

# Hyperspectral Image Classification Based on Expansion Convolution Network

Cuiping Shi<sup>1</sup>, Member, IEEE, Diling Liao, Tianyu Zhang, and Liguo Wang<sup>2</sup>, Member, IEEE

**Abstract**—In recent years, convolutional neural networks (CNNs) have achieved excellent performance in hyperspectral image classification and have been widely used. However, the convolution kernel used in traditional CNN has the limitation of single scale, which is not conducive to the improvement of hyperspectral classification performance. In addition, training a classification network of high-dimensional data based on limited labeled samples is still one of the challenges of hyperspectral image classification. To solve the above problems, a hyperspectral image classification method based on expansion convolution network (ECNet) is proposed. The expansion convolution injects holes into the standard convolution kernel to expand the receptive field (RF), so as to extract more context features. Because the shallow features of hyperspectral images contain more location and detail information, while the deep features contain stronger semantic information, in order to further enhance the correlation between deep and shallow information, inspired by ResNet, a similar feedback block (SFB) is introduced on the basis of ECNet, and the deep features and shallow features are fused through this feedback mechanism. Thus, an improved version of ECNet method is obtained, which is called feedback ECNet (FECNet). This study was tested on four commonly used hyperspectral datasets [i.e., Indian Pine (IP), Pavia University (UP), Kennedy Space Center (KSC), and Salinas Valley (SV)] and on a higher resolution and complexly distributed land cover dataset (University of Houston (HT)). The experimental results show that the proposed method has better classification performance than some state-of-the-art methods, which shows that FECNet has a certain potential in hyperspectral image classification.

**Index Terms**—Convolutional neural network (CNN), expansion convolution block (ECB), hyperspectral image (HSI) classification, similar feedback block (SFB).

## I. INTRODUCTION

IN THE past few decades, hyperspectral technology has been developed rapidly and widely used in many fields, for example, vegetation, estimation of soil salinity, geophysical exploration, and so on [1]–[5].

Manuscript received October 7, 2021; revised March 21, 2022 and April 30, 2022; accepted May 7, 2022. Date of publication May 10, 2022; date of current version May 23, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 41701479 and Grant 62071084, in part by the Heilongjiang Science Foundation Project of China under Grant LH2021D022, and in part by the Fundamental Research Funds in Heilongjiang Provincial Universities of China under Grant 135509136. (*Corresponding author: Cuiping Shi*)

Cuiping Shi, Diling Liao, and Tianyu Zhang are with the Department of Communication Engineering, Qiqihar University, Qiqihar 161000, China (e-mail: scp1980@126.com; 2020910228@qqhr.edu.cn; 2019910178@qqhr.edu.cn).

Liguo Wang is with the College of Information and Communication Engineering, Dalian Nationalities University, Dalian 116000, China (e-mail: wangliguo@hrbeu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2022.3174015

Hyperspectral image (HSIs) is a field with great research vitality and has also received extensive attention in the field of remote sensing [6]. In the early stage of HSIs research, many spectrum-based feature extraction methods were proposed successively, including support vector machines (SVMs) [7], multiple logistic regression [8], [9], and dynamic or random subspace [10], [11]. In addition, some methods based on feature extraction or dimensionality reduction have also attracted people's attention, such as principal component analysis (PCA) [12], independent component analysis (ICA) [13], and linear discriminant analysis (LDA) [14]. However, the above classification results based on pixelwise classifier are not satisfactory. In order to better classify hyperspectral images, some effective spatial–spectral feature representation methods are proposed [15], [16]. Among them, the classical methods based on spatial–spectral feature extraction include extended morphological profiles (EMPs) [17] and multiple kernel learning [18], [19]. In [20]–[22], the sparse representation model considers the adjacent spatial information. In addition, in [23]–[25], HSIs are divided into multiple superpixels to explore spatial consistency according to the similarity of texture. Although these methods based on spatial–spectral feature extraction are more effective, they are difficult to provide better classification performance in the case of high interclass similarity and large intraclass differences. Therefore, obtaining more discriminative features is the key to further improve the classification performance.

With the advent of the big data era, deep learning has developed rapidly in the past few years and has been applied in many fields, such as image processing [26], natural language processing [27], and so on. In the early development of deep learning, stacked autoencoder (SAE) [28] and recursive autoencoder (RAE) [29] were proposed and achieved good performance. However, this method can only deal with 1-D vectors, which leads to the destruction of the spatial information of HSIs. Subsequently, in [30], the restricted Boltzmann machine and deep belief network were used to extract features and pixel classification, retaining most of the feature information of HSIs. In addition, some methods based on 2-D convolutional neural network (CNN) have been proposed successively, including rolling guidance filter and vertex component analysis network (R-VCANet) [31], 2-D CNN [32], and so on. In order to better represent data, HSIs are generally regarded as a 3-D cube. Therefore, using the method based on 2-D CNN to deal with HSIs will lead to the complexity of convolution. In order to make up for the shortcomings of 2-D CNN, some methods of 3-D CNN are proposed. Lee and Kwon [33]

proposed a contextual deep CNN (CDCNN), which can jointly extract the spatial-spectral information of HSIs. However, with the deepening of the network, Hughes [34] phenomenon is likely to appear. In order to alleviate this problem, He *et al.* [35] proposed a new network structure ResNet. In addition, Zhong *et al.* [36] proposed a spatial-spectral-based residual network (SSRN). Paoletti *et al.* [37] proposed a pyramid residual network (PyResNet), which adds additional links to CNN and gradually increases the dimension of feature map in all Conv layers. In [37], a CNN based on dense connections (DenseNet) was proposed. By introducing dense connections into the network, feature propagation was strengthened and better classification performance was obtained. Because the features obtained by single-scale convolution kernel are not rich enough, many methods based on multiscale convolution kernel are used to extract richer features, which improve the classification performance of HSIs [39]–[41]. Recently, some new methods for hyperspectral image classification have been proposed. For example, Roy *et al.* [42] proposed a hybrid 3-D–2-D CNN model [Hybrid-spectral network (SN)], which reduces the complexity of 3-D-CNN by combining 2-D CNN and 3-D CNN. Meng *et al.* [43] proposed a two mixed link network, which can extract more discriminative features in hyperspectral images. Meanwhile, Roy *et al.* [44] proposed an improved residual attention network ( $A^2S^2K$ -ResNet) based on adaptive spatial-spectral kernel, which uses adaptive convolution kernel to expand receptive field (RF), so as to extract more effective features. Although standard convolution can provide better image classification performance, it often brings a lot of parameters and computation. Therefore, it is also a great challenge to build a network with low computing cost and excellent classification performance under limited data samples. Recently, Zhang *et al.* [45] proposed a lightweight 3-D asymmetric inception network (AINet). The network improves the classification performance by emphasizing the spectral features in the context of HSIs data space and the learning strategy of data fusion and migration. However, the classification performance of AINet is poor in the case of limited data samples. Inspired by GoogLeNet and PeleNet, Li and Duan [46] proposed a lightweight network (Lite-DenseNet). The number of training parameters was reduced through group convolution, and a 3-D two-way dense layer was constructed to capture the local and global features of the input. However, group convolution may affect classification accuracy due to ignoring the relationship between channels. Therefore, Cui *et al.* [47] proposed a new network structure (LiteDepthwiseNet), which greatly reduced the number of training parameters by decomposing the standard convolution into depth convolution and pointwise convolution. Ma *et al.* [48] proposed an end-to-end deep deconvolution network, which uses unpooling and deconvolution methods to recover the lost feature information of pooling operation, so as to retain most of the original information of HSIs. Yu and Koltun [49] proposed an expanded convolution network based on multiscale context aggregation. Similarly, in order to solve the problem of resolution and coverage loss caused by pooling layer, Pan *et al.* [50] proposed a semantic segmentation network (DSSNet) based on expanded convolution, which shows

that expanded convolution has certain potential for hyperspectral image processing. At the same time, with the development of CNN architecture, there are many excellent popular models in the field of hyperspectral image classification. In [51], a generative adversarial network (GAN) is proposed, which uses spatial spectrum classifier to fine tune the features. Roy *et al.* [52] proposed a 3-D-hyperGAMO model using a generative adversarial minority oversampling, which shows excellent data generation ability and significantly improves the classification performance. In addition, in order to better represent the sequence attributes of spectral features, Hong *et al.* [53] used transformer architecture in the task of hyperspectral image classification. This is also the first practice in the field of hyperspectral image classification based on transformer architecture. Although the big data samples of the network have obtained relatively good classification results, the classification performance is still poor under limited samples.

In recent years, attention mechanism also shows great potential in computer vision. In cognitive science, human beings tend to pay more attention to more important information and ignore other information. Attention mechanism can be regarded as the imitation of human vision and has been widely used in many fields of computer vision [54]–[56]. Wang *et al.* proposed a squeeze-and-excitation (SE) module [57] embedded in ResNet network. Ma *et al.* [58] proposed a dual branch and multiattention network (DBMA) to extract spatial-spectral information. Similarly, Li *et al.* proposed a dual attention network (DANet) [59] and achieved good classification results. In order to further improve the classification performance of HSIs, Li *et al.* [60] proposed double branch and double attention network (DBDA). However, because the spatial attention and channel attention of attention mechanism are usually separated from each other, Cui *et al.* [61] proposed a new dual triple attention network (DTAN), which can effectively classify hyperspectral images by capturing cross-dimensional interactive information.

Although the existing methods based on deep learning can effectively extract hyperspectral image features, hyperspectral image classification still faces many challenges, for example, limited training samples [62] and huge computing costs [63]. In order to solve these problems, a hyperspectral image classification method based on expansion convolution network (ECNet) is proposed in this article. Since the computational cost of expansion convolution will not increase with the increase of RF, the expanded convolution is adopted as the feature extraction unit. In order to further improve the classification performance of ECNet, a similar feedback block (SFB) is introduced to obtain an improved network feedback ECNet (FECNet).

The main contributions of this article are as follows.

- 1) In order to fully extract the spectral features of HSIs, an expanded convolution block (ECB) is proposed. This module not only increases the RF but also does not increase the computational amount, which alleviates the computational cost of training. In addition, the ECB-based method has good generalization ability.
- 2) A kind of similar feedback mechanism (SFM) is proposed. Specifically, SFM is a mechanism that generates

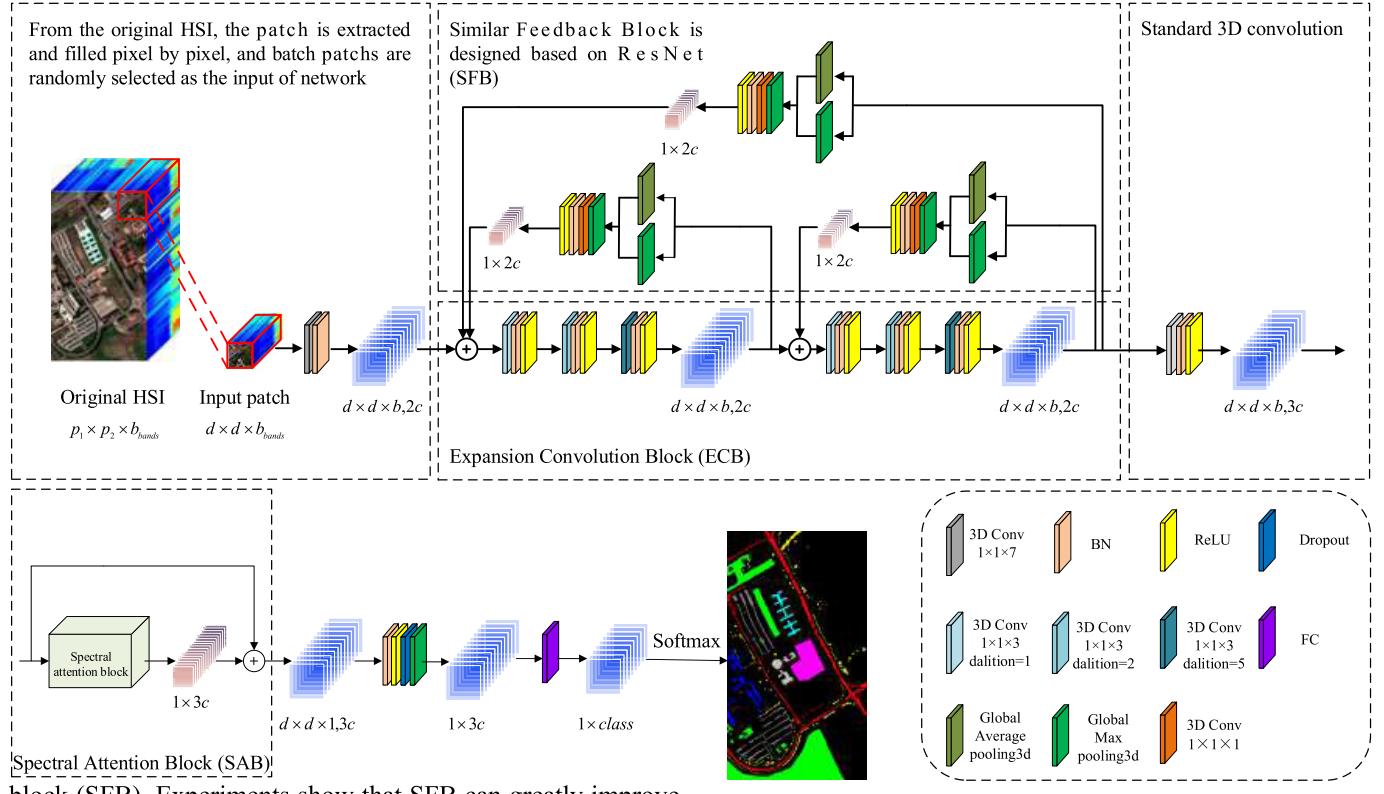


Fig. 1. It is a hierarchical diagram of the proposed FECNet. It is mainly composed of three blocks: SFM based on ResNet idea, ECB which can expand convolution RF, and SAB with emphasis ability. In addition, in order to classify more effectively, the original input HSI is the processed element by element and used as the input of the network (upper left corner). In particular, the designed ECNet does not have SFM of FECNet.

the attention maps with deep features through SFM and feedback to shallow features. Both the deep rich semantic information and the features with large RF are used to calculate the attention map, which significantly improves the classification performance.

