Sandwich Convolutional Neural Network for Hyperspectral Image Classification Using Spectral Feature Enhancement

Hongmin Gao, Zhonghao Chen, and Chenming Li

***Abstract*—Recently, convolutional neural networks (CNNs) have been used to extract spectral and spatial features of hyperspectral images (HSIs) for hyperspectral image classification (HSIC) because of their excellent performance in extracting and analyzing complex data. However, due to the limited labeled samples and existing mixed pixels, it is difficult to extract features effectively, which further leads to the problem of overfitting of the model. On the other hand, to improve the extraction ability of the CNN, the depth of the model, and the complexity of the convolution kernel often need to be increased. In this article, a sandwich CNN based on spectral feature enhancement (SFE-SCNN) is proposed for HSIC. The proposed method, SFE-SCNN, introduces the spectral feature enhancement operation, which makes the data reflect more discriminative spectral feature details to suppress the interference of mixed pixels. Furthermore, according to the preprocessed data structure features, a lightweight sandwich convolution neural network is proposed. To fully extract the spectral features, the spectral feature re-extraction operation is used for the first time. Experimental results on three real hyperspectral datasets demonstrate that the proposed method achieves better classification performance than other state-of-the-art methods.**

***Index Terms*—Convolutional neural network, hyperspectral image (HSI) classification, lightweight sandwich convolution neural network, spectral feature enhancement, spectral feature re-extraction.**

I. Introduction

HYPERSPECTRAL images (HSIs) possess hundreds of continuous spectral bands so that it contains rich spectral information while containing spatial information [1], [2]. It is precisely because of these characteristics that HSIs have been widely valued in environmental monitoring, agriculture, military, and other fields. Especially, the research of HSI classification (HSIC) technology has been a hot topic in the field of HSI processing [3].

In general, HSIC is based on the rich detail features of the HSIs in the spectral domain and the spatial characteristics of ground objects to achieve the classification of each pixel in the images. In the early stage of the development of hyperspectral classification technology, HSIs have attracted the attention of a large number of researchers due to the rich features reflected by their high spectral resolution. Therefore, many HSI classifiers have been proposed, which are designed by analyzing original spectral features. Such as multinomial logistic expression [4], *k*-nearest neighbor [5], weighted Markov random Fields [6], and support vector machine (SVM) [7]. Then, in order to further improve the classification performance of traditional SVM, the subspace-based support vector machines (SVMsub) [8] were proposed. Experiments have proved that by analyzing the spectral information of HSIs, these methods have achieved some performance in classification. However, since the phenomenon of the same material may present spectral dissimilarity and different materials may have indistinguishable spectral features, it is very difficult to realize the accurate classification of objects only by using the original spectral features [9]. In addition, the lack of consideration of the spatial correlation of individual pixels is also a significant defect in traditional classification algorithms [10].

In the last decades, a series of targeted methods have been proposed to address the shortcomings of the aforementioned traditional algorithms. Especially, for the feature redundancy and Hughes phenomenon caused by the higher spectral dimension of HSIs, a series of dimensionality reduction and band selection methods have been proposed [11]–[13]. In addition to effectively solve the problems caused by high-dimensional data, these methods can also increase the difference between classes. To further improve the classification performance, some classification methods based on joint extraction of spectral-spatial features have been proposed gradually [14]–[18]. For instance, due to the spectral-spatial information of HSIs in a neighboring region can be coalesced into a sparse model, sparse representation [14] method was introduced to improve the classification performance. In [15], multiple morphological operations were applied for assembling spectral-spatial features of HSIs. Multiple kernel learning [16] based on spectral-spatial information also demonstrates to be a potent tool for improving the SVM classifier. Zhang *et al.* [17] proposed a new feature selection and extraction algorithm which can map spectral-spatial features into a common feature space, so as to find a low-dimensional representation of joint features in the spectral and spatial domain. Furthermore, Liu *et al.* [18] introduced a kernel

Manuscript received January 4, 2021; revised February 8, 2021; accepted February 24, 2021. Date of publication March 2, 2021; date of current version March 19, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 62071168 and Grant 61701166, in part by the National Key R&D Program of China under Grant 2018YFC1508106, and in part by the Fundamental Research Funds for the Central Universities of China under Grant B200202183. *(Corresponding author: Chenming Li.)*

The authors are with the College of Computer and Information, Hohai University, Nanjing 211100, China (e-mail: gaohongmin@hhu.edu.cn; chenzhonghao@hhu.edu.cn; lcm@hhu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3062872

low-rank representation method, where the similarity of local features was used to improve the performance of classification.

In recent years, deep-learning (DL) based methods, especially deep convolution neural networks (CNN), have shown unique superiority in feature extraction, which makes them widely studied in many fields (e.g., image classification [19], [20], natural language processing [21], target location [22], medical diagnosis [23]). Compared with the hand-crafting features extracted by traditional methods, deep learning provides an end-to-end solution to automatically extract deep features. Different from shallow machine learning methods, the deep network structure of deep learning has better feature mining performance. Therefore, many researchers have applied CNN to HSIC and demonstrated that CNN can show exceedingly promising performance [24]–[36]. For example, in [24], the spatial and spectral features of the original HSI are simultaneously extracted by 3D-CNN. Then, to make full use of spatial information to improve the recognition of the target spatial structure, Zhang *et al.* [25] proposed diverse region-based neural networks, and Sun *et al.* [26] introduced an attention mechanism in neural networks. However, overattention to the boundary of the ground object may cause classification noise within the same category in the face of low spatial resolution and mixed pixels. In [27] and [28], the residual structure was applied to the feature extraction network to alleviate the vanishing gradient caused by the network over depth. Inspired by the residual structure, Zhang e*t al.* [29] and Yang *et al.* [30] proposed the use of densely connected networks to improve the model's utilization of features. Recently, to ensure that spectral and spatial features are extracted while maintaining a small scale of the model, Roy *et al.* [31] proposed a hybrid model for feature extraction, which is composed of 2D-CNN and 3D-CNN. More recently, Li *et al.* [32] proposed a new dual-channel spectral-spatial features extraction fusion network, through which the local and global information of the target pixel points can be extracted and fused. In addition, in order to mine the contribution of the spatial position of the objects in remote sensing images to the classification task, graph convolutional network [33] and global spatial based [34] classification method were proposed. Then, invariant attribute profiles [35] are proposed to improve the recognition capacity of the same object in different scenes or locations. Furthermore, Hong *et al.* [36] proposed a multimodal deep learning framework, which can effectively improve the classification accuracy and classification refinement of HSIs. In general, the current task of HSIC faces the following two challenges: 1) The credibility of original spectral information is damaged by mixed pixels; and 2) Overcomplicated models more often than not have poor computational efficiency.

