13	3.56	3.78	3.66	2.26

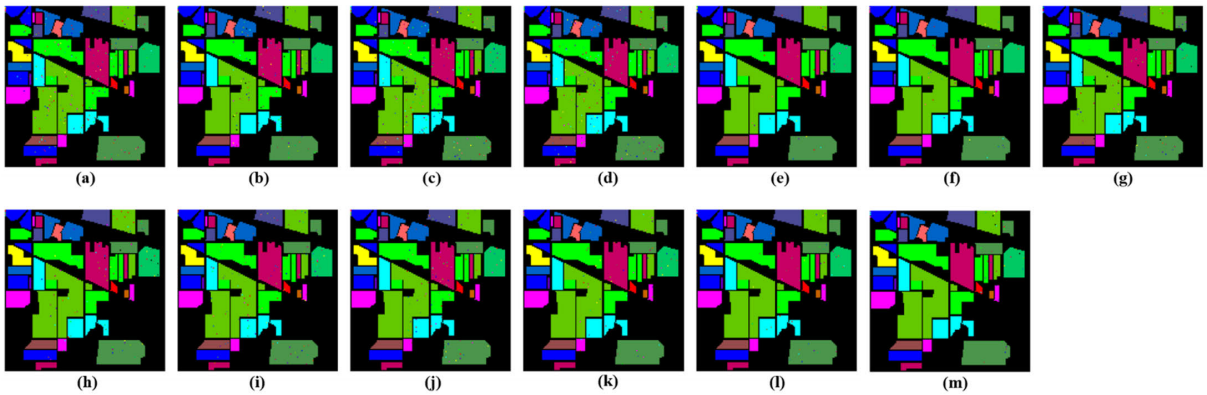

FIGURE 12. Classification maps for the IN data set. (a) SVM. (b) CNN. (c) DFFN. (d) MCNN. (e) SSRN. (f) FDSSC. (g) LSMSC. (h) ASSMN. (i) DBMA. (j) DBDA. (k) CASSN. (l) A2S2K ResNet. (m) SPFS-RSSAN.

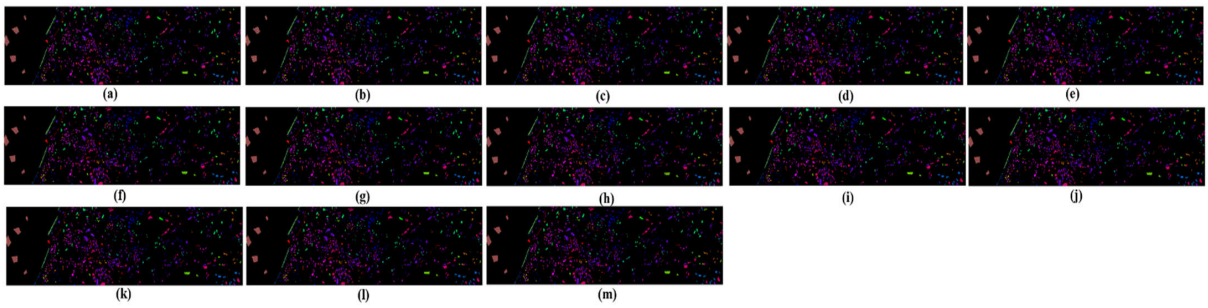

FIGURE 13. Classification maps for the LK data set. (a) SVM. (b) CNN. (c) DFFN. (d) MCNN. (e) SSRN. (f) FDSSC. (g) LSMSC. (h) ASSMN. (i) DBMA. (j) DBDA. (k) CASSN. (l) A2S2K ResNet. (m) SPFS-RSSAN.

smooth classification maps than these methods. Compared with other methods, the SPFS-RSSAN model generated the most accurate and smooth classification maps, especially in the boundary of two different classes. Our model uses the

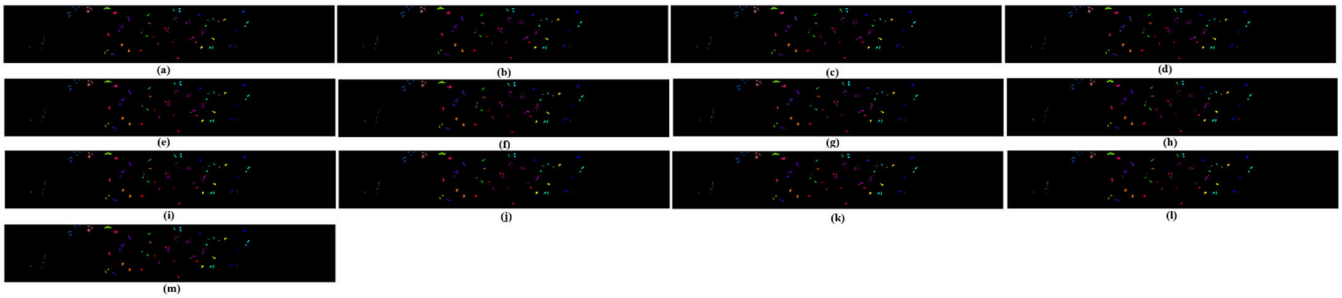

FIGURE 14. Classification maps for the BW data set. (a) SVM. (b) CNN. (c) DFFN. (d) MCNN. (e) SSRN. (f) FDSSC. (g) LSMSC. (h) ASSMN. (i) DBMA. (j) DBDA. (k) CASSN. (l) A2S2K ResNet. (m) SPFS-RSSAN.