- 3) In order to further enhance the propagation of features, SFM is densely connected to form a kind of SFB. Experiments show that SFB can greatly improve the classification performance of hyperspectral images.

The remaining parts of this article are organized as follows. Section II gives a detailed description of the proposed method ECNet and the improved version FECNet. In Section III, some experimental results are given and analyzed in detail. Section IV gives the conclusions.

## II. METHODOLOGY

The framework of the proposed FECNet is shown in Fig. 1. It is mainly composed of several components: ECB, spectral attention block (SAB), and SFB. In particular, the difference between the designed FECNet and ECNet is that FECNet adds SFB. Among them, ECB is used to obtain more contextual information. In the case of the same RF, this block requires fewer training parameters than the standard convolution block. In order to better extract important features and suppress useless features, SAB is adopted to emphasize the effectiveness of spectral bands. In order to further improve the representation ability of shallow features and deep features, SFB is designed to fuse deep features into shallow features

for feature extraction again. Next, Sections II-A and II-B will introduce the components of ECNet and the improved method FECNet, respectively.

### A. Proposed ECNet Method

The structure of ECNet is mainly composed of ECB and attention mechanism block (SAB). In order to describe the ECB proposed in this article clearly, the principles of 2-D expansion convolution and 3-D expansion convolution will be briefly introduced, and then, the component ECB and the SAB in ECNet will be introduced in detail.

*1) 2-D and 3-D Expansion Convolution:* In recent years, CNN has been widely used in the field of deep learning because of its strong feature extraction ability. However, due to the limitations of the traditional standard convolution, many different convolution methods have been derived. Among them, expansion convolution has also been widely used in deep learning because it can obtain a larger RF.

In order to illustrate the difference between 2-D expansion convolution and standard convolution, taking convolution kernel with size of  $3 \times 3$  as an example, the process of the 2-D expansion convolution is shown in Fig. 2 (where  $p \times p$  represents the size of space). From the perspective of the size of RF, the RF increases with the increase of expansion rate of expansion convolution. From the perspective of computational complexity, compared with the standard convolution, in the case of the same RF (excluding the expansion rate equal to 1), the parameters required for the expansion convolution training

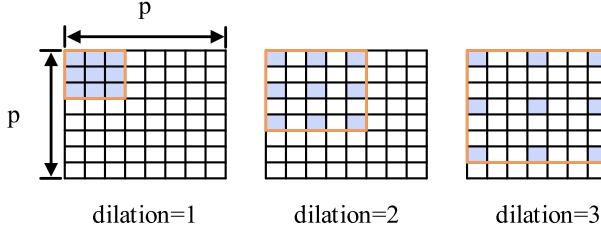


Fig. 2. Analytical diagram of 2-D expansion convolution. From left to right, the expansion rates of expansion convolution are 1, 2, and 3, respectively. Where, the yellow box is the equivalent convolution kernel size. The blue dot indicates the value of the corresponding position of the convolution kernel when the expansion rate is equal to 1. The white point in the yellow box represents the injected hole, and the value is equal to 0.

are less than the standard convolution and the greater the difference between the two parameters with the increase of the expansion rate. In order to more generally express the equivalent convolution kernel relationship, it is assumed that the size of convolution kernel is  $r \times r$ , the expansion rate is  $d$ , and the size of equivalent convolution kernel is  $r' \times r'$

$$r' = r + (r - 1)(d - 1). \quad (1)$$

It can be seen that when the expansion rate is equal to 1, the expansion convolution results are consistent with the standard convolution results. When the expansion rate is equal to 2, the RFs of the convolution kernel with the size of  $3 \times 3$  of expansion convolution and the convolution kernel with the size of  $5 \times 5$  of standard convolution are consistent. Similarly, when the expansion rate is equal to 3, the RFs of the convolution kernel with the size of  $3 \times 3$  of expansion convolution and the convolution kernel with the size of  $7 \times 7$  of standard convolution are consistent. Therefore, the general expression of RF is

$$R_{i+1} = R_i + (r' - 1)S_i \quad (2)$$

$$S_i = \prod_{i=1}^i \text{stride}(i) \quad (3)$$

where  $R_i$  is the RF of layer  $i$ th,  $R_{i+1}$  is the RF of layer  $i+1$ th, and  $S_i$  is the product of all strides of the previous layer  $i$ th.

Different from 2-D expansion convolution, the principle and corresponding relationship of 3-D expansion convolution are based on 3-D space, as shown in Fig. 3. Similarly, in order to illustrate the working process of 3-D expansion convolution, take the convolution kernel with the size of  $3 \times 3 \times 3$  as an example. The relationship between 3-D expansion convolution and 3-D standard convolution still follows the law of 2-D convolution. From the perspective of the size of RF, the RF still increases with the increase of expansion rate. From the point of view of computational complexity, when the expansion convolution and the standard convolution have the same RF (excluding the expansion rate of 1), the parameters required for the expansion convolution training are still less than those for the standard convolution and the greater the difference between the parameters required for expansion convolution and standard convolution with the increase of expansion rate.

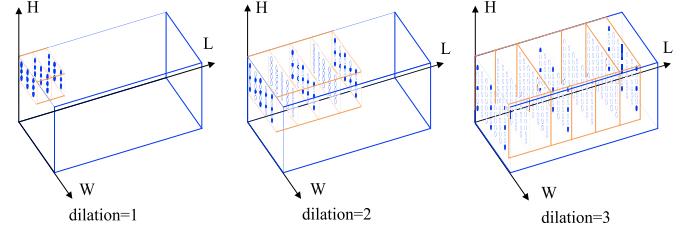


Fig. 3. Analytical diagram of 3-D expansion convolution. From left to right, the expansion rates of expansion convolution are 1, 2, and 3, respectively. Where, the yellow box is the equivalent convolution kernel size. The blue dot indicates the value of the corresponding position of the convolution kernel when the expansion rate is equal to 1. The white point in the yellow box represents the injected hole, and the value is equal to 0.

2) *Component ECB*: Convolution is a very effective feature extraction method in deep learning. Specifically, if the input is  $X \in R^{H \times W \times L}$  (where  $H$ ,  $W$ , and  $L$  are the input height, width, and number of channels, respectively) and the output is  $Y \in R^{H' \times W' \times L'}$ , and the input is convoluted

$$Y = X * W + b \quad (4)$$

where  $W \in R^{r \times r \times M \times N}$  is the weighted tensor (where  $r \times r$  is the space size of convolution kernel, and  $M$  and  $N$  is the number of input and output feature maps, respectively), and  $b$  is the bias term. In order to further compare the expansion convolution with the standard convolution, assuming that the expansion rate of the expansion convolution is  $d$ , the parameter and calculation quantity of the standard convolution are as follows:

$$f_{\text{STD}} = (r + 2(d - 1))^2 MNH'W' \quad (5)$$

$$F_{\text{STD}} = HW(r + 2(d - 1))^2 MNH'W' \quad (6)$$

where  $f_{\text{STD}}$  represents the parameter quantity of standard convolution, and  $F_{\text{STD}}$  represents the calculation quantity of standard convolution. It can be seen from the above introduction that expansion convolution expands RF without reducing the image resolution and adding additional parameters and calculation. The parameter quantity and calculation quantity of expansion convolution are as follows:

$$f_{\text{EXP}} = r^2 MNH'W' \quad (7)$$

$$F_{\text{EXP}} = HWr^2 MNH'W'. \quad (8)$$

It can be seen that standard convolution is a special form of expansion convolution. Where  $f_{\text{EXP}}$  represents the parameter quantity of expansion convolution, and  $F_{\text{EXP}}$  represents the calculation quantity of expansion convolution. In order to more intuitively show the difference between the calculation quantity and parameter quantity of standard convolution and expansion convolution when the standard convolution RF is consistent with the equivalent RF of expansion convolution, Fig. 4 shows the relationship between them. As can be seen from Fig. 4, with the increase of RF, the quantity of calculation and parameters of expansion convolution do not change, while the quantity of calculation and parameters of standard convolution increase exponentially.

For hyperspectral images, extracting multiscale features can greatly improve the classification performance. However,

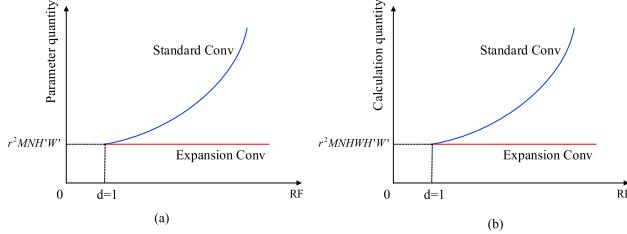


Fig. 4. Relationship between calculation quantity and parameter quantity of standard convolution and expansion convolution. (a) Relationship of parameter quantity. (b) Relationship of calculation quantity.

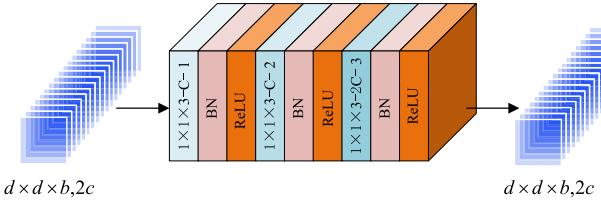


Fig. 5. Designed ECB.

there are still some contradictions in network design, such as the contradiction between RF and computational complexity and the contradiction between large RF and small RF. In order to alleviate the above contradictions, this article uses expansion convolution as the main feature extraction method of the network. The designed ECB is shown in Fig. 5. The block is mainly composed of three parts: expansion convolution layer, batch normalization (BN) layer, and activation function rectified linear unit (ReLU). Among them, the expansion convolution layer is expressed as “the size of convolution kernel—the number of output feature maps—expansion rate.” For example,  $1 \times 1 \times 3\text{-}C\text{-}1$  indicates that the size of the convolution kernel is  $1 \times 1 \times 3$ , the number of output feature maps is  $C$ , and the expansion rate is equal to 1.

In order to further prove the advantages of expansion convolution, multiple expansion convolution layers are connected sequentially to build the ECB. The block can not only obtain the detailed information through the expansion convolution with small expansion rate but also obtain the context information through the expansion convolution with large expansion rate, which makes the acquired features richer and has stronger discrimination ability. According to the principle of expansion convolution, this method can maximize the RF. In short, the RF size of the current layer is the expanded RF of the current layer superimposed on the RF of the previous layer. Since the space size of the convolution kernel of the block is  $1 \times 1$ , the RF calculation process of the 3-D expansion convolution with the convolution kernel of  $1 \times 1 \times 3$  is similar to that of the 2-D expansion convolution.

3) Standard 3-D Convolution: CNNs have strong representation ability. Because hyperspectral images contain spatial and spectral features, the use of 2-D convolution usually destroys the correlation among image samples. Therefore, the standard convolution of the proposed network adopts 3-D standard convolution. The calculation process of 3-D standard

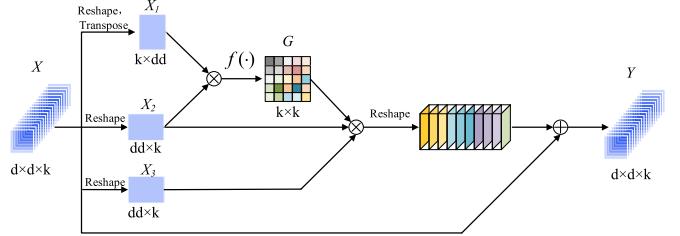


Fig. 6. Spectral attention mechanism block (SAB).

convolution is as follows:

$$v_{ij}^{xyz} = f \left( \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)} + b_{ij} \right) \quad (9)$$

where  $i$  represents the current layer,  $v_{ij}^{xyz}$  represents the output of the  $j$ th feature map of layer  $i$  at position  $(x, y, z)$ ,  $m$  represents the number of feature maps connected between layer  $i$  and layer  $(i - 1)$ ,  $w_{ijm}^{pqr}$  represents the weight of the  $m$ th feature map at position  $(p, q, r)$ ,  $b_{ij}$  represents the bias term,  $f$  represents the activation function, and  $P$ ,  $Q$ , and  $R$  represent the length, width, and height of the convolution kernel, respectively.

4) Component SAB: In neural network, attention mechanism can dynamically manage information flow and features, so as to improve learning effect. This mechanism filters out irrelevant stimuli and helps the network deal with long-term dependence. In order to focus on useful features, a spectral attention mechanism block (SAB) is designed in the network. The process of the designed SAB is introduced as follows.