To address the above issues, in this article, a novel 3-D data preprocessing method based on HSI data structure features is proposed, and a new network structure based on the preprocessed data features is designed, which is called sandwich convolutional neural network (SCNN). Different from the traditional way of data augmentation methods (i.e., flip, rotation, and noise) applied in the daily images, due to the 3-D characteristics of HSI, the spectral information can be structurally exchanged with the spatial information dimensions to realize spectral feature enhancement (SFE). Then, being motivated by the procedure of pretreatment, higher spectral discriminative features in HSIs can be effectively extracted by SCNN with only fewer parameters.

The main contributions are summarized as follows.

1) To reduce the limitation of mixed pixels on classification performance, the data preprocessing method of SFE is proposed, which can enhance the recognition of spectral features of heterogeneous ground objects.
2) A novel SCNN is proposed to extract the spectral-spatial features from processed HSI cubes. It uses a series of small convolution kernels of different sizes to achieve multiscale extraction of 3-D features, which greatly reduces the number of parameters. Therefore, the model can achieve a faster classification.
3) To extract more useful abstract features, multiscale pointwise convolution is used in spatial feature extraction, and then the spectral feature re-extraction mechanism is used for the first time to extract more abstract features after spatial feature extraction and fusion.

The rest of this article is summarized as follows. Section II describes the proposed method. Section III analyzes the experimental results. Finally, the conclusion is drawn in Section IV.

II. Proposed Method

Fig. 1 shows the architecture of the proposed HSIC model. First, principal component analysis (PCA) will be applied to the original hyperspectral data to extract the major features from the spectral dimension to reduce its dimension. Second, the neighborhood data cubes of the target pixels are acquired and the SFE operation is applied to them. Third, the proposed SCNN model is used to extract the features of the preprocessed training set to train the model parameters. Finally, the unlabeled test pixels are predicted by the model and classified by the softmax function.

A. Data Preprocessing for Spectral Feature Enhancement

In general, the spectral-spatial information of HSI can be extracted simultaneously by using the neighborhood pixel block centered on the target pixel as the input of the neural network. However, for some images with low original spatial resolution and the deterioration of spatial resolution caused by pooling and strides convolution operation [37], the actual improvement of spatial information on the classification effect is lower than the theoretical level. In contrast, the spectral information of HSI is quite rich. Therefore, to further improve the classification performance of HSIC, it is necessary to enhance the mining of HSI spectral information and reduce the damage of spatial resolution. Inspired by Gao *et al.* [38], a method of SFE is proposed, which not only enhances spectral feature but also preserves spatial information. Let $\mathrm{C} \in \mathrm{R}^{P\times Q\times B}$ represent the original HSI cube, where P and Q represent the spatial dimension of the data cube, and B represents the number of bands of HSI. Furthermore, $\mathrm{c}^n \in R^{1\times B}$ represents the information of each pixel and $\mathrm{X}^n \in R^{W\times W\times B}$ represents the neighborhood cube centered on, where W and $n \in (1, \ldots, N)$ represent the patch size of neighborhood cubes and the number of samples.

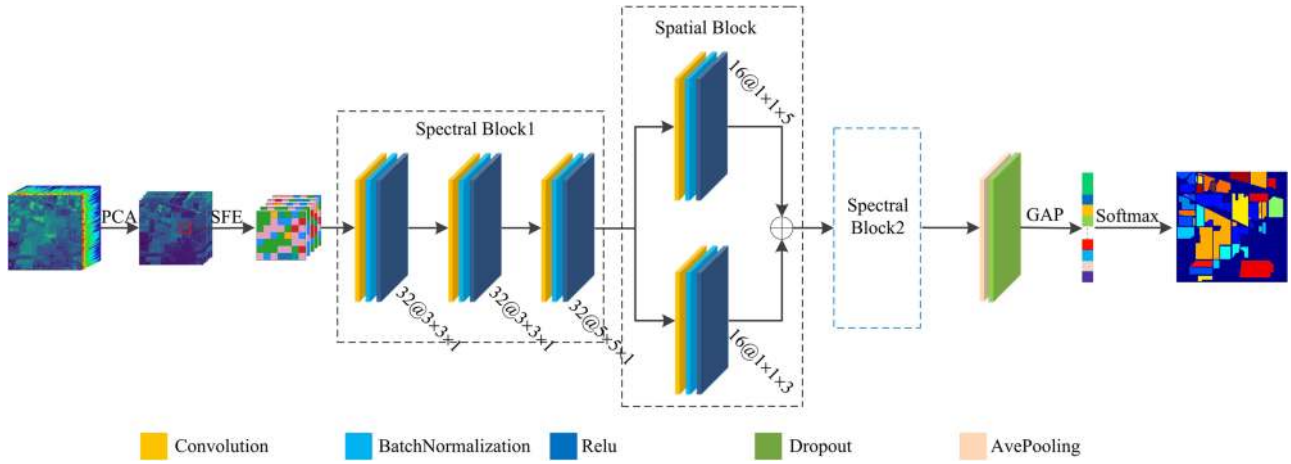

Fig. 1. Schematic of the proposed SFE-SCNN. "SFE" denotes spectral feature enhancement, and "+"denotes concatenation. "GAP" refers to global average pooling operation. For convenience, the structure of Spectral Block2 is omitted, which is the same as the structure of Spectral Block.