spectral-spatial attention to learn the relationship between the center pixel and its surrounding pixels. So, the proposed model can correctly label almost all categories.

TABLE 9. Classification results on the disjoint train–test data set for the IN data set.

Class	SPFS-RSSAN
1	100.00
2	98.76
3	99.09
4	100.00
5	100.00
6	99.32
7	100.00
8	100.00
9	92.31
10	99.74
11	99.06
12	100.00
13	100.00
14	99.41
15	100.00
16	100.00
OA(%)	99.37
AA(%)	99.23
Kappa×100	99.28

F. EXPERIMENTAL RESULTS ON DISJOINT TRAIN/TEST SAMPLES

The results of the disjoint train-test [76] for the IN data set provided by the IEEE GRSS DASE are shown in Table 9. The superior classification performances demonstrate the excellent classification ability of the proposal.

IV. CONCLUSION

In this article, a novel SPFS-RSSAN model is proposed for HSIC, which uses the RSeAM, RSaAM, and SPFSM to implement the selection of spectral bands and spatial information. The RSeAM is used to distribute the weight of each spectral band of different original input data. The RSaAM enhances the spatial information related to the central pixel while suppressing the unnecessary, which can improve the recognition of central pixel. Compared with the existing methods, the proposal is improved in terms of accuracy and efficiency. In particular, a proximity selection-based SPFSM is played a key role. In the SPFSM, a tanh-derivative activation function is used to convert each spatial similarity to the appropriate weight. Such an operation can analyzes the characteristics of land-cover more sufficiently. The experimental results on three public data sets demonstrate the effective classification performances of the proposal. In the future, we plan to reduce the time cost brought by the iterative process to enhance the efficiency of the algorithm.

REFERENCES

[1] P. Ghamisi, M. D. Mura, and J. A. Benediktsson, "A survey on spectral–spatial classification techniques based on attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2335–2353, May 2015.

[2] B. Uzkent, A. Rangnekar, and M. J. Hoffman, "Aerial vehicle tracking by adaptive fusion of hyperspectral likelihood maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 233–242.

[3] J. Degerickx, M. Hermy, and B. Somers, "Mapping functional urban green types using hyperspectral remote sensing," in *Proc. Joint Urban Remote Sens. Event (JURSE)*, Mar. 2017, pp. 1–4.

[4] S. Zhong, C.-I. Chang, and Y. Zhang, "Iterative support vector machine for hyperspectral image classification," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3309–3312.

[5] U. Ergul and G. Bilgin, "Classification of hyperspectral images with multiple kernel extreme learning machine," in *Proc. 26th Signal Process. Commun. Appl. Conf. (SIU)*, May 2018, pp. 1–4.

[6] M. Khodadadzadeh, P. Ghamisi, C. Contreras, and R. Gloaguen, "Subspace multinomial logistic regression ensemble for classification of hyperspectral images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 5740–5743.

[7] L. Drumetz, M.-A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3890–3905, Aug. 2016.

[8] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[9] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[10] X. Zheng, Y. Yuan, and X. Lu, "Hyperspectral image denoising by fusing the selected related bands," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2596–2609, May 2019.