The structure of SAB is shown in Fig. 6. The SAB can understand the relationship between spectral channels and setting each input element to a threshold of 0 to 1, which can reflect the importance or dependence of the element in the feature. Specifically, assuming that the input is  $P \in R^{d \times d \times k}$  (where  $d \times d$  is the space size and  $k$  represents the number of channels), in order to facilitate the calculation of the dependence between different positions, the input is transformed or transposed to obtain  $X_1$ ,  $X_2$ , and  $X_3$ , and then,  $X_1$  and  $X_2$  are matrix multiplied and the results perform  $f(\cdot)$  operation, where

$$f(\cdot) = \text{Soft max}(\cdot) \quad (10)$$

where  $f(\cdot)$  represents the activation function layer, which can organize the attention map into a probability distribution with the weighted sum of each channel as 1, which is recorded as  $G \in R^{k \times k}$

$$g_{ji} = \frac{\exp(X_i^T \cdot X_j)}{\sum_{\forall j} (X_i^T \cdot X_j)} \quad (11)$$

where  $g_{ji}$  is the weight coefficient of the  $i$ th channel to the  $j$ th channel, that is, the importance of the  $i$ th channel to the  $j$ th channel, and  $X_n (n = 1, 2, \dots, k)$  represents the  $n$ th channel of  $X$ . Set  $\alpha$  as the attention parameter (if  $\alpha = 0$ , the attention mechanism does not work)

$$Y_j = \alpha \sum_{\forall j} g_{ji} X_j + X_j \quad (12)$$

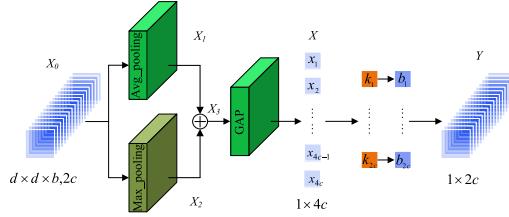


Fig. 7. Structure of SFM.

where  $Y_n (n = 1, 2, \dots, k)$  represents the  $n$ th channel of  $Y \in R^{d \times d \times k}$ .

### B. Improved Method FECNet of ECNet

Based on ECNet, the SFB is introduced, which is composed of multiple SFMs.

For a neural network, because shallow features contain more location and detail information, while deep features contain stronger semantic information, the extraction of shallow features can contribute to the extraction of deep features, and deep features can also provide feedback for shallow features. Therefore, the fusion of deep multiscale features and shallow features is an important way to improve the classification performance of hyperspectral images. Therefore, this article designs an SFM, which can feedback the deep features to the shallow features in the form of attention maps, and the shallow features are fused with the feedback attention maps, so as to realize the deep fusion of deep features and shallow features. The structure of SFM will be described in detail below.

The structure of SFM is shown in Fig. 7. Suppose that  $X \in R^{d \times d \times b}$  (where  $d \times d$  is the space size of the cube and  $b$  is the number of channels), and  $X_0$  generates two feature maps after passing through a max pooling layer and an average pooling layer, which are  $X_1$  and  $X_2$ , respectively. In order to balance the local information and global invariance, the obtained  $X_1$  and  $X_2$  are cascaded to obtain  $X_3$ . Then,  $X_3$  is passed through a global average pooling layer to obtain  $X$ . Finally,  $X$  is convoluted with  $K$  to obtain the correction matrix  $Y$ . That is

$$Y = G(X) = X * K + b \quad (13)$$

where  $G(\cdot)$  represents the convolution function, and  $b$  represents the offset term of the convolution. In SFB,  $X = \{x|x_1, x_2, \dots, x_{4c}\} \in R^{1 \times 1 \times 4c}$ ,  $K = \{k|k_1, k_2, \dots, k_{2c}\} \in R^{1 \times 1 \times 2c}$ ,  $b = [b_1, b_2, \dots, b_{2c}]^T$ , and  $Y \in R^{1 \times 2c}$ .

In order to fully feedback the semantic information of the later layer, SFM is densely connected to form SFB, as shown in Fig. 8. This dense connection method is inspired by the idea of ResNet, which feedback the deep features to all the previous layers. After the deep feature passes through the SFB module, a feedback correction matrix is generated, which can represent the semantic information of the deep feature. The obtained correction matrix is loaded into the shallow feature, so that the shallow feature can obtain the initial weight value related to the deep feature. This kind of feedback block can well combine the deep and shallow feature information and

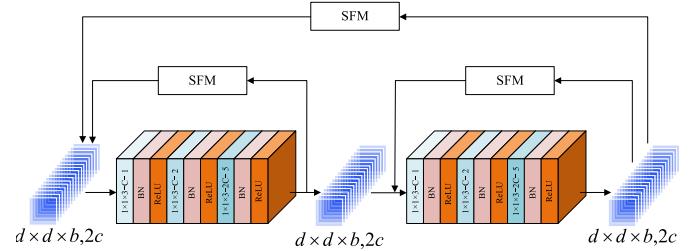
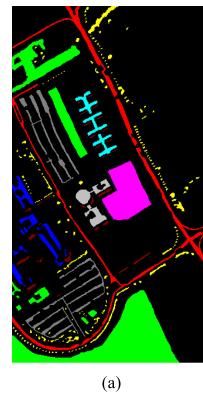


Fig. 8. Structure of SFB. The structure consists of three SFMs.

Class	Color	Land cover	Samples
C1	Red	Alfalfa	46
C2	Green	Corn-notill	1428
C3	Blue	Corn-mintill	830
C4	Yellow	Corn	237
C5	Cyan	Grass/pasture	483
C6	Magenta	Grass/trees	730
C7	Grey	Grass/pasture-mowed	28
C8	Dark Grey	Hay-windrowed	478
C9	Maroon	Oats	20
C10	Pink	Soybean-notill	972
C11	Dark Green	Soybean-mintill	2455
C12	Purple	Soybean-clean	593
C13	Dark Blue	Wheat	205
C14	Dark Orange	Woods	1265
C15	Orange	Bldg-Grass-Tree-Drivers	386
C16	Yellow	Stone-Steel-Towers	93
TOTAL			10249

(a) (b)

Fig. 9. Real land cover of IP scene. (a) Ground truth. (b) Label of IP scene.



Class	Color	Land cover	Samples
C1	Red	Asphalt	6631
C2	Green	Meadows	18649
C3	Blue	Gravel	2099
C4	Yellow	Trees	3064
C5	Cyan	Painted metal sheets	1345
C6	Magenta	Bare Soil	5029
C7	Grey	Bitumen	1330
C8	Dark Grey	Self-Blocking Bricks	3682
C9	Maroon	Shadows	947
TOTAL			42776

(a) (b)

Fig. 10. Real land cover of UP scene. (a) Ground truth. (b) Label of UP scene.

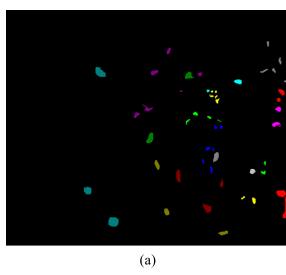
form the direct correlation between deep and shallow features, which is very beneficial to effective feature extraction.

## III. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, in order to verify the effectiveness of the proposed ECNet and the improved FECNet, a large number of experiments are carried out on five HSIs datasets.

### A. Experimental Data Description

The experiment used four commonly used public datasets and a higher resolution dataset, including Indian pine (IP), Pavia University (UP), Kennedy Space Center (KSC), Salinas Valley (SV), and University of Houston (HT). The color composite image of the dataset, the real classification map of features, and the details of each category are shown in Figs. 9–13.

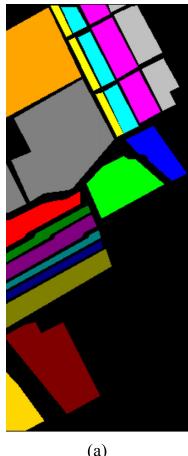


(a)

Class	Color	Land cover	Samples
C1	Red	Scrub	761
C2	Green	Willow-swamp	243
C3	Blue	CP-hammock	256
C4	Yellow	Slash-pine	252
C5	Cyan	Oak/Broadleaf	161
C6	Magenta	Hardwood	229
C7	Grey	Swap	105
C8	Dark Grey	Graminoid-marsh	431
C9	Dark Red	Spartina-marsh	520
C10	Pink	Cattail-marsh	404
C11	Gold	Salt-marsh	419
C12	Purple	Mud-flats	503
C13	Dark Green	Water	927
<b>TOTAL</b>			<b>5211</b>

(b)

Fig. 11. Real land cover of KSC scene. (a) Ground truth. (b) Label of KSC scene.

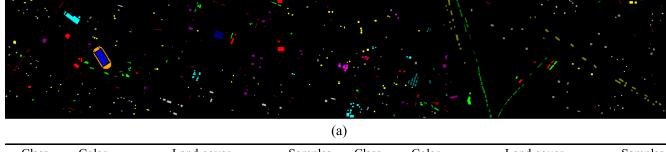


(a)

Class	Color	Land cover	Samples
C1	Red	Broccoli-green-weeds_1	2009
C2	Green	Broccoli-green-weeds_2	3726
C3	Blue	Fallow	1976
C4	Yellow	Fallow-rough-plow	1394
C5	Cyan	Fallow-smooth	2678
C6	Magenta	Stubble	3959
C7	Grey	Celery	3579
C8	Dark Grey	Grapes-untrained	11271
C9	Dark Red	Soil-vinyard-develop	6203
C10	Pink	Corn-senesced-green-weeds	3278
C11	Gold	Lettuce-romaine-4wk	1068
C12	Purple	Lettuce-romaine-5wk	1927
C13	Dark Green	Lettuce-romaine-6wk	916
C14	Dark Blue	Lettuce-romaine-7wk	1070
C15	Orange	Vinyard-untrained	7268
C16	Yellow	Vinyard-vertical-trellis	1807
<b>TOTAL</b>			<b>54129</b>

(b)

Fig. 12. Real land cover of SV scene. (a) Ground truth. (b) Label of SV scene.



(a)

Class	Color	Land cover	Samples	Class	Color	Land cover	Samples
C1	Red	Healthy grass	1251	C9	Dark Red	Road	1252
C2	Green	Stressed grass	1254	C10	Pink	Highway	1227
C3	Blue	Synthetic grass	697	C11	Gold	Railway	1235
C4	Yellow	Trees	1244	C12	Purple	Parking Lot 1	1233
C5	Cyan	Soil	1242	C13	Dark Green	Parking Lot 2	469
C6	Magenta	Water	325	C14	Dark Blue	Tennis Court	428
C7	Grey	Residential	1268	C15	Orange	Running Track	660
C8	Dark Grey	Commercial	1244	<b>TOTAL</b>			<b>15029</b>

(b)

Fig. 13. Real land cover of HT scene. (a) Ground truth. (b) Label of HT scene.

- IP: The IP dataset was captured by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor in June 1992. The space size of the dataset is  $145 \times 145$ , the number of bands is 220, and the wavelength range is  $0.2\text{--}2.4 \mu\text{m}$ . Among them, except for the number of bands with water absorption and low signal-to-noise ratio (bands 108–112, 150–163, and 220), there are still 200 bands available for experiments. The real category of the image is 16 (as shown in Fig. 9).
- University of Pavia (UP): The UP dataset was obtained by the Reflective Optical Spectral Imaging System

(ROSI-3) sensor. The dataset contains nine real categories with an image size of  $610 \times 340$ , with a spatial resolution of  $1.3 \text{ m}$ , 115 bands, and a wavelength range of  $0.43\text{--}0.86 \mu\text{m}$ . In addition to 13 noise bands, 103 bands are left for the experiment (as shown in Fig. 10).

- KSC: The acquisition tool of Kennedy aviation center dataset is the same as that of IP dataset, which was captured by AVIRIS sensor in Florida in 1996. The space size of the image is  $512 \times 614$ . After removing the water absorption band, 176 bands are left for the experiment. The spatial resolution of the image is  $20 \text{ m}$ , and the spectral range includes  $400\text{--}2500 \text{ nm}$ . The image contains 13 different land cover categories (as shown in Fig. 11).
- SV: The SV dataset was also collected by AVIRIS sensors. The dataset is characterized by the regular distribution of different ground cover categories. The image contains 16 categories and the space size is  $512 \times 217$ , except for the water absorption bands (108–112, 154–167, and 224), and there are 204 bands left for the experiment (as shown in Fig. 12).
- HT: The HT dataset was acquired by the Compact Airborne Spectral Imager (CASI) sensor on the HT Campus in June 2012. The space size of this scene is  $349 \times 1905$ , the number of bands is 114, and the wavelength range is  $380\text{--}1050 \text{ nm}$ . The image contains 15 land cover categories (as shown in Fig. 13).

## B. Experimental Setup

The batch size and training epochs of the design network are set to 16 and 200, respectively, and Adam algorithm is used for optimization. During the experiment, the setting range of learning rate is 0.001, 0.005, 0.0001, 0.0005, and 0.00005. Through multiple experiments on each learning rate, the learning rate is finally set to 0.0005. The designed network is designed and implemented by pytorch framework. All data results are the average of 30 experimental results, and all experiments are implemented in the same configuration. Among them, the hardware platform of the experiment is Intel<sup>1</sup> Core<sup>2</sup> i9-9900K CPU, NVIDIA GeForce RTX 2080Ti GPU, and 32-GB RAM. The experimental software platform is based on windows10 VSCode operating system, including CUDA 10.0, Pytorch 1.2.0, and Python 3.7.4.