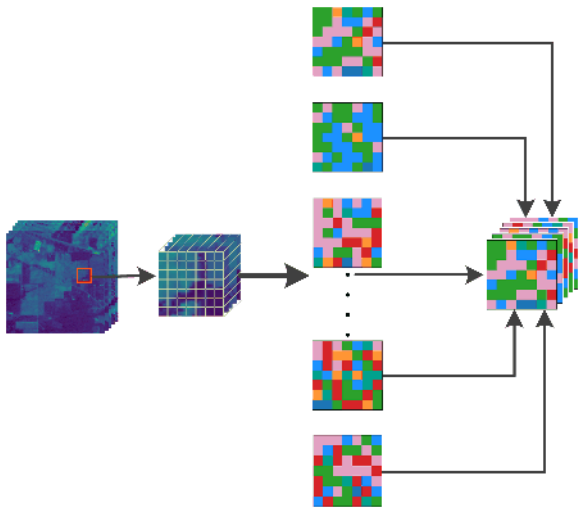

Fig. 2. Illustration of the SEF operation.

Fig. 2 shows the flowchart of the SFE operation. First of all, PCA will be applied to the original HSI to extract main components in the spectral dimension and realize dimension reduction to obtain $\mathrm{C}' \in R^{P\times Q\times R}$, where M represents the number of principal components. To meet the needs of spectral feature enhancement, M is the square number. It can be represented as follows:

$$M = \mathrm{W}^2 \tag{1}$$

where M is the number of main components. Second, obtaining the neighborhood data cube $\mathrm{X}^{n'} \in R^{W\times W\times M}$ of the target pixel, which includes both spectral and spatial information. Third, the spectral data $\mathrm{x}^i \in R^{1\times 1\times M}$ of each pixel in the neighborhood cube is converted to $\mathrm{x}^{i'} \in R^{W\times W\times 1}$, where $i, i' \in (1, \ldots, M)$. Fourth, these 2-D data are reassembled into a new neighborhood cube $\mathrm{T}^n \in R^{W\times W\times M}$, where the spectral data and spatial data exchange the spatial location compared with $\mathrm{X}^{n'}$. The advantage

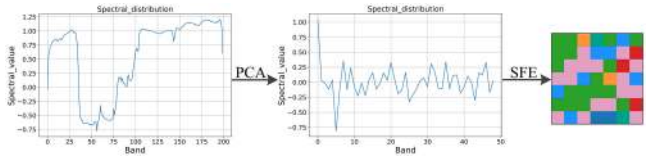

Fig. 3. Illustration of the validity of SEF operation.

of doing so can be seen from Fig. 3 that when the spectral information changes from 1-D to 2-D, the complexity of the spectral information of a single pixel increases, which can reflect more characteristics so that the separability of interclass is ameliorated and the intense correlation of interband is weakened to a certain degree.

B. Proposed SCNN Model for HSIC

The 3-D convolution kernel enables the network to effectively extract spectral-spatial features synchronously, which is very useful for improving classification performance. However, conventional 3-D convolution is more complex than 1-D and 2-D convolution and requires consumption of more computing resources. Therefore, according to preprocessed data features, by designing a multiscale 3-D step-by-step feature extraction strategy and spectral feature re-extraction mechanism, the proposed SCNN can ensure that the spectral and spatial features are efficiently extracted while having higher computational efficiency.

Spectral block 1: There are three 3-D convolution operations in the spectral block to extract spectral information and each convolution operation contains batch normalization and relu function. Specifically, the convolution kernel structure of the spectral block is $\mathrm{k}\times\mathrm{k}\times 1(\mathrm{Height}\times\mathrm{Width}\times\mathrm{Channel})$, where K is the size of the extraction window. The stride of convolution is $1\times 1\times 1$, and the number of convolution kernels is 32. Moreover, to obtain more distinguishable features, convolution kernel structures with different sizes of extraction

window are applied to the spectral dimension to ensure that features of different scales are extracted.

Spatial block: After the preliminary extraction of spectral features, the spatial feature extraction block will be applied to extract the spatial features of hyperspectral data. Different from the convolution kernel structure used in the spectral block 1, the pointwise convolutional strategy is used to extract the spatial information in hyperspectral data blocks, which convolution kernel structure is $1 \times 1 \times S$, and the stride of convolution is $1 \times 1 \times 1$. Especially, considering that different sizes of spatial information may include different features, a multiscale feature fusion mechanism is adopted in the spatial block.

Spectral block 2: This block is consistent with spectral block 1 in composition and structure. When extracting the spatial information, the pointwise convolution makes the spectral information of space objects in different locations will be weighted fusion and more abstract features will appear. Therefore, it is necessary to extract the spectral information of these multiposition fusions, which are exploited by the spectral block 2. The experimental results in Section III show that the SCNN which uses the spectral information re-extraction mechanism can achieve better performance.

When the features of HSI are extracted by SCNN, 3-D average pooling and dropout operations will be used to reduce the redundancy of extracted features. The size of the pool kernel is $2 \times 2 \times 2$ and the ratio of dropout is 0.1. Then, the 3-D spectral-spatial feature map is transformed into a 1-D feature vector by global average pooling (GAP). Finally, we use softmax function to get the classification result vector $\widehat{Y} = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_L,]$, where L is the number of categories of land-cover and $\hat{Y}$ is a one-hot label vector.

III. Experiments and Discussion

In this section, three real hyperspectral datasets with different characteristics (unbalanced training data, high spatial resolution, and small sample data) will be used to corroborate the dependability of the proposed method. In addition, to effectively evaluate the classification results, the overall accuracy (OA), average accuracy (AA), and kappa coefficient (κ) are used as the indicators of classification accuracy. Especially, the F1_score is used to evaluate the performance of methods on the imbalanced dataset. This is because the F1_score takes into account the accuracy and recall of a few categories, so it can measure the performance of the model under unbalanced data. All experiments are carried on a desktop computer with NVIDIA GeForce 1660 graphical processing unit (GPU) and 16 GB RAM.