[11] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Neural ordinary differential equations for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1718–1734, Mar. 2020.
[12] S. Jia, H. Xie, and X. Deng, "Extended morphological profile-based Gabor wavelets for hyperspectral image classification," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 782–787.
[13] L. He, J. Li, A. Plaza, and Y. Li, "Discriminative low-rank Gabor filtering for spectral–spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1381–1395, Mar. 2017.
[14] W. Wang, Y. Qian, and Y. Y. Tang, "Hypergraph-regularized sparse NMF for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 681–694, Feb. 2016.
[15] X. Zhang, Z. Gao, L. Jiao, and H. Zhou, "Multifeature hyperspectral image classification with local and nonlocal spatial information via Markov random field in semantic space," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1409–1424, Mar. 2018.
[16] X. Cao, J. Yao, X. Fu, H. Bi, and D. Hong, "An enhanced 3-D discrete wavelet transform for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 6, pp. 1104–1108, Jun. 2021.
[17] Y. Y. Tang, Y. Lu, and H. Yuan, "Hyperspectral image classification based on three-dimensional scattering wavelet transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2467–2480, May 2015.
[18] L. He, C. Liu, J. Li, Y. Li, S. Li, and Z. Yu, "Hyperspectral image spectral–spatial-range Gabor filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4818–4836, Jul. 2020.
[19] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.
[20] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, "Perceptual generative adversarial networks for small object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1951–1959.
[21] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114417.
[22] P. Wang, E. Fan, and P. Wang, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *Pattern Recognit. Lett.*, vol. 141, pp. 61–67, Jan. 2021.
[23] Z. Meng, F. Zhao, and M. Liang, "SS-MLP: A novel spectral–spatial MLP architecture for hyperspectral image classification," *Remote Sens.*, vol. 13, no. 20, p. 4060, Oct. 2021.
[24] C. Shi and C.-M. Pun, "Multiscale superpixel-based hyperspectral image classification using recurrent neural networks with stacked autoencoders," *IEEE Trans. Multimedia*, vol. 22, no. 2, pp. 487–501, Feb. 2020.
[25] L. D. Medus, M. Saban, J. V. Francés-Víllora, M. Bataller-Mompeán, and A. Rosado-Muñoz, "Hyperspectral image classification using CNN: Application to industrial food packaging," *Food Control*, vol. 125, Jul. 2021, Art. no. 107962.
[26] J. Feng, L. Liu, X. Cao, L. Jiao, T. Sun, and X. Zhang, "Marginal stacked autoencoder with adaptively-spatial regularization for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3297–3311, Sep. 2018.
[27] X.-H. Han, B. Shi, and Y. Zheng, "SSF-CNN: Spatial and spectral fusion with CNN for hyperspectral image super-resolution," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2506–2510.
[28] J. Feng, N. Zhao, R. Shang, X. Zhang, and L. Jiao, "Self-supervised divide-and-conquer generative adversarial network for classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5536517.
[29] N. He, M. E. Paoletti, J. M. Haut, L. Fang, S. Li, A. Plaza, and J. Plaza, "Feature extraction with multiscale covariance maps for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 755–769, Feb. 2019.
[30] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, pp. 1–12, Jul. 2015, Art. no. 258619.
[31] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, Feb. 2017.
[32] H. Wu and S. Prasad, "Convolutional recurrent neural networks forHyperspectral data classification," *Remote Sens.*, vol. 9, no. 3, p. 298, Mar. 2017.
[33] H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1259–1270, Mar. 2018.
[34] X. Hu, Y. Zhong, X. Wang, C. Luo, J. Zhao, L. Lei, and L. Zhang, "SPNet: Spectral patching end-to-end classification network for UAV-borne hyperspectral imagery with high spatial and spectral resolutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503417.
[35] F. Tong and Y. Zhang, "Spectral–spatial and cascaded multilayer random forests for tree species classification in airborne hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4411711, doi: 10.1109/TGRS.2022.3177935.
[36] T. Song, Y. Wang, C. Gao, H. Chen, and J. Li, "MSLAN: A two-branch multidirectional spectral–spatial LSTM attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5528814, doi: 10.1109/TGRS.2022.3176216.
[37] M. Ahmad, A. M. Khan, M. Mazzara, S. Distefano, M. Ali, and M. S. Sarfraz, "A fast and compact 3-D CNN for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5502205.
[38] B. Praveen and V. Menon, "Study of spatial–spectral feature extraction frameworks with 3-D convolutional neural network for robust hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1717–1727, 2021.
[39] J. Lin, L. Mou, X. X. Zhu, X. Ji, and Z. J. Wang, "Attention-aware pseudo-3-D convolutional neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7790–7802, Feb. 2021.
[40] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017.
[41] 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.
[42] L. Mou, P. Ghamisi, and X. X. Zhu, "Unsupervised spectral–spatial feature learning via deep residual Conv–Deconv network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 391–406, Jan. 2018.
[43] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
[44] X. Zhang, E. Pan, Y. Ma, X. Dai, J. Huang, F. Fan, Q. Du, H. Zheng, and J. Ma, "Spectral–spatial self-attention networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5512115, doi: 10.1109/TGRS.2021.3102143.
[45] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Apr. 2016.
[46] J. Yue, W. Zhao, S. Mao, and H. J. R. S. L. Liu, "Spectral–spatial classification of hyperspectral images using deep convolutional neural networks," *Remote Sens. Lett.*, vol. 6, nos. 4–6, pp. 468–477, May 2015.
[47] M. Zhang, W. Li, and Q. Du, "Diverse region-based CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2623–2634, Jun. 2018.
[48] S. Mei, J. Ji, J. Hou, X. Li, and Q. Du, "Learning sensor-specific spatial–spectral features of hyperspectral images via convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4520–4533, Aug. 2017.
[49] Y. Xu, L. Zhang, B. Du, and F. Zhang, "Spectral–spatial unified networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5893–5909, Oct. 2018.
[50] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Learning and transferring deep joint spectral–spatial features for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4729–4742, Aug. 2017.
[51] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multisource remote sensing data classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 937–949, Feb. 2018.
[52] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 277–281, Jun. 2020.
[53] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. Int. Conf. Mach. Learn.*, Feb. 2015, pp. 2048–2057.