In order to evaluate the classification performance of different methods, the overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa) are used as evaluation indexes.

## C. Classification Results

In order to verify the effectiveness of ECNet and FECNet in hyperspectral image classification, the proposed method is compared with two kinds of methods: one is the classical method SVM and the other is ten methods based on deep learning, including SSRN, CDCNN, PyResNet, DBMA, DBDA,

<sup>1</sup>Register trademark.

<sup>2</sup>Trademarked.

TABLE I  
CLASSIFICATION RESULTS OF IP DATASET USING 3% TRAINING SAMPLES (%)

Class	SVM [7]	SSRN [36]	CDCNN [33]	PyResNet [37]	DBMA [53]	DBDA [55]	Hybrid-SN [42]	$A^2S^2K$ -ResNet [44]	DSSNet [48]	3D-HyperGAMO[52]	SpectralFormer[53]	ECNet	FECNet
C1	36.62±0.00	82.54±8.88	36.96±1.96	26.67±5.88	82.05±5.25	97.49±0.55	81.79±2.93	93.43±0.50	25.16±14.62	74.94±5.19	38.16±1.52	97.99±0.33	96.92±1.03
C2	55.49±0.00	89.19±1.53	58.11±3.42	80.92±4.12	85.73±1.00	93.25±1.85	69.12±6.24	93.01±2.37	50.84±2.56	90.30±0.87	68.56±2.29	78.87±2.87	95.65±0.11
C3	62.55±0.38	87.67±0.88	48.36±14.64	81.24±8.79	88.44±4.26	92.60±1.07	91.00±0.81	90.25±0.37	43.09±6.07	92.23±0.06	57.29±1.14	95.01±0.35	94.82±1.51
C4	42.54±0.00	84.28±1.23	37.41±13.85	62.17±7.15	87.79±2.27	93.63±1.07	84.87±8.4	89.94±1.70	49.27±14.71	89.73±2.62	56.19±0.81	94.03±0.83	95.76±0.91
C5	85.05±0.00	97.77±0.37	69.81±13.94	91.75±1.81	94.85±1.38	98.76±0.27	90.73±2.91	97.78±0.27	31.6±13.16	96.70±0.75	87.52±1.46	98.44±0.27	98.39±0.01
C6	83.32±0.00	96.43±0.58	84.03±5.64	94.26±1.31	97.33±0.44	97.85±0.84	88.59±1.95	98.25±1.12	52.56±12.11	98.40±0.11	87.72±3.81	98.04±0.38	98.02±0.44
C7	59.87±0.00	86.99±2.60	20.21±19.78	19.75±17.5	50.91±3.85	66.62±3.37	83.62±18.6	81.80±0.97	7.96±4.97	72.16±1.42	16.87±3.31	70.73±4.79	72.49±1.40
C8	89.67±0.00	96.76±0.61	87.9±1.68	100.00±0.00	98.62±0.41	99.75±0.24	87.24±4.60	99.20±0.30	70.84±8.12	93.78±2.48	99.79±0.30	99.70±0.42	100±0.00
C9	39.45±0.29	72.15±11.9	46.57±32.58	69.09±2.11	51.31±0.74	84.42±6.05	60.44±7.11	64.65±4.28	17.15±3.3	80.13±3.75	13.24±2.87	80.35±18.24	77.8±0.86
C10	62.32±0.00	85.92±3.51	45.47±11.12	82.96±1.43	84.22±5.24	87.47±0.79	86.25±2.08	89.08±1.02	40.76±0.31	92.04±0.52	68.35±2.91	91.61±0.14	91.77±0.50
C11	63.73±1.73	89.27±1.20	69.47±6.45	89.59±0.74	87.51±1.68	94.12±1.65	88.95±3.83	90.52±0.93	54.91±0.88	94.14±0.52	74.96±3.12	95.23±0.10	96.66±0.23
C12	50.55±0.00	86.33±0.88	36.42±5.2	59.82±2.27	81.18±1.41	92.22±4.95	79.03±2.3	93.66±2.9	37.93±0.43	89.83±0.97	64.90±3.34	92.37±0.05	93.12±0.49
C13	86.74±0.00	99.14±0.13	80.06±5.43	80.07±2.03	94.8±1.89	97.69±0.22	93.64±3.99	98.74±0.62	46.49±14.1	91.17±5.35	90.50±3.97	97.84±0.83	97.49±0.25
C14	88.67±0.00	95.54±0.52	87.44±8.09	96.31±1.56	95.52±0.75	97.15±0.31	92.65±0.71	95.68±1.34	69.46±5.73	97.91±0.25	93.38±1.34	98.27±0.02	98.04±0.18
C15	61.82±0.00	89.64±1.67	63.19±10.18	86.36±4.18	83.19±0.59	93.37±1.19	88.83±3.65	91.86±2.11	43.40±7.28	93.34±1.85	55.82±2.46	94.58±0.08	94.27±1.00
C16	98.66±0.00	95.47±1.20	83.84±13.07	90.37±4.63	93.47±0.51	91.83±0.66	92.23±2.54	94.27±0.45	52.41±14.32	84.86±7.76	82.55±2.45	94.52±0.30	94.47±2.45
OA	68.76±0.00	90.25±0.42	64.86±3.76	85.65±1.45	87.95±1.07	93.58±0.55	82.18±1.50	92.55±0.11	48.32±2.41	93.50±0.07	71.32±2.31	95.33±0.17	95.81±0.00
AA	66.73±0.00	89.69±0.97	58.67±8.31	75.67±1.27	84.8±0.61	92.17±0.25	84.31±1.61	91.29±0.25	43.36±4.07	89.48±0.06	66.08±1.38	93.07±0.60	93.48±0.26
K <sub>100</sub>	63.98±0.00	88.87±0.48	59.53±4.99	83.6±1.64	86.24±1.21	92.69±0.64	79.85±1.42	91.48±0.12	40.7±2.93	92.59±0.07	67.31±2.50	94.68±0.21	95.22±0.00

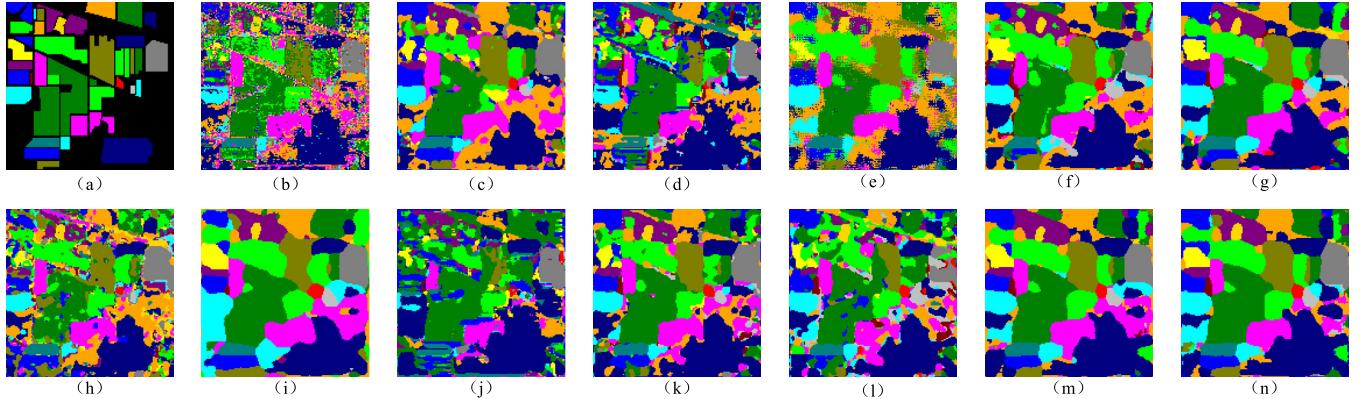


Fig. 14. Classification maps obtained by different methods on IP dataset (Sample proportion is 3%). (a) Ground truth map. (b) SVM (68.76%). (c) SSRN (90.25%). (d) CDCNN (64.86%). (e) PyResNet (85.65%). (f) DBMA (87.95%). (g) DBDA (93.58%). (h) Hybrid-SN (82.18%). (i)  $A^2S^2K$ -ResNet (92.55%). (j) DSSNet (48.32%). (k) 3-D-HyperGAMO (93.50%). (l) SpectralFormer (71.32%). (m) ECNet (95.33%). (n) FECNet (95.81%).

Hybrid-SN,  $A^2S^2K$ -ResNet, DSSNet, 3-D-HyperGAMO, and SpectralFormer.

SVM is a spectral classifier based on radial basis function (RBF) [7]. CDCNN is a deep context CNN, which can optimally explore the local context by combining the spatial-spectral relationship of a single pixel vector. PyResNet is an improvement of ResNet. It adds additional links to ordinary CNN and gradually increases the dimension of feature map in all Conv layers. Different from the above methods, DBMA and DBDA are two branch networks, which are used to extract spatial-spectral features, respectively. Hybrid-SN is a hybrid model of 3-D-CNN and 2-D-CNN. The model uses 3-D-CNN and 2-D-CNN to extract spectral-spatial features, respectively.  $A^2S^2K$ -ResNet improves ResNet by using adaptive spatial-spectral kernel and captures more discriminative spatial-spectral features through end-to-end training. DSSNet is a segmentation network based on expansion convolution, which aims to solve the problem that pooling operation may lead to the loss of resolution and coverage. 3-D-HyperGAMO

is a GAN. The data generated by the generator are trained with the original data to further improve the performance of the classifier. SpectralFormer is the first method applied in the field of hyperspectral image classification based on transformer.

Different space input sizes have a great impact on classification performance. For fair comparison, the input space size of all methods is set to  $9 \times 9$ . In addition, the percentage of training samples for IP, UP, KSC, SV, and HT datasets is set to 3%, 0.5%, 5%, 0.5%, and 2%.

1) *Classification Results of IP Dataset:* Table I and Fig. 14 show the numerical accuracy and visual results of ECNet and FECNet methods compared with other methods. It can be seen in Table I that the three indicators OA, AA, and Kappa of ECNet and FECNet are higher than those of other methods. Among them, FECNet obtained the best OA (95.81%), AA (93.48%), and Kappa (95.22%), while ECNet obtained OA, AA, and Kappa only 0.48%, 0.41%, and 0.54% lower than FECNet. In addition, compared with other methods, the increases of OA obtained by the proposed ECNet

TABLE II  
CLASSIFICATION RESULTS OF UP DATASET USING 0.5% TRAINING SAMPLES (%)

Class	SVM [7]	SSRN [36]	CDCNN [33]	PyResNet [37]	DBMA [53]	DBDA [55]	Hybrid-SN [42]	A <sup>2</sup> S <sup>2</sup> K-ResNet [44]	DSSNet [48]	3D-HyperGAMO [52]	SpectralFormer [53]	ECNet	FECNet
C1	81.26±0.00	94.10±2.21	86.77±0.47	88.11±6.5	89.82±1.38	93.50±0.86	74.94±2.81	81.61±6.34	65.42±6.68	85.26±0.30	89.08±2.43	97.00±0.93	96.51±0.90
C2	84.53±0.00	96.66±0.79	93.72±0.38	97.77±1.61	96.08±0.05	99.08±0.16	91.45±1.91	91.26±2.12	81.79±3.69	97.37±0.02	90.09±1.77	99.45±0.12	99.24±0.28
C3	56.56±0.00	76.75±5.41	64.27±0.76	30.97±18.97	76.09±6.26	88.85±3.32	68.9±0.62	76.49±8.05	10.57±4.34	82.13±5.19	62.09±0.19	90.07±4.69	93.59±0.81
C4	94.34±0.00	99.29±0.08	95.12±0.83	84.79±9.42	95.70±1.50	97.26±0.25	85.51±6.6	99.05±0.38	49.73±1.4	95.09±0.95	95.66±0.18	97.66±0.33	98.12±0.42
C5	95.38±0.00	99.64±0.20	96.52±0.79	96.64±4.42	98.45±0.50	98.83±0.32	94.18±6.6	99.3±0.50	62.31±1.91	97.14±0.88	99.22±0.47	98.34±0.4	98.68±0.06
C6	80.66±0.00	93.85±2.60	88.61±6.95	54.3±13.16	92.65±1.19	97.46±0.85	86.22±3.86	94.00±1.37	34.24±7.43	94.34±0.40	69.95±4.02	100±0.0	100±0.0
C7	49.13±0.00	86.48±4.29	77.29±3.54	38.3±25.69	86.72±12.62	91.61±6.65	62.59±4.43	95.99±5.58	33.13±2.24	83.71±0.52	66.01±2.09	99.23±0.05	99.34±0.33
C8	71.16±0.00	83.71±3.29	79.52±0.3	75.5±18.22	80.18±2.36	88.42±2.27	57.66±2.91	65.54±0.66	47.04±8.37	75.46±0.96	67.84±0.34	89.59±1.82	91.37±0.66
C9	99.04±0.00	98.97±0.31	91.04±0.57	91.15±8.5	94.38±1.41	97.48±0.77	62.55±2.26	92.94±0.62	65.68±5.51	83.22±0.44	89.36±1.53	98.88±0.35	98.16±0.69
OA	82.06±0.00	92.50±1.33	87.94±0.13	83.01±1.89	91.80±0.56	96.01±0.03	82.09±0.33	86.81±1.19	57.9±0.68	91.32±0.27	82.12±0.48	<b>97.12±0.01</b>	<b>97.50±0.05</b>
AA	79.22±0.00	92.16±1.32	85.32±0.19	73.06±3.50	90.01±2.64	94.72±0.59	76.05±0.01	87.96±1.22	49.99±1.34	88.19±0.68	81.03±0.51	<b>96.69±0.21</b>	<b>97.22±0.22</b>
K×100	75.44±0.00	90.89±1.64	83.95±0.16	76.90±2.64	89.04±0.75	94.71±0.04	76±0.31	82.18±1.54	44.44±0.95	88.45±0.35	76.93±0.52	<b>96.19±0.02</b>	<b>96.68±0.06</b>