A. Dataset Description

The first dataset is Indian Pines (IP), which contains 145×145 pixels of 16 categories. Specifically, excluding the contaminated 20 bands, it contains 200 spectral bands and its spatial resolution is 20 m. The single band map, ground truth map, and color code board are displayed in Fig. 4. The number of samples in the training set and testing set is listed in Table I.

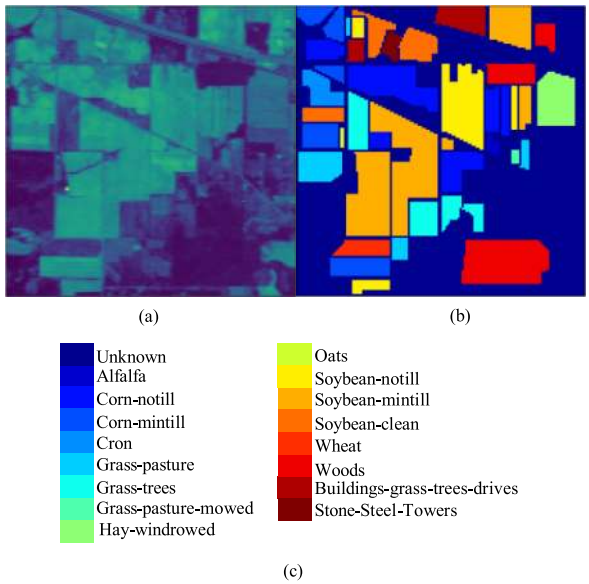

Fig. 4. IPs dataset. (a) Single band map. (b) Ground truth map. (c) Color code board.

TABLE I
Number of Training Samples and Testing Samples on the IPs Dataset

Class	Name	Training_Num	Testing_Num
1	Alfalfa	4	42
2	Corn-notill	144	1248
3	Corn-mintill	83	747
4	Corn	23	214
5	Grass-pasture	48	435
6	Grass-t	73	657
7	Grass-p-m	2	26
8	Hay-w	48	430
9	Oats	2	18
10	Soybean-notill	98	874
11	Soybean-mintill	247	2208
12	Soybean-clean	59	534
13	Wheat	20	185
14	Woods	127	1138
15	Buildings-g-t-d	38	348
16	Stone-s-t	9	84
	Total	1025	9924

The second dataset is Salinas Valley (SA), which contains 512×217 pixels and can be divided into 16 ground-truth classes. It has 224 effective spectral bands and 20 water absorption bands, which spatial resolution is 3.7 m. Fig. 5 shows the single band map and a corresponding reference map of the SA image. The number of samples in the training set and testing set is listed in Table II.

As shown in Fig. 6, the third dataset is Kennedy Space Center (KSC), which contains 512×614 pixels with a resolution of 18 m by a pixel. In total, 13 feature types included in this dataset, and the number of effective bands in KSC is 176. The number of samples in training set and testing set is listed in Table III. More specifically, different data partition strategies are used for different datasets to verify the generalization ability of the proposed model. For the IP dataset, we randomly select 10% of the samples of each of the 16 classes in the dataset as the

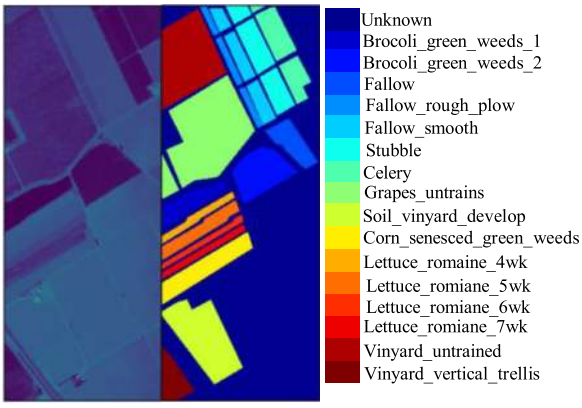

Fig. 5. SV dataset. (a) Single band map. (b) Ground truth map. (c) Color code board.

TABLE II
NUMBER OF TRAINING SAMPLES AND TESTING SAMPLES ON THE SV DATASET

Class	Name	Training_Num	Testing_Num
1	Brocoli_g_w_1	40	1969
2	Brocoli_g_w_2	75	3651
3	Fallow	39	1937
4	Fallow_r_p	28	1366
5	Fallow_s	53	2625
6	Stubble	79	3880
7	Celery	72	3507
8	Grapes_u	226	11045
9	Soil_v_d	124	6079
10	Corn_s_g_w	66	3212
11	Lettuce_r_4wk	21	1047
12	Lettuce_r_5wk	38	1889
13	Lettuce_r_6wk	18	898
14	Lettuce_r_7wk	21	1049
15	Vinyard_u	146	7122
16	Vinyard_v_t	36	1771
	Total	1082	53047

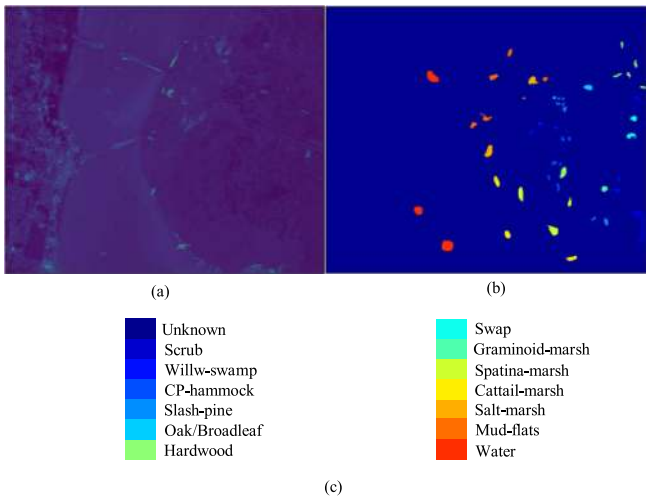

Fig 6. KSC dataset. (a) Single band map. (b) Ground truth map. (c) Color code board.