[54] Z. Yang, X. He, J. Gao, L. Deng, and A. Smola, "Stacked attention networks for image question answering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 21–29.
[55] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6450–6458.
[56] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141.
[57] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
[58] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6298–6306.
[59] J. Park, S. Woo, J.-Y. Lee, and I. So Kweon, "BAM: Bottleneck attention module," 2018, *arXiv:1807.06514*.
[60] C. Yu, R. Han, M. Song, C. Liu, and C.-I. Chang, "Feedback attention-based dense CNN for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501916.
[61] J. Feng, J. Chen, Q. Sun, R. Shang, X. Cao, X. Zhang, and L. Jiao, "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.
[62] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615.
[63] S. Kumar Roy, A. Deria, D. Hong, B. Rasti, A. Plaza, and J. Chanussot, "Multimodal fusion transformer for remote sensing image classification," 2022, *arXiv:2203.16952*.
[64] S. R. Dubey, S. Chakraborty, S. K. Roy, S. Mukherjee, S. K. Singh, and B. B. Chaudhuri, "DiffGrad: An optimization method for convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4500–4511, Nov. 2020.
[65] B. Waske, S. van der Linden, J. Benediktsson, A. Rabe, and P. Hostert, "Sensitivity of support vector machines to random feature selection in classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2880–2889, Jul. 2010.
[66] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Milan, Italy, Jul. 2015, pp. 4959–4962.
[67] W. Song, S. Li, L. Fang, and T. Lu, "Hyperspectral image classification with deep feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3173–3184, Jun. 2018.
[68] R. Li, Z. Pan, Y. Wang, and P. Wang, "A convolutional neural network with mapping layers for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3136–3147, May 2020.
[69] W. Wang, D. Shuguang, J. Zhongmin, and S. Liujie, "A fast dense spectral–spatial convolution network framework for hyperspectral images classification," *Remote Sens.*, vol. 10, no. 7, p. 1068, Jul. 2018.
[70] G. Sun, X. Zhang, X. Jia, J. Ren, A. Zhang, Y. Yao, and H. Zhao, "Deep fusion of localized spectral features and multi-scale spatial features for effective classification of hyperspectral images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 91, Sep. 2020, Art. no. 102157.
[71] D. Wang, B. Du, L. Zhang, and Y. Xu, "Adaptive spectral–spatial multiscale contextual feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2461–2477, Mar. 2021.
[72] H. Gao, Y. Yang, S. Lei, C. Li, H. Zhou, and X. Qu, "Multi-branch fusion network for hyperspectral image classification," *Knowl.-Based Syst.*, vol. 167, pp. 11–25, Mar. 2019.
[73] R. Li, S. Zheng, C. Duan, Y. Yang, and X. Wang, "Classification of hyperspectral image based on double-branch dual-attention mechanism network," *Remote Sens.*, vol. 12, no. 3, p. 582, Feb. 2020.
[74] K. Yang, H. Sun, C. Zou, and X. Lu, "Cross-attention spectral–spatial network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.
[75] S. K. Roy, S. Manna, T. Song, and L. Bruzzone, "Attention-based adaptive spectral–spatial kernel ResNet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7831–7843, Sep. 2021.
[76] M. Ahmad, S. Shabbir, S. K. Roy, D. Hong, X. Wu, J. Yao, A. M. Khan, M. Mazzara, S. Distefano, and J. Chanussot, "Hyperspectral image classification—Traditional to deep models: A survey for future prospects," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 968–999, 2022.

XINSHENG ZHANG received the B.S. degree in software engineering from Hainan Tropical Ocean University, Sanya, China, in 2020. He is currently pursuing the M.S. degree in electronic information with the School of Computer Science and Technology, Hainan University, Haikou, China. His research interests include hyperspectral image processing and analysis and deep learning.

ZHAOHUI WANG (Member, IEEE) received the M.S. degree in image processing from the University of Derby, U.K., in 2004, and the Ph.D. degree from the University of Leeds, U.K., in 2008. Then, he joined the Norwegian Colour and Visual Computing Laboratory, Gjøvik, Norway, to work on visual computing and multispectral color imaging research projects. He joined Hainan University, China, in 2013. He is currently a Professor of computer science with the Faculty of Computer Science and Technology. His current research interests include hyperspectral image processing and analysis, remote sensing image processing and its applications, computer vision, and deep learning. His professional memberships include IS&T, SPIE, and CCF.