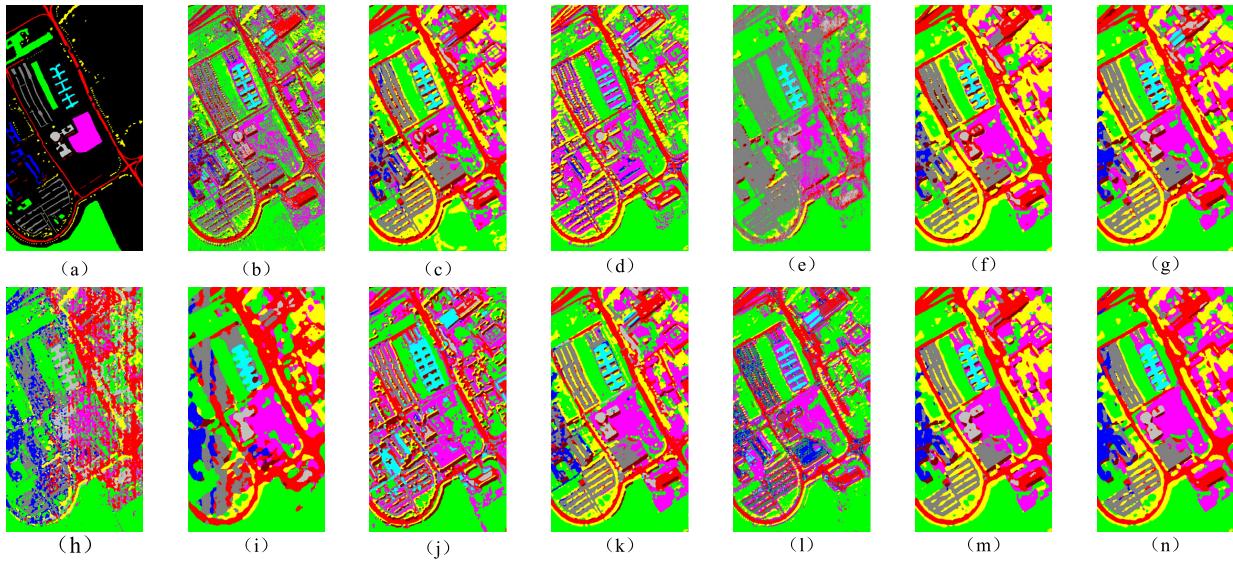


Fig. 15. Classification maps obtained by different methods on UP dataset (sample proportion is 0.5%). (a) Ground truth map. (b) SVM (82.06%). (c) SSRN (92.50%). (d) CDCNN (87.94%). (e) PyResNet (83.01%). (f) DBMA (91.80%). (g) DBDA (96.01%). (h) Hybird-SN (82.09%). (i) A<sup>2</sup>S<sup>2</sup>K-ResNet (86.81%). (j) DSSNet (57.90%). (k) 3-D-HyperGAMO (91.32%). (l) SpectralFormer (82.12%). (m) ECNet (97.12%). (n) FECNet (97.50%).

are 26.57% (SVM), 5.08% (SSRN), 30.47% (CDCNN), 9.68% (PyResNet), 7.38% (DBMA), 1.75% (DBDA), 13.15% (Hybrid-SN), 2.78% (A<sup>2</sup>S<sup>2</sup>K-ResNet), 47.01% (DSSNet), 5.80% (3-D-HyperGAMO), and 24.01% (SpectralFormer). Fig. 14 shows the visualization results. Due to the serious mixing between different categories of the original IP dataset, the classification results of many advanced comparison methods are not satisfied, such as Hybrid-SN, DSSNet, and SpectralFormer. Similarly, it can be seen from the classification diagram in Fig. 14 that there is more noise in the classification maps of SVM, CDCNN, PyResNet, and DSSNet. In addition, there are some wrong classifications in Hybrid-SN and SpectralFormer. However, the proposed methods ECNet and FECNet can obtain good classification results, as shown in Fig. 14(m) and (n).

2) *Classification Results of up Dataset:* Table II and Fig. 15 show the specific classification results. The UP dataset has nine categories, which is seven less than the IP dataset. The number of UP spectral bands is about half of the IP, but from the classification results, the UP dataset is easier to classify.

It can be seen from Table II that the classification results of DBMA and DBDA based on double branch structure on the UP dataset are better than those of SSRN, CDCNN, PyResNet, Hybrid-SN, and DSSNet. In addition, satisfactory results were obtained by A<sup>2</sup>S<sup>2</sup>K-ResNet based on adaptive attention and 3-D-HyperGAMO based on generative adversarial. However, the OA, AA, and Kappa results of the proposed ECNet and FECNet are higher than that of all the above comparison methods. Among the two methods of ECNet and FECNet, the three scores OA, AA, and Kappa of ECNet are slightly lower than those of FECNet. From the classification diagram, as shown in Fig. 15, the methods with more label classification errors are SVM, PyResNet, Hybird-SN, DSSNet, and SpectralFormer. The same conclusion can be obtained from the classification results in Table II. However, the classification results obtained by the proposed ECNet and FECNet are the best.

3) *Classification Results of KSC Dataset:* Table III and Fig. 16 show the classification results of all methods, respectively. It can be seen from Table III that compared with the classification results of UP dataset in Table II,

TABLE III  
CLASSIFICATION RESULTS OF KSC DATASET USING 5% TRAINING SAMPLES (%)

Class	SVM [7]	SSRN [36]	CDCNN [33]	PyResNet [37]	DBMA [53]	DBDA [55]	Hybrid-SN [42]	$A^2S^2K$ -ResNet [44]	DSSNet [48]	3D-HyperGAMO[52]	SpectralFormer[53]	ECNet	FECNet
C1	92.43±0.00	98.40±0.48	96.81±0.69	99.86±0.14	99.39±0.39	99.67±0.16	88.08±5.95	100±0.00	96.65±1.2	98.38±1.75	98.27±0.04	99.83±0.01	99.87±0.18
C2	87.14±0.00	94.52±1.92	83.65±1.41	92.93±7.05	93.80±2.36	96.58±0.43	76.94±3.25	99.13±0.39	82.12±2.41	90.57±5.50	70.1±0.74	98.53±0.75	99.53±0.66
C3	72.47±0.00	85.20±5.46	83.92±2.96	84.22±5.84	80.20±1.62	88.72±2.03	33.37±2.94	87.81±0.62	66.34±0.95	90.83±3.39	59.04±1.46	95.1±1.43	96.82±1.00
C4	54.45±0.00	74.55±2.39	58.61±1.53	44.63±15.75	75.31±1.08	80.82±0.59	71.36±7.40	95.53±0.02	69.34±6.01	89.55±1.32	62.64±7.04	97.01±0.48	98.34±2.35
C5	64.11±0.00	75.13±11.77	52.83±3.21	72.98±12.98	69.6±6.22	78.14±2.55	83.99±4.44	92.36±0.20	45.12±0.42	82.86±1.23	51.89±4.9	96.6±1.85	97.62±3.37
C6	65.23±0.00	94.35±0.72	77.17±0.29	89.91±10.33	95.06±3.41	97.75±1.82	73.62±12.16	99.92±0.11	64.94±11.38	96.64±0.78	80.68±0.83	99.9±0.07	99.98±0.04
C7	75.50±0.00	84.64±4.05	75.34±2.14	98.33±1.53	87.08±1.09	95.15±1.22	63.01±14.69	95.85±1.99	63.88±5.3	81.52±7.95	65.8±2.03	97.29±0.81	93.9±3.99
C8	87.33±0.00	96.97±1.44	85.83±0.11	94.30±7.86	95.4±1.88	99.08±0.76	76.35±7.53	99.41±0.60	76.6±2.81	98.67±0.56	82.46±0.28	99.94±0.05	99.95±0.07
C9	87.94±0.00	97.83±0.82	91.65±0.29	99.87±0.23	96.21±1.07	99.98±0.03	74.5±23.64	99.76±0.05	84.1±1.93	99.65±0.22	85.08±6.18	100±0.00	100±0.00
C10	96.01±1.73	98.84±1.00	93.87±0.09	97.05±3.76	96.13±1.85	99.92±0.07	80.07±3.30	100±0.00	83.01±6.87	99.62±0.35	89.12±0.93	100±0.00	100±0.00
C11	96.03±0.00	99.14±0.37	98.77±0.17	98.24±1.65	99.64±0.29	98.92±0.34	94.41±4.86	100±0.00	97.1±1.16	99.02±1.39	99.3±0.22	98.75±0.05	99.39±0.86
C12	93.75±0.01	99.17±0.28	94.08±1.85	99.37±0.63	98.19±0.04	98.95±0.18	71.55±0.20	99.64±0.11	87.78±1.56	99.70±0.23	91.83±0.15	99.35±0.21	99.42±0.11
C13	99.72±0.00	100±0.00	99.8±0.13	100±0.00	100±0.00	99.97±0.05	91.96±0.11	100±0.00	98.33±0.71	100±0.00	98.52±0.76	100±0.00	100±0.00
OA	87.96±0.00	94.52±0.90	89.33±0.65	93.97±2.44	94.12±0.27	96.76±0.51	79.72±4.31	98.34±0.46	83.4±0.58	96.90±0.95	86.01±0.06	<b>99.12±0.02</b>	<b>99.27±0.06</b>
AA	82.55±0.00	92.15±1.87	84.03±0.95	90.13±3.65	91.23±0.75	94.90±0.20	78.17±4.24	97.65±0.08	78.1±1.82	94.38±3.10	79.6±0.59	<b>98.65±0.07</b>	<b>98.45±0.21</b>
K $\times$ 100	86.59±0.00	93.9±1.00	88.13±0.73	93.29±2.71	93.45±0.31	96.40±0.57	77.34±4.70	98.84±0.46	81.51±0.66	96.55±1.06	84.41±0.06	<b>99.02±0.02</b>	<b>99.16±0.18</b>

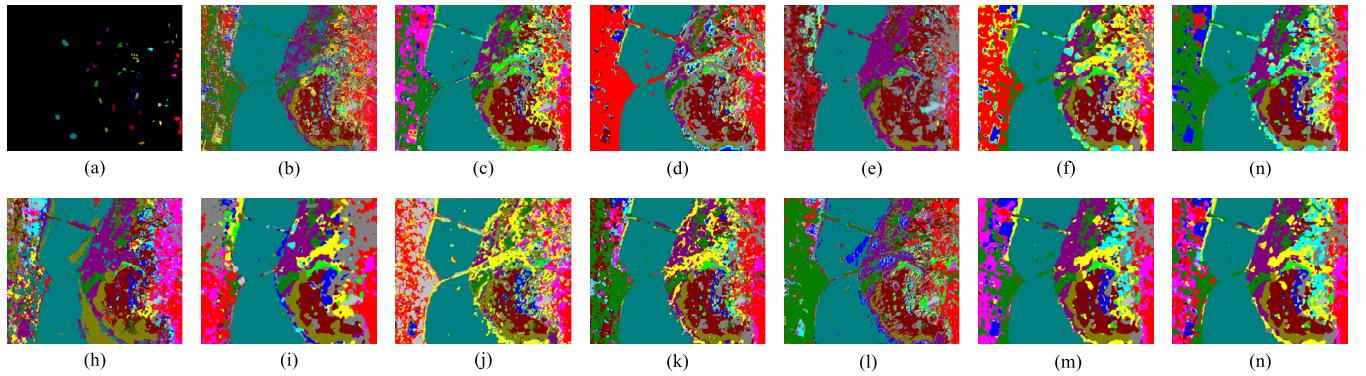


Fig. 16. Classification maps obtained by different methods on KSC dataset (sample proportion is 5%). (a) Ground truth map. (b) SVM (87.96%). (c) SSRN (94.52%). (d) CDCNN (89.33%). (e) PyResNet (93.97%). (f) DBMA (94.12%). (g) DBDA (96.76%). (h) Hybird-SN (79.72%). (i)  $A^2S^2K$ -ResNet (98.34%). (j) DSSNet (83.40%). (k) 3-D-HyperGAMO (96.90%). (l) SpectralFormer (86.01%). (m) ECNet (99.12%). (n) FECNet (99.27%).

the classification results of KSC dataset are greatly improved, which is due to less noise in KSC dataset and higher spatial resolution of KSC dataset than UP. In addition, it is worth noting that ECNet and FECNet still get the highest OA, AA, and Kappa scores. Among them, the OA, AA, and Kappa scores of FECNet are still slightly higher than ECNet. Compared with other comparison methods, the increases of OAs obtained by the proposed ECNet are 11.16% (SVM), 4.60% (SSRN), 9.79% (CDCNN), 5.15% (PyResNet), 5.00% (DBMA), 2.36% (DBDA), 19.40% (Hybrid-SN), 0.78% ( $A^2S^2K$ -ResNet), 15.72% (DSSNet), 2.22% (3-D-HyperGAMO), and 13.11% (SpectralFormer). The classification results of KSC datasets by different methods are shown in Fig. 16. Oak/broadleaf (C5) is a category that is difficult to classify in KSC dataset. For easier observation, enlarge some areas of the classification map, as shown in the red box mark in Fig. 16. It can be seen that the classification map of SVM, Hybrid-SN, DSSNet, and SpectralFormer contains more noise. Obviously, the classification map obtained by SSRN, DBMA, DBDA,  $A^2S^2K$ -ResNet, and 3-D-HyperGAMO is better than them. However, compared with these methods, the classification map of ECNet and FECNet is smoother.