TABLE III
NUMBER OF TRAINING SAMPLES AND TESTING SAMPLES ON THE KSC DATASET

Class	Name	Training_Num	Testing_Num
1	Scrub	6	755
2	Willw_s	6	237
3	CP_h	6	250
4	Slash-pine	6	246
5	Oak/Broadlea	6	155
6	Hardwood	6	223
7	Swap	6	99
8	Graminoid_m	6	425
9	Spatina_m	6	514
10	Cattail_m	6	398
11	Salt_m	6	413
12	Mud_f	6	497
13	Water	6	921
	Total	78	5133

TABLE IV
PROPOSED NETWORK ARCHITECTURE

Blocks	Layers	Kernel Size	Filters
Spectral1	1-2	$3 \times 3 \times 1$	32
	3	$5 \times 5 \times 1$	32
Spatial	4	$1 \times 1 \times 3$	16
		$1 \times 1 \times 5$	16
Spectral2	5-6	$3 \times 3 \times 1$	32
	7	$5 \times 5 \times 1$	32
AvePool	8	$2 \times 2 \times 2$	-
Drop_layer	9	-	-
GAP	10	-	-
FC	11	-	-
Total Trainable Parameters: 84,432			

training set, and the remaining 90% as the test set. Because the IP dataset itself has the problem of data imbalance, under the above data division method, some classes have only two samples as training samples, which poses certain challenges to the learning capacity of the model. For the SA dataset, the percentages of the training set samples and the test set samples for each class are 1% and 99%, respectively. For the KSC dataset, we only select six samples for each class as the training set samples to further verify the classification performance of the model under the condition of small samples.

B. Experiment Setting

In our experiments, the detail of our proposed network architecture is shown in Table IV. The hyperparameters of the proposed model are set as follows. The influence of the size of sample patches (S) on the proposed method is shown in Fig 7. When S is 7×7, 9×9, and 11×11 , respectively, the OAs

TABLE V
CLASSIFICATION ACCURACIES (IN PERCENTAGES) ON IP DATASET USING PROPOSED METHOD AND STATE-OF-THE-ART METHODS

Class	Method					
	3D-CNN	DRN	HybirdSN	SSRN	DFFN	SFE-SCNN
1	**100±0**	97.73±0.04	**100±0**	95.45±0.34	98.12±1.41	97.62±0.59
2	91.49±1.91	95.43±0.27	**97.77**±1.33	97.90±1.37	97.73±1.56	**99.30±0.05**
3	95.66±0.94	98.20±0.89	96.85±1.21	98.33±1.42	97.24±1.75	**99.20±0.09**
4	64.35±0.37	94.59±0.04	97.67±1.13	91.89±5.55	**100±0**	99.53±0.65
5	84.86±1.29	**97.79±0.19**	96.42±2.42	96.46±4.26	92.31±3.45	95.86±0.41
6	96.19±0.35	99.41±0.05	99.25±0.34	**99.41±0.15**	96.4±2.34	99.24±0.30
7	**100±0**	66.67±3.34	61.54±3.56	74.07±4.31	**100±0**	**100±0**
8	**100±0**	**100±0**	**100±0**	**100±0**	99.61±0.26	**100±0**
9	**100±0**	63.16±2.67	56.25±2.16	78.95±3.47	72.3±5.37	**100±0**
10	93.64±0.70	95.16±5.19	97.14±1.34	96.70±2.14	**97.51±0.32**	96.45±2.18
11	94.84±2.28	98.34±0.04	98.19±0.69	98.39±0.31	98.70±1.14	**99.32±0.29**
12	71.18±1.85	89.91±0.59	95.49±2.67	90.99±0.79	94.34±2.39	**99.44±0.10**
13	94.97±0.20	**100±0**	**100±0**	**100±0**	94.16±1.67	**100±0**
14	99.27±1.33	99.75±0.06	**100±0**	99.41±0.18	**100±0**	**100±0**
15	97.33±0.99	98.61±0.13	96.25±1.62	**100±0**	97.66±0.22	97.70±2.81
16	81.32±1.90	**100±0**	**100±0**	**100±0**	97.62±0.43	98.81±0.61
AA(%)	91.56±0.36	93.42±0.52	93.30±1.27	94.87±0.27	95.85±0.19	**98.90±0.20**
OA(%)	92.67±0.49	97.25±1.39	97.86±0.39	97.74±0.56	97.63±0.24	**98.93±0.53**
κ×100	91.63±0.53	96.86±1.64	97.56±0.51	97.43±1.08	97.31±0.27	**98.78±0.59**
F1_score	92.72±0.51	97.55±0.87	97.88±0.64	97.82±0.35	97.72±0.23	**98.94±0.37**

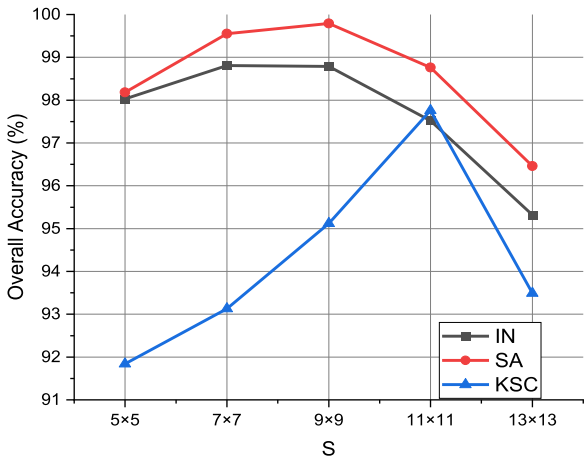

Fig 7. Effect of the size of sample window on classification accuracies in the case of IP, SA, and KSC.

of the proposed method are optimal. For training settings, the initial learning rates selected for IP, SA, and KSC are 0.00025, 0.00015, and 0.001, and the learning rate decreases by 0.00001, 0.00001, and 0.0001, respectively, for each update of the Adam optimizer. The training batch is 16, and each experiment trained 200 epochs.

C. Classification Results

In this section, several DL-based classification methods including 3D-CNN [21], SSRN [24], DRN [25], HybridSN [28], and DFFN [29] are used for comparison. For a fair comparison, the parameter settings of these methods are consistent with the default values of their original works. Figs. 8 –10 show the classification maps of different state-of-the-art methods with the corresponding accuracy scores (OA). Tables V –VII show the classification results in terms of class-specific accuracy, OA, AA, and Kappa coefficient.

The first experiment is conducted on the IP image. Because of the serious data imbalance in the IPs image, 10% of the labeled data was randomly selected for model training, and the rest for testing ("Oats" and "Grass_pasture_mowed" have only two samples for training.). From Table V, it can be observed that the proposed method outperforms the compared methods in the IP database. Specifically, the proposed method achieves OA 98.93%, with the gains of 6.26%, 1.68%, 2.60%, 1.19%, and 1.30% over the 3D-CNN, DRN, HybirdSN, SSRN, and DFFN method, respectively. In addition, it is worth noting that for the category with only two training samples (Class 7 and Class 9). When faced with only two training samples in IP, other methods have relatively serious classification errors. Conversely, the proposed method shows superior performance in the "Oats" and "Grass_pasture_mowed," which demonstrated that the SFE-SCNN can more fully extract the characteristics of these two samples. In addition, according to the F1_score in Table V, it can be observed that the proposed method can still achieve competitive results when the dataset is imbalanced. Fig. 8 shows the classification map of IP. Visually, the proposed method produces less internal noise in the Classification map, which is closer to the reference classification results.

The second experiment is conducted on the SA image. For SA, 2% and 98% of labeled data were randomly selected for model training and testing, respectively. Table VI presents detailed classification results of different methods. Similarly, the results on the SA dataset also indicate that the proposed SFE-SCNN is in the first place comparing to other methods. It is notable that compared with the IP dataset, the SA dataset has a higher spatial resolution. Therefore, it can be seen from SSRN, DFFN,

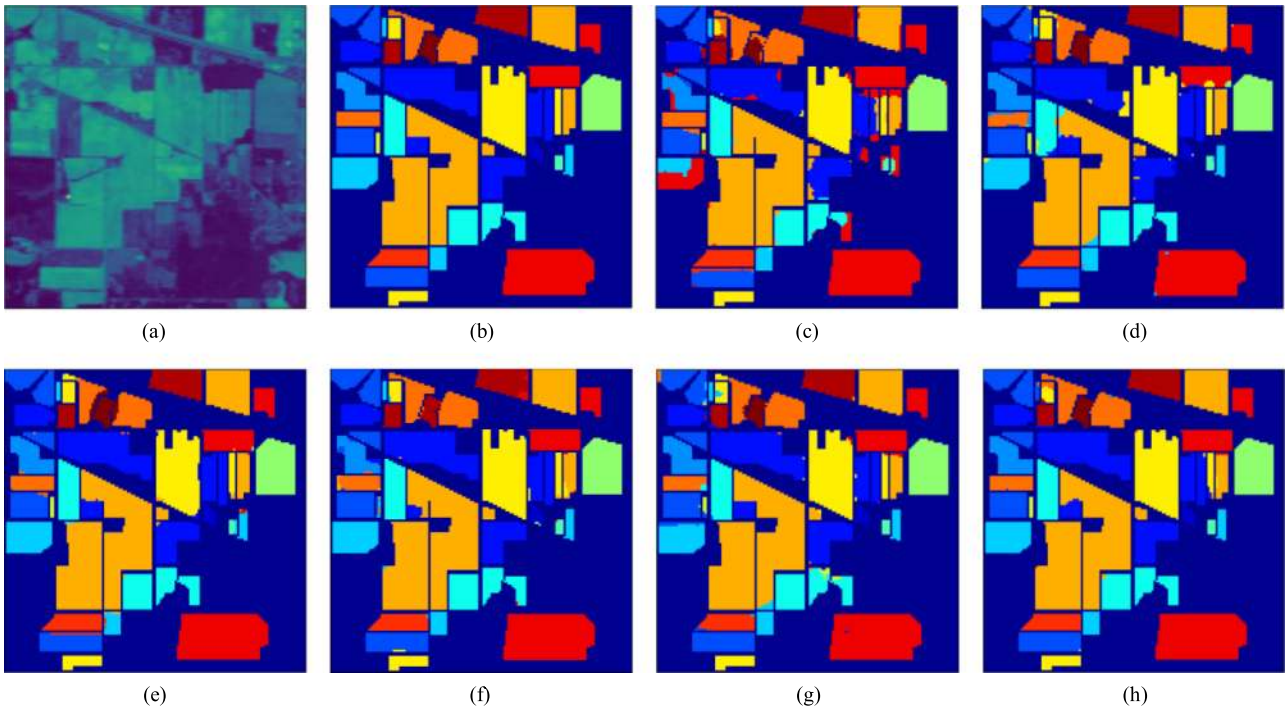

Fig. 8. Classification maps for IP. (a) Single band map. (b) Ground truth. (c)–(h) Predicted classification maps for 3D-CNN (OA = 92.67%), DRN (OA = 97.25%), HybirdSN(OA = 97.86%), SSRN(OA = 97.74%), DFFN(97.63%), and proposed SFE-SCNN(OA = 98.93%).

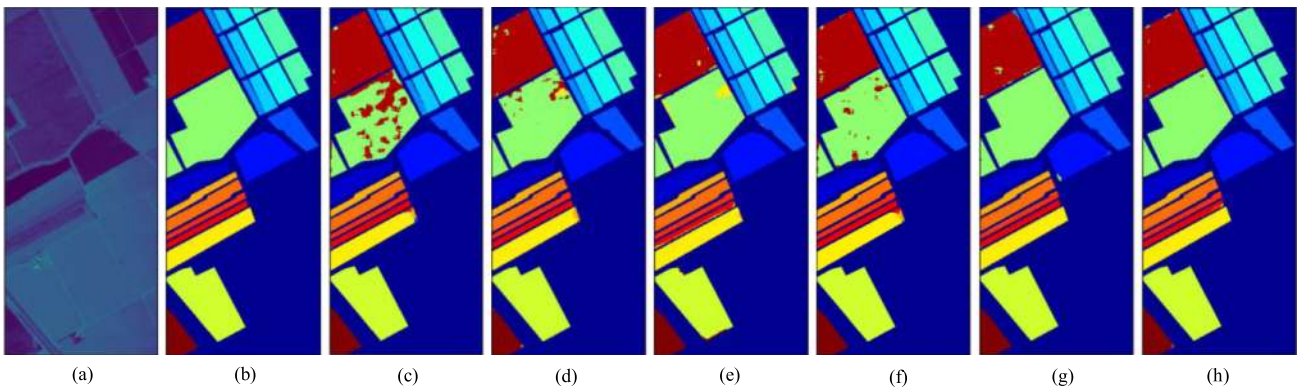

Fig. 9. Classification maps for SA. (a) Single band map. (b) Ground truth. (c)–(h) Predicted classification maps for 3D-CNN (OA = 94.29%), DRN (OA = 97.93%), HybirdSN(OA = 98.19%), SSRN(OA = 98.27%), DFFN(98.83%), and proposed SFE-SCNN(OA = 99.58%).