**4) Classification Results of SV Dataset:** The classification results of each method are shown in Table IV and Fig. 17. It can be seen from Table IV that for Grapes-untrained (C8), our methods ECNet and FECNet can get better classification results, while other methods have poor classification results for this category, which shows that the proposed method can still effectively extract features for categories that are difficult to classify, which further shows that the proposed method has strong robustness. In addition, compared with other comparison methods, the increases of OAs obtained by the proposed ECNet are 10.43% (SVM), 5.37% (SSRN), 9.05% (CDCNN), 4.68% (PyResNet), 4.46% (DBMA), 3.67% (DBDA), 9.63% (Hybrid-SN), 2.26% ( $A^2S^2K$ -ResNet), 28.01% (DSSNet), 1.92% (3-D-HyperGAMO), and 10.25% (SpectralFormer). The visualization results of all methods are shown in Fig. 17. It can be seen that the classification maps of ECNet and FECNet are smoother than other classification maps, which proves that the method proposed in this article has more advantages in datasets with similar categories and regular land cover.

**5) Classification Results of HT Dataset:** Table V and Fig. 18 show the classification results of all methods. It can be seen from Table V that the proposed method can provide

TABLE IV  
CLASSIFICATION RESULTS OF SV DATASET USING 0.5% TRAINING SAMPLES (%)

Class	SVM [7]	SSRN [36]	CDCNN [33]	PyResNet [37]	DBMA [53]	DBDA [55]	Hybrid-SN [42]	$A^2S^2K$ -ResNet [44]	DSSNet [48]	3D-HyperGAMO[52]	SpectralFormer[53]	ECNet	FECNet
C1	99.42±0	97.18±2.47	96.74±3.04	88.79±17.63	98.52±2.52	99.73±0.23	90.72±8.56	99.99±0.02	70.72±2.38	97.22±3.80	93.52±2.47	100±0	100±0
C2	98.79±0	98.86±1.11	96.48±0.56	95.17±8.36	99.62±0.32	99.17±0.82	96.7±1.71	99.9±0.03	82.75±7	99.09±0.57	95.8±0.07	98.89±1.4	99.97±0.04
C3	87.98±0	94.25±2.13	89.53±1.78	85.99±18.68	96.81±0.54	97.47±0.31	81.44±9.43	94.95±3.28	79.87±0.52	94.91±1.51	91.91±3.27	94.89±2.6	98.01±0.06
C4	97.54±0	97.64±0.12	95.55±0.06	94.15±8.46	92.15±1.52	94.3±0.8	62.87±3.32	98.09±0.21	84.54±14.56	91.88±2.25	96.36±2.80	95.46±0.94	95.02±1.17
C5	95.06±0.06	97.26±1.5	96.08±2.51	99.13±0.79	96.74±0.94	98.14±1.23	92.72±1.86	98.6±0.1	76.84±8.39	98.01±0.33	88.81±2.02	99.55±0.64	97.89±2.18
C6	99.9±0	99.94±0.04	97.34±0.45	99.99±0.02	99.32±0.48	99.86±0.16	99.39±0.04	99.9±0.12	88.7±8.57	99.29±0.84	99.85±0.11	99.86±0.16	99.92±0.08
C7	95.6±0.01	99.34±0.35	92.89±4.01	99.63±0.64	97.68±0.65	98.32±0.27	98.06±1.26	99.97±0.04	75.26±1.06	99.55±0.19	91.98±7.9	99.69±0	99.95±0.04
C8	72.16±0.71	85.27±4.58	80.44±0.35	83.76±10.17	89.38±1.17	91.82±2.63	83.59±0.3	88.07±0.01	67.11±1.1	93.15±0.42	75.34±3.95	94.63±1.53	95.92±0.2
C9	98.08±0	99.38±0.12	98.59±0.11	99.6±0.34	99.15±0.25	99.07±0.07	98.46±0.71	99.9±0.01	89.53±10.51	99.21±0.42	98.23±0.81	99.63±0.13	99.89±0.01
C10	85.39±0	95.36±0.6	86.82±0.84	95.07±1.68	93.89±0.86	97.52±0.85	93.79±0.31	97.35±1.44	60.1±8.9	98.26±0.27	92.96±1.45	98.87±0.7	99.18±0.06
C11	86.98±0	95.81±0.26	82.65±2.27	88.65±10.85	93.62±0.85	95.74±0.26	50.68±5.42	97.33±0.42	39.73±5.15	91.74±4.89	51.99±2.03	96.54±0.65	97.37±0.3
C12	94.2±0	98±0.42	95.78±0.57	99.93±0.06	97.77±1.6	98.84±0.69	84.6±2.76	98.51±0.23	79.2±16.36	98.80±0.30	96.08±5.47	99.68±0.21	99.77±0.13
C13	93.43±0	98.23±1.07	96.88±0.44	99.16±1	98.27±0.91	99.49±0.23	41.48±4.08	97.77±2.49	74.94±28.83	78.37±1.08	85.29±1.88	99.98±0.03	99.88±0.01
C14	92.03±0	96.8±1.46	92.21±0.18	99.34±0.49	95.94±0.54	95.54±0.41	38.65±6.53	95.61±2.31	76.61±1.41	81.95±3.36	97.46±1.33	96.03±1.37	95.69±0.48
C15	71.02±0	82.34±3.5	72.84±1.73	87.93±5.54	83.02±1.06	83.22±4.71	85.96±0.16	88.44±0.74	64.84±7.78	90.67±3.53	78.84±4.36	94.86±3.08	94.97±0.57
C16	97.82±0	99.54±0.29	97.8±0.78	94.26±6.17	99.03±0.28	99.98±0.01	93.74±0.08	99.63±0.08	76.12±5.1	99.13±0.30	74.82±1.4	99.79±0.3	99.54±0.65
OA	86.98±0	92.04±0.96	88.36±0.28	92.73±1.9	92.95±0.33	93.74±0.74	87.78±0.62	95.15±0.31	69.4±2.61	95.49±1.71	87.16±0.94	<b>97.41±0.08</b>	<b>97.85±0.08</b>
AA	91.56±0	95.95±0.21	91.95±0	94.41±0.63	95.68±0.2	96.76±0.17	80.79±1.15	97.13±0.32	74.21±1.47	94.45±2.96	88.08±0.54	<b>98.03±0.25</b>	<b>98.31±0.04</b>
K×100	85.45±0	91.14±1.08	87.05±0.3	91.92±2.09	92.16±0.34	93.05±0.8	86.37±0.68	94.6±0.34	65.74±2.81	94.98±1.90	85.71±1.02	<b>97.11±0.08</b>	<b>97.61±0.09</b>

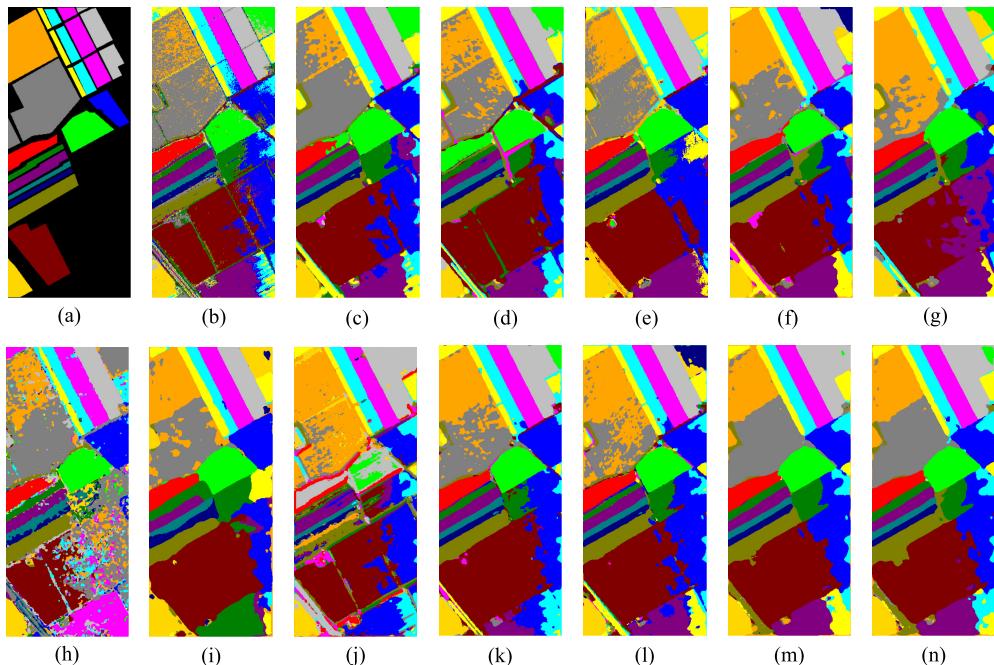


Fig. 17. Classification maps obtained by different methods on SV dataset (sample proportion is 0.5%). (a) Ground truth map. (b) SVM (86.98%). (c) SSRN (92.04%). (d) CDCNN (88.36%). (e) PyResNet (92.73%). (f) DBMA (92.95%). (g) DBDA (93.74%). (h) Hybird-SN (87.78%). (i)  $A^2S^2K$ -ResNet (95.15%). (j) DSSNet (69.40%). (k) 3-D-HyperGAMO (95.49%). (l) SpectralFormer (87.16%). (m) ECNet (97.41%). (n) FECNet (97.85%).

the highest OA, AA, and Kappa on HT dataset. Among them, compared with  $A^2S^2K$ -ResNet with the highest three indexes in other comparison methods, the OA, AA, and Kappa scores of FECNet are 1.25%, 0.37%, and 1.34% higher, respectively. Fig. 18 shows the visualization results of all methods. When dealing with objects with small sizes, spatial features alone may lead to over smoothing phenomenon or even misclassify the image, for example, CDCNN and DSSNet. However, the classification map of ECNet and FECNet is not only less misclassified but also maintains good boundaries for the objects.

In conclusion, the ECNet and FECNet methods proposed in this article get the best classification results on five datasets, which fully prove that the proposed method has strong generalization ability. In IP datasets with uneven land cover distribution and HT datasets with higher resolution, ECNet and FECNet can still obtain better classification accuracy than other methods. In addition, in the SV dataset with similar categories and the UP dataset with fewer bands, our method can realize high-precision classification of the categories that are difficult to distinguish.

TABLE V  
CLASSIFICATION RESULTS OF HT DATASET USING 2% TRAINING SAMPLES (%)