and SFE-SCNN that spatial information fusion can improve the classification results obviously. Fig. 9 shows the classification map obtained by the proposed method and state-of-the-art methods. Obviously, the proposed method can achieve better classification performance.

The third experiment is performed on the KSC image. To test the performance of the model under the condition of the small sample and low spatial resolution, for KSC, we only randomly select six samples from each category for model training. In addition, due to the low spatial resolution of the KSC dataset itself, it brings great challenges to the classification work. Actually, it can be clearly seen from Table VII that when facing a few shots training set and low-spatial resolution situation, compared to other methods, the 3D-CNN get the worst performance because of its shallow architecture and robust feature extraction method. Contrariwise, according to Table VII, the proposed SFE-SCNN yields outperforming results than other methods in most categories. The main reason is that the proposed method has a better ability to extract the spectral information. The classification maps of the proposed method and state-of-the-art methods can be compared in Fig. 10 visually. Evidently, the proposed method gains a superior visual result on the KSC dataset.

It is generally known that DL-learning HSIC methods often contain a large number of parameters to be trained, which requires a large number of computational resources. From Table VIII, it can be seen that the proposed SFE-SCNN has better

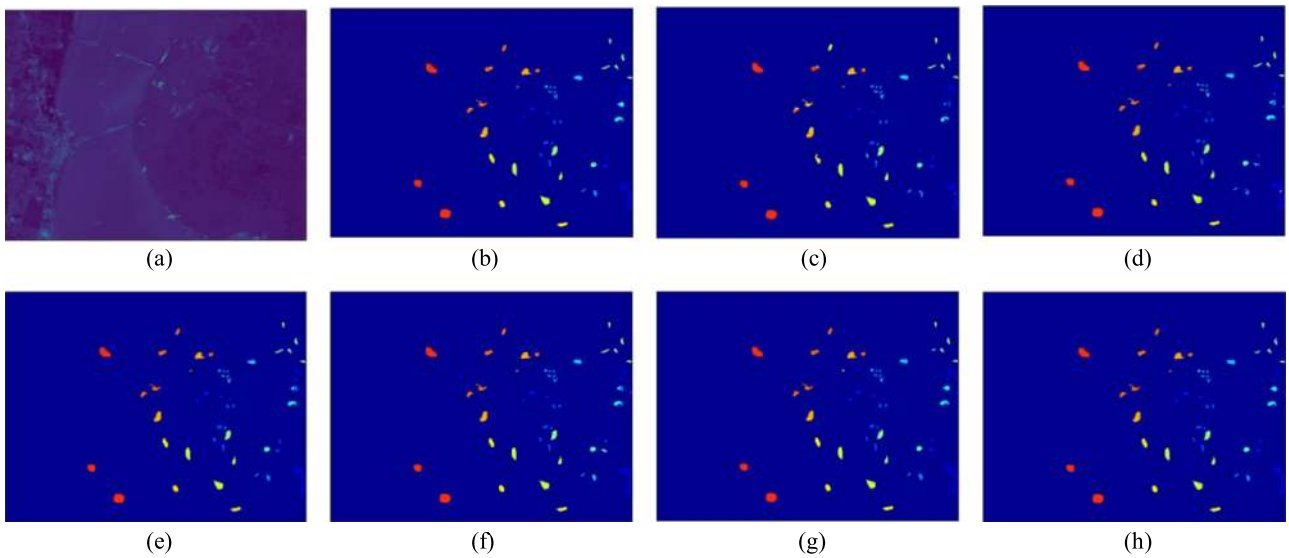

Fig. 10. Classification maps for KSC. (a) Single band map. (b) Ground truth. (c)–(h) Predicted classification maps for 3D-CNN (OA = 84.51%), DRN (OA = 92.48%), HybirdSN(OA = 96.26%), SSRN(OA = 95.48%), DFFN(95.73%), and proposed SFE-SCNN(OA = 97.76%).

TABLE VI
CLASSIFICATION ACCURACIES (IN PERCENTAGES) ON SA DATASET USING PROPOSED METHOD AND STATE-OF-THE-ART METHODS

Class	Methods					
	3D-CNN	DRN	HybirdSN	SSRN	DFFN	SFE-SCNN
1	**100±0**	**100±0**	**100±0**	99.92±0.56	**100±0**	**100±0**
2	**100±0**	**100±0**	99.86±0.13	**100±0**	98.56±1.14	**100±0**
3	**100±0**	**100±0**	**100±0**	99.87±0.24	99.16±0.26	**100±0**
4	98.55±0.01	98.77±0.38	**99.93±0.21**	98.91±0.88	94.32±1.75	99.89±0.04
5	99.28±0.07	99.85±0.42	99.06±0.34	99.43±022	98.21±1.14	**99.94±0.08**
6	**100±0**	**100±0**	99.92±0.06	**100±0**	99.63±0.70	99.99±0.03
7	**100±0**	**100±0**	**100±0**	**100±0**	99.50±0.62	99.99±0.03
8	75.81±0.23	95.05±0.77	94.31±1.43	96.94±1.26	99.12±0.31	**99.61±0.17**
9	**100±0**	**100±0**	**100±0**	99.87±0.13	99.63±0.32	99.83±0.19
10	94.51±0.43	94.27±2.29	95.68±0.51	97.23±0.98	99.69±0.26	**99.42±0.31**
11	99.62±0.73	99.81±0.07	**99.81±0.17**	95.79±1.46	97.42±1.41	99.08±0.12
12	**100±0**	93.87±0.94	99.90±0.09	99.55±1.46	95.37±1.56	99.26±0.06
13	96.91±1.52	**99.89±0.14**	99.45±0.11	98.00±0.64	92.32±0.84	96.72±0.16
14	97.83±0.58	93.87±0.15	**99.62±0.14**	95.07±0.30	98.73±0.55	96.55±0.37
15	99.36±0.38	96.14±0.48	98.39±1.46	95.50±5.62	98.13±0.62	**99.36±0.13**
16	97.76±0.11	99.83±0.12	98.55±0.33	**100±0**	99.57±0.16	99.43±0.09
AA(%)	97.47±1.28	98.59±0.70	99.02±0.35	98.29±0.08	98.08±0.19	**99.27±0.44**
OA(%)	94.29±0.40	97.93±0.33	98.19±0.23	98.27±0.49	98.82±0.24	**99.58±0.27**
κ×100	93.67±1.42	97.69±0.77	97.77±0.46	98.07±0.10	98.68±0.27	**99.53±0.31**

classification performance and higher computational efficiency. This is due to the lightweight feature of the proposed network and the efficient feature extraction capability.

D. Efficiency of SFE and Spectral Feature Re-extraction Operation

To verify the effectiveness of SFE operation, we compared the performance of the proposed model without SFE operation to the version with SFE operation and the effectiveness of spectral feature re-extraction operation. The results are shown in Table IX. Clearly, for three datasets, the SFE-SCNN model performs better than the SCNN model in terms of OA. Among them, SFE has a more obvious improvement on the dataset with a lower spatial resolution (IP and SA). This is because SFE operation can effectively improve the spectral feature recognition of the data and reduce the dependence on spatial information to some extent to improve the classification performance. Furthermore, the effectiveness of spectral feature re-extraction operation is shown in Table IX. Obviously, for three datasets, the SFE-SCNN model also outperforms the SFE-CNN model in terms of OA.

TABLE VII
CLASSIFICATION ACCURACIES (IN PERCENTAGES) ON KSC DATA SET USING PROPOSED METHOD AND STATE-OF-THE-ART METHODS

Class	Method					
	3D-CNN	DRN	HybirdSN	SSRN	DFFN	SFE-SCNN
1	99.60±0.32	95.63±0.16	**100±0**	99.60±0.06	99.56±0.39	**100±0**
2	98.73±0.67	82.28±0.98	94.51±0.33	77.64±1.24	99.53±0.08	**100±0**
3	80.00±3.46	93.20±4.11	**100±0**	98.80±0.17	**100±0**	**100±0**
4	54.47±5.47	63.41±5.88	74.80±5.29	62.60±2.21	**83.56±7.8**	72.36±2.60
5	75.48±6.45	77.42±3.33	98.06±0.35	80.65±0.43	**100±0**	**100±0**
6	49.78±4.43	68.61±7.85	**100±0**	78.03±0.15	99.51±0.31	**100±0**
7	**100±0**	**100±0**	**100±0**	**100±0**	99.19±0.64	97.98±1.72
8	83.06±3.96	97.18±0.62	80.09±3.51	**100±0**	85.29±3.63	93.18±0.23
9	97.28±1.95	99.22±0.76	**100±0**	**100±0**	84.13±1.59	**100±0**
10	80.65±3.08	92.96±0.80	**100±0**	**100±0**	99.24±0.63	**100±0**
11	84.50±2.42	**100±0**	99.60±0.17	**100±0**	99.21±0.48	96.13±0.89
12	49.70±6.57	88.93±1.30	**100±0**	99.60±0.01	97.44±1.51	**100±0**
13	**100±0**	**100±0**	**100±0**	**100±0**	99.59±0.46	**100±0**
AA(%)	81.01±0.81	89.14±0.60	95.45±2.08	92.07±0.39	95.86±0.34	**96.89±0.39**
OA(%)	84.51±1.01	92.48±1.03	96.26±1.41	95.48±0.57	95.73±0.63	**97.76±0.31**
κ×100	82.65±0.90	91.61±0.66	95.87±1.36	94.96±0.43	95.24±0.37	**97.50±0.44**

TABLE VIII
TRAINING AND TESTING TIME (IN SECONDS) OF DIFFERENT METHODS FOR IP DATASET

Methods	Train Time	Test Time	OA (%)
3D-CNN	7579.2	38.61	91.41
DRN	993.7	5.74	97.12
HybirdSN	289.8	2.43	91.21
SSRN	2913.6	16.29	97.94
DFFN	5927.4	29.08	97.63
Proposed	440.4	3.66	98.93

TABLE IX
EFFECT OF SFE AND SPECTRAL FEATURE REEXTRACTION ON CLASSIFICATION PERFORMANCE

dataset	Network		
	SCNN	SFE-CNN	SFE-SCNN
IP	95.88%	98.16%	**98.72%**
SA	96.69%	98.24%	**99.56%**
KSC	91.26%	95.32%	**96.56%**

IV. CONCLUSION

In this study, a novel SFE-SCNN model was proposed for the HSI classification. By transforming the structure of the hyperspectral data, the spectral features of the HSI are enhanced, and the proposed SCNN can efficiently extract spatial information and richer spectral features from it. Three standard hyperspectral datasets have demonstrated that the proposed method outperforms several recent state-of-the-art methods, especially under small sample and low spatial resolution conditions. Furthermore, efficient convolution kernel design and feature extraction methods make the proposed model more computationally efficient.

REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. 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.

[2] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[3] B. Rasti *et al.*, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep (overview and toolbox)," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Dec. 2020.

[4] 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.

[5] L. Samaniego, A. Bardossy, and K. Schulz, "Supervised classification of remotely sensed imagery using a modified k-NN technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2112–2125, Jul. 2008.

[6] L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei, "