Class	SVM [7]	SSRN [36]	CDCNN [33]	PyResNet [37]	DBMA [53]	DBDA [55]	Hybrid-SN [42]	A <sup>2</sup> S <sup>2</sup> K-ResNet [44]	DSSNet [48]	3D-HyperGAMO[52]	SpectralFormer[53]	ECNet	FECNet
C1	92.96±0.86	86.44±6.14	68.71±14.61	87.92±1	88.51±2.26	89.61±1.7	88.07±2.87	90.72±2.43	68.26±2.72	87.76±3.89	87.96±1.81	91.67±1.54	92.17±0.7
C2	94.04±0.9	93.87±3.7	85.75±10.24	91.71±3.5	95.57±1.56	97.12±2.38	95.97±1.97	97.62±1.31	66.76±16.84	94.61±2.11	97.51±0.01	95.41±4.37	97.44±0.29
C3	99.65±0	99.8±0.29	71.68±12.81	98.02±1.97	100±0	100±0	97.79±0.47	99.63±0.1	61.25±7.76	95.95±5.73	99.63±0.26	99.87±0.16	99.75±0.1
C4	98.58±0	96.35±0.86	90.77±7.39	93.32±1.32	98.51±0.44	98.47±0.37	94.38±2.05	96.51±1.88	74.88±20.5	96.76±1.03	98.09±0.27	98.87±0.41	98.45±0.47
C5	91.41±0	94.5±0.91	95.26±0.55	91.87±1.01	96.58±1.76	97.74±0.01	94.88±1.28	95.99±0.28	60.92±15.16	96.34±1.13	98.33±1.1	96.71±0.51	97.61±0.64
C6	99.56±0	100±0	87.59±12.66	95.65±0.49	99.66±0.24	98.83±0.44	96.22±1.88	98.57±1.32	67.82±1.41	94.36±2.16	99.6±0.31	99.16±0.76	98.65±1.51
C7	75.97±0	80.62±1.82	70.86±2.05	78.25±1.56	85.32±1.57	87.2±2.24	87.03±0.33	93.46±0.18	67.69±2.17	88.69±1.96	89.38±3.2	87.75±2.57	87.84±1.62
C8	75.86±0	86.33±0.66	76.44±0.36	93.27±3.3	94.41±0.78	95.54±1.9	87.41±2.06	95.25±1.22	47.99±1.8	86.16±0.74	85.97±1.96	96.11±3.9	96.61±2.17
C9	73.68±0	91.02±0.2	73.96±3.54	73.53±4.19	85.83±0.37	86.86±0.34	81.96±0.47	87.48±0.65	56.8±15.58	85.63±0.01	89.22±1.62	88.9±2.36	89.05±0.06
C10	74.88±0	78.69±1.75	58.22±9.91	65.26±10.76	90.19±1.33	82.11±1.58	83.04±0.38	78.42±0.66	39.26±11.31	78.82±4.84	85.3±4.33	87.39±0.01	87.38±1.92
C11	76.63±0	84.48±0.33	67.85±0.4	65.56±7.57	86±0.68	93.95±2.81	87.89±5.56	90.87±1.77	54.41±0.42	91.74±2.62	88.27±3.55	94.71±1.1	94.59±1.46
C12	73.56±0	84.01±5.97	65.38±6.84	70.12±12.14	88.75±2.38	90.12±1.46	86±0.59	91.47±0.43	38.27±5.61	78.52±3.83	91.61±0.15	92.91±3.73	95.19±2.19
C13	53.28±0	88.35±0.54	84.17±0.16	93.03±8.68	85.89±0.84	90.66±0.13	93.33±1.41	92.03±2.42	68.94±8.65	93.17±2.51	88.53±2.63	83.16±5.62	81.4±1.1
C14	88.57±0	95.29±4.29	80.89±3.87	94.41±0.78	98.86±0.11	98.52±0.01	91.43±0.78	97±0.35	46.39±11.8	94.85±3.02	98.73±1.53	95.21±0.59	94.76±1.28
C15	99.19±0	96.76±0.31	87.82±0.17	94.83±4.04	95.91±0.15	96.15±0.14	96.59±1.79	98.27±1.03	39.23±6.84	97.02±0.33	95.05±0.25	97.44±0.01	97.81±0.4
OA	84.12±0	88.09±2.02	74.64±0.01	80.09±1.66	90.73±0.95	92.17±0.08	89.31±0.77	92.18±0.74	52.61±9.96	88.48±1.05	91.44±0.45	92.9±0.16	93.43±0.21
AA	84.52±0	90.43±1.15	77.69±0.96	85.79±0.54	92.33±0.65	93.53±0.01	90.87±0.77	93.55±0.41	57.26±8.1	90.69±1.07	92.84±0.41	93.68±0.3	93.92±0.09
K×100	82.81±0	87.12±2.18	72.76±0.25	78.45±1.8	89.98±1.03	91.53±0.09	88.43±0.83	91.55±0.8	48.75±10.78	87.54±1.15	90.75±0.49	92.32±0.17	92.89±0.22

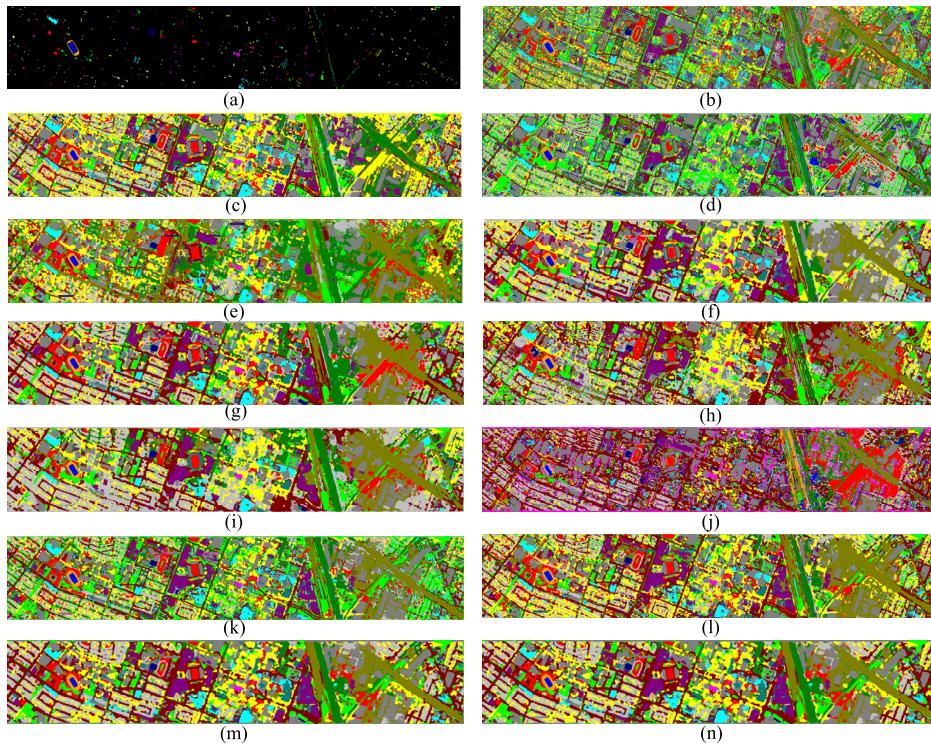


Fig. 18. Classification maps obtained by different methods on HT dataset (sample proportion is 2%). (a) Ground truth map. (b) SVM (84.12%). (c) SSRN (88.09%). (d) CDCNN (74.64%). (e) PyResNet (80.09%). (f) DBMA (90.73%). (g) DBDA (92.17%). (h) Hybrid-SN (89.31%). (i) A<sup>2</sup>S<sup>2</sup>K-ResNet (92.18%). (j) DSSNet (52.61%). (k) 3-D-HyperGAMO (88.48%). (l) SpectralFormer (91.44%). (m) ECNet (92.90%). (n) FECNet (93.43%).

#### D. Analysis of the Proposed Method

In this section, we will analyze the parameters of the proposed method, and the model is suitable for the selection of parameters.

1) *Contribution of Each Block of the Proposed Method:* In order to verify the effectiveness of each block in the method proposed in this article, some ablation experiments are carried out on each block, and the results are shown in Table VI. It can be seen that compared with standard CNN, the proposed ECB can provide higher OA score, which shows that it is

TABLE VI  
COMPARISON OF OA SCORES OF DIFFERENT BLOCKS (%)

component	IP	UP	KSC	SV	HT
Standard CNN	94.92	96.76	98.63	96.85	90.91
Standard CNN + SAB	95.23	97.03	99.05	97.14	91.22
ECB+SAB	95.33	97.12	99.12	97.41	92.90
SFB+ECB+SAB	95.81	97.50	99.27	97.85	93.43

more effective to expand the RF of spectral feature extraction through expansion convolution. In addition, we add the SFB block on the basis of ECB, which can feedback the deep

TABLE VII  
RUNNING TIME (S), PARAMETERS REQUIRED, AND MEMORY REQUIREMENT (MB) BY ALL THE METHODS COMPARED WITH THE PROPOSED METHOD

method	IP		UP		KSC		SV		HT		Memory
	Time	Params	Time	Params	Time	Params	Time	Params	Time	Params	
SSRN	106	364K	71	216K	73	327K	129	370K	76	278K	10.27
CDCNN	<b>24</b>	1.1M	<b>42</b>	628K	<b>41</b>	955K	<b>34</b>	1.1M	<b>20</b>	812K	5.81
PyResNet	56	<b>22.4M</b>	61	<b>22.1M</b>	63	<b>22.3M</b>	65	<b>21.7M</b>	73	<b>22.2M</b>	85.63
DBMA	222	609K	96	324K	174	539K	230	621K	253	447K	37.13
DBDA	194	382K	93	202K	129	338K	225	389K	107	280K	25.89
Hybrid-SN	37	403K	71	1.7M	45	3.1M	112	3.6M	33	2.5M	11.83
A <sup>2</sup> S <sup>2</sup> K-ResNet	40	373K	<b>182</b>	222K	296	335K	72	<b>83K</b>	154	285K	12.02
DSSNet	303	366K	125	363K	140	365K	330	367K	108	366K	<b>90.47</b>
3D-HyperGAMO	458	1.1M	168	1.1M	256	1.1M	596	1.1M	189	1.1M	4.18
SpectralFormer	<b>1511</b>	351K	<b>2513</b>	167K	<b>656</b>	295K	<b>4545</b>	361K	<b>1702</b>	231K	<b>1.34</b>
ECNet	121	<b>308K</b>	58	<b>165K</b>	97	<b>272K</b>	118	314K	162	<b>224K</b>	50.37
FECNet	145	<b>317K</b>	59	<b>174K</b>	127	<b>281K</b>	132	323K	197	<b>234K</b>	50.42

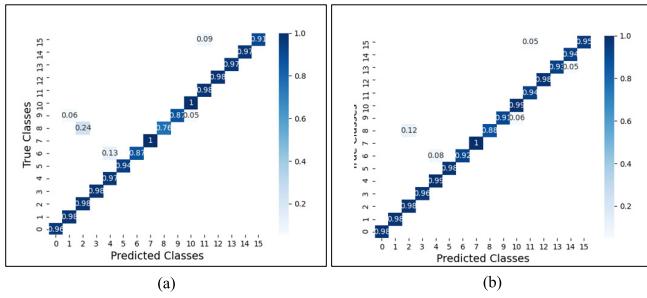


Fig. 19. Visualization results of the two blocks. (a) Confusion matrix obtained using standard CNN. (b) Confusion matrix obtained using ECB.

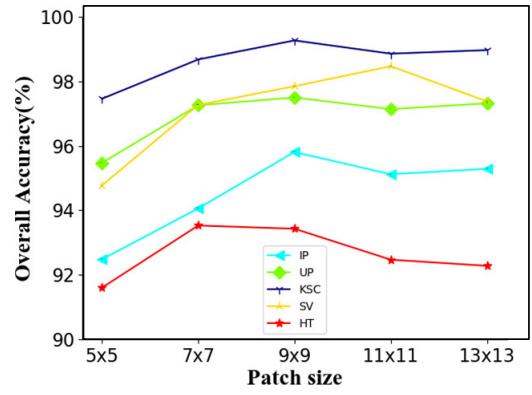


Fig. 20. Comparing different input space sizes of FECNet.

features to the shallow features, so that the shallow features can adjust themselves. Experimental results show that SFB can further improve the classification performance. For datasets IP and HT, when the SFB block is added to the network, the classification effect is improved more obviously.

In order to further verify the effectiveness of ECB, the visualization results of the proposed FECNet using standard CNN and ECB are shown in Fig. 19(a) and (b), respectively. The ordinate represents the real category label, the abscissa represents the predicted category label, and the diagonal represents the probability that the predicted category label is the same as the real category label. Compared with the results in Fig. 19(b), it is obvious that there are more misclassification results in Fig. 19(a), which proves that the effectiveness of the proposed network with ECB is better than that with standard CNN.

2) *Comparison of Different Input Sizes:* The setting of parameters can affect the classification performance, and the input space size has a great impact on the classification performance of the network. In order to further verify the impact of different input space sizes on performance, the space input with the size of  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ ,  $11 \times 11$ , and  $13 \times 13$  was carried out in the experiment. Fig. 19 shows the comparison results of the different input space sizes of FECNet. It can be seen that from Fig. 20: first, when the input space size is small, the OA score obtained by training is relatively low. Second, for IP, UP, and KSC datasets, after OA reach better performance,

the OA score changes little with the increase of space size. For SV and HT datasets, OA increased first and then decreased. Third, for the datasets IP, UP, and KSC, the space size is  $9 \times 9$  to achieve the best performance. Although the optimal classification performance was not obtained at  $9 \times 9$  on SV and HT, the suboptimal classification performance was also obtained. Therefore, the input space size of all datasets in this experiment is  $9 \times 9$ .

3) *Comparison of Running Time, Parameters, and Memory Requirement of Different Methods:* Table VII shows the comparison results of running time, parameter quantity, and memory requirement of all methods. Because PyResNet adds additional links to CNN and gradually increases the dimension of feature map in all Conv layers, the parameters required for training the model are large. ECNet and FECNet build models based on expansion convolution, so that the network requires relatively few parameters. Similarly, although the required training parameters and running time of DBMA and DBDA are similar to those of the proposed method, the performance of ECNet and FECNet is better. The parameters of DSSNet based on expansion convolution are similar to ECNet and FECNet, and the training time is long, especially for IP and SV datasets. It is worth noting that SpectralFormer based on transformer framework needs a long running time due to its slow convergence speed. In addition, the memory required

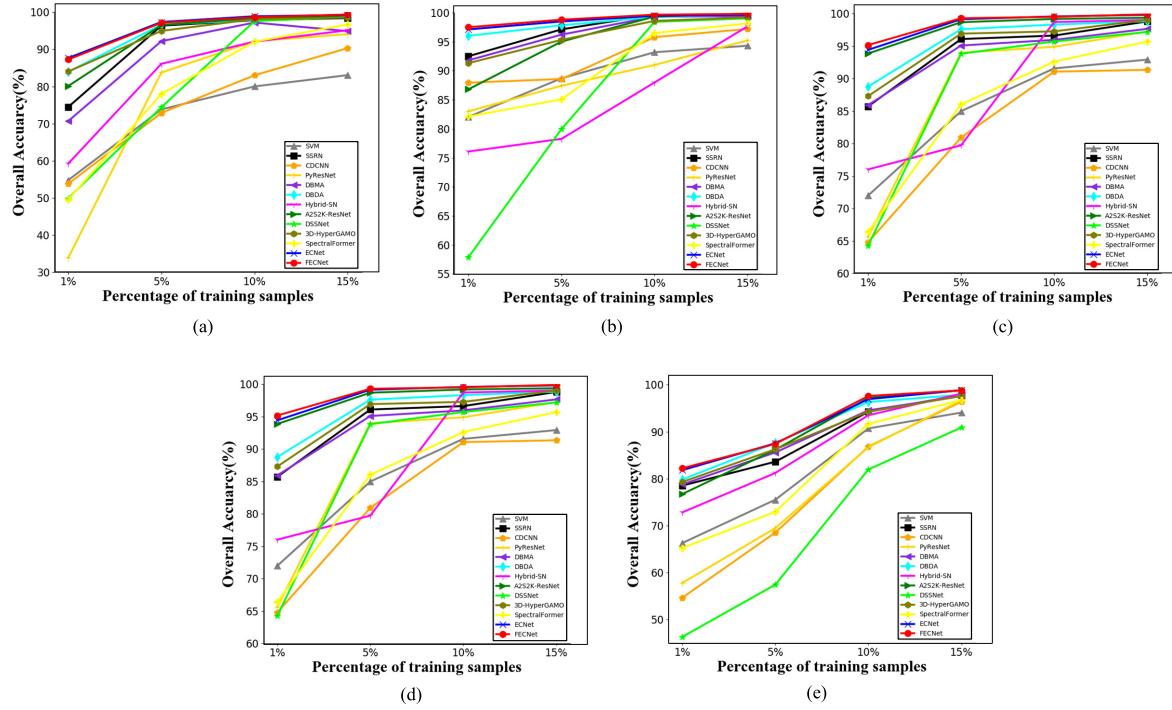


Fig. 21. Comparison of OAs of all methods with different training proportions. (a) OA comparison on IP datasets. (b) OA comparison on UP dataset. (c) OA comparison on KSC dataset. (d) OA comparison on SV dataset. (e) OA comparison on HT dataset.

by the proposed method is relatively moderate. Among all methods, SpectralFormer based on transformer architecture requires the least memory, while PyResnet and DSSNet require more memory. In conclusion, compared with other methods, ECNet and FECNet require less parameters, moderate running time, and memory on five datasets.

**4) Comparison of OAs of Different Methods With Different Training Proportions:** The comparison results of OAs of all methods under different training proportions on five datasets are shown in Fig. 21. Here, the ordinate represents the OA and the abscissa represents the proportion of training samples. As can be seen from Fig. 20, the OAs of ECNet and FECNet with all training proportions are higher than that of other methods, which shows that the proposed method can not only achieve good classification in small samples but also maintain the good classification performance in large samples, which further proves the effectiveness of the proposed method.

#### IV. CONCLUSION

In this article, we propose an ECNet based on expansion convolution, which expands the RF through the superposition of expansion convolution layers, so as to obtain more discriminative features. This method not only alleviates the computational burden caused by the increase of RF but also avoids the overfitting phenomenon under small training samples. In addition, when the equivalent convolution kernel of the expanded convolution is the same as that of the standard convolution, the expanded convolution module not only requires fewer parameters than the standard convolution module but also provides higher classification accuracy. In order

to further enhance the propagation of features, we introduce SFB based on ECNet. The experimental results show that the addition of SFB can significantly improve the classification performance and fully prove the effectiveness of the block in hyperspectral image classification. The proposed ECNet and FECNet are tested on four classical datasets and a higher resolution dataset (HT) and compared with a variety of the latest classification methods. The experimental results show that ECNet and FECNet can provide the best classification performance and have strong generalization ability. Although the proposed method achieves higher classification accuracy than other state-of-the-art methods on five public datasets, there are still some challenges. For example, for IP datasets with more categories and complex ground distribution and HT datasets with higher resolution, the classification accuracy of the method based on expansion convolution under limited training samples still needs to be further improved. In the future work, we will consider to combine the expansion convolution block of space and spectrum to extract hyperspectral image features more effectively, so as to further improve the classification performance of hyperspectral images.

#### ACKNOWLEDGMENT

The authors would like to thank the editor-in-chief, an associate editor, and the reviewers for their insightful comments and suggestion. In addition, the authors would like to thank the Hyperspectral Image Analysis Laboratory, University of Houston, and the IEEE GRSS Image Analysis and Data Fusion Technical Committee for providing the University of Houston dataset.

## REFERENCES

- [1] L. Liang *et al.*, "Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method," *Remote Sens. Environ.*, vol. 165, pp. 123–134, Aug. 2015.
- [2] X. Yang and Y. Yu, "Estimating soil salinity under various moisture conditions: An experimental study," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2525–2533, May 2017.
- [3] N. Yokoya, J. C.-W. Chan, and K. Segl, "Potential of resolution-enhanced hyperspectral data for mineral mapping using simulated EnMAP and sentinel-2 images," *Remote Sens.*, vol. 8, no. 3, pp. 172–189, Feb. 2016.
- [4] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018.
- [5] S. Zhang, J. Li, Z. Wu, and A. Plaza, "Spatial discontinuity-weighted sparse unmixing of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5767–5779, Oct. 2018.
- [6] P. Ghamisi *et al.*, "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.
- [7] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [8] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
- [9] J. Li, J. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.
- [10] B. Du and L. Zhang, "Random-selection-based anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1578–1589, May 2011.
- [11] B. Du and L. Zhang, "Target detection based on a dynamic subspace," *Pattern Recognit.*, vol. 47, no. 1, pp. 344–358, Jan. 2014.
- [12] G. Licciardi, P. R. Marpu, J. Chanussot, and J. A. Benediktsson, "Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 447–451, May 2012.
- [13] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.
- [14] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.
- [15] P. Ghamisi *et al.*, "New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 3, pp. 10–43, Sep. 2018.
- [16] 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.
- [17] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [18] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
- [19] M. Faufel, J. Chanussot, and J. A. Benediktsson, "A spatial-spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognit.*, vol. 45, no. 1, pp. 381–392, Jan. 2012.
- [20] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
- [21] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.
- [22] L. Fang, C. Wang, S. Li, and J. A. Benediktsson, "Hyperspectral image classification via multiple-feature-based adaptive sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1646–1657, Jul. 2017.
- [23] S. Li, T. Lu, L. Fang, X. Jia, and J. A. Benediktsson, "Probabilistic fusion of pixel-level and superpixel-level hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7416–7430, Dec. 2016.
- [24] T. Lu, S. Li, L. Fang, X. Jia, and J. A. Benediktsson, "From subpixel to superpixel: A novel fusion framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4398–4411, Aug. 2017.
- [25] L. Fang, N. He, S. Li, P. Ghamisi, and J. A. Benediktsson, "Extinction profiles fusion for hyperspectral images classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1803–1815, Mar. 2018.
- [26] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. 1, pp. 110–122, Sep. 2009.
- [27] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, "Joint learning of words and meaning representations for open-text semantic parsing," in *Proc. Int. Conf. Art. Intell. Stat.*, 2012, pp. 127–135.
- [28] B. Rasti *et al.*, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Dec. 2020.
- [29] X. Zhang, Y. Liang, C. Li, N. Huyan, L. Jiao, and H. Zhou, "Recursive autoencoders-based unsupervised feature learning for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 1928–1932, Nov. 2017.
- [30] T. Li, J. Zhang, and Y. Zhang, "Classification of hyperspectral image based on deep belief networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 5132–5136.
- [31] B. Pan, Z. Shi, and X. Xu, "R-VCANet: A new deep-learning-based hyperspectral image classification method," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1975–1986, May 2017.
- [32] X. Cao, F. Zhou, L. Xu, D. Meng, Z. Xu, and J. Paisley, "Hyperspectral image classification with Markov random fields and a convolutional neural network," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2354–2367, May 2018.
- [33] H. Lee and H. Kwon, "Going deeper with contextual CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4843–4855, Oct. 2017.
- [34] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.
- [35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [36] 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.
- [37] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral-spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 740–754, Feb. 2019.
- [38] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269.
- [39] P. Duan, X. Kang, S. Li, and P. Ghamisi, "Noise-robust hyperspectral image classification via multi-scale total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1948–1962, Jun. 2019.
- [40] S. Fang, D. Quan, S. Wang, L. Zhang, and L. Zhou, "A two-branch network with semi-supervised learning for hyperspectral classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 3860–3863.
- [41] B.-S. Liu and W.-L. Zhang, "Multi-scale convolutional neural networks aggregation for hyperspectral images classification," in *Proc. 13th Symp. Piezoelectricity, Acoustic Waves Device Appl. (SPAWDA)*, Jan. 2019, pp. 1–6.
- [42] 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, Feb. 2020.
- [43] Z. Meng, L. Jiao, M. Liang, and F. Zhao, "Hyperspectral image classification with mixed link networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2494–2507, 2021.

- [44] S. K. Roy *et al.*, “Attention-based adaptive spectral-spatial kernel ResNet for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 1–13, Dec. 2020.
- [45] H. Zhang, Y. Liu, B. Fang, Y. Li, L. Liu, and I. Reid, “Hyperspectral classification based on 3D asymmetric inception network with data fusion transfer learning,” 2020, *arXiv:2002.04227*.
- [46] R. Li and C. Duan, “LiteDenseNet: A lightweight network for hyperspectral image classification,” 2020, *arXiv:2004.08112*.
- [47] B. Cui, X.-M. Dong, Q. Zhan, J. Peng, and W. Sun, “LiteDepthwiseNet: A lightweight network for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.
- [48] X. Ma, A. Fu, J. Wang, H. Wang, and B. Yin, “Hyperspectral image classification based on deep deconvolution network with skip architecture,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4781–4791, Aug. 2018.
- [49] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” 2015, *arXiv:1511.07122*.
- [50] B. Pan, X. Xu, Z. Shi, N. Zhang, H. Luo, and X. Lan, “DSSNet: A simple dilated semantic segmentation network for hyperspectral imagery classification,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 11, pp. 1968–1972, Nov. 2020.
- [51] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, “Generative adversarial networks for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5046–5063, Sep. 2018.
- [52] S. K. Roy, J. M. Haut, M. E. Paoletti, S. R. Dubey, and A. Plaza, “Generative adversarial minority oversampling for spectral-spatial hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5500615, doi: [10.1109/TGRS.2021.3052048](https://doi.org/10.1109/TGRS.2021.3052048).
- [53] D. Hong *et al.*, “SpectralFormer: Rethinking hyperspectral image classification with transformers,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615, doi: [10.1109/TGRS.2021.3130716](https://doi.org/10.1109/TGRS.2021.3130716).
- [54] A. Vaswani *et al.*, “Attention is all you need,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
- [55] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
- [56] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, “ECA-Net: Efficient channel attention for deep convolutional neural networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11534–11542.
- [57] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional module attention module,” in *Proc. Eur. Conf. Comput. Vis.*, 2018, p. 17.
- [58] W. Ma, Q. Yang, Y. Wu, W. Zhao, and X. Zhang, “Double-branch multi-attention mechanism network for hyperspectral image classification,” *Remote Sens.*, vol. 11, no. 11, p. 1307, Jun. 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/11/1307>.
- [59] J. Fu *et al.*, “Dual attention network for scene segmentation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3146–3154.
- [60] 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. [Online]. Available: <https://www.mdpi.com/2072-4292/12/3/582>.
- [61] Y. Cui, Z. Yu, J. Han, S. Gao, and L. Wang, “Dual-triple attention network for hyperspectral image classification using limited training samples,” *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [62] D. Erhan, A. Courville, Y. Bengio, and P. Vincent, “Why does unsupervised pre-training help deep learning?” *J. Mach. Learn. Res.*, vol. 11, pp. 625–660, Feb. 2010.
- [63] M. Z. Alom *et al.*, “A state-of-the-art survey on deep learning theory and architectures,” *Electron.*, vol. 8, no. 3, p. 292, Mar. 2019.



**Cuiping Shi** (Member, IEEE) received the M.S. degree from Yangzhou University, Yangzhou, China, in 2007, and the Ph.D. degree from the Harbin Institute of Technology ( HIT ), Harbin, China, in 2016.

From 2017 to 2020, she held a Post-Doctoral Research with the College of Information and Communications Engineering, Harbin Engineering University, Harbin. She is currently a Professor with the Department of Communication Engineering, Qiqihar University, Qiqihar, China. She has published two academic books about remote sensing more than 60 articles in journals and conference proceedings. Her main research interests include remote sensing image processing, pattern recognition, and machine learning.

Dr. Shi’s doctoral dissertation won the Nomination Award of Excellent Doctoral Dissertation of HIT in 2016.



**Diling Liao** received the bachelor’s degree from the Zhuhai College of Jilin University, Zhuhai, China, in 2019. He is currently pursuing the master’s degree with Qiqihar University, Qiqihar, China.

His research interests include hyperspectral image processing and machine learning.



**Tianyu Zhang** received the bachelor’s degree from Qufu Normal University, Qufu, China, in 2019. She is currently pursuing the master’s degree with Qiqihar University, Qiqihar, China.

Her research interests include hyperspectral image processing and machine learning.



**Liguo Wang** (Member, IEEE) received the M.S. and Ph.D. degrees in signal and information processing from the Harbin Institute of Technology, Harbin, China, in 2002 and 2005, respectively.

From 2006 to 2008, he held a Post-Doctoral Research Position with the College of Information and Communications Engineering, Harbin Engineering University, Harbin, where he is currently a Professor. From 2020, he worked with the College of Information and Communication Engineering, Dalian Nationalities University, Dalian, China.

He has published two books about hyperspectral image processing and more than 130 articles in journals and conference proceedings. His main research interests include remote sensing image processing and machine learning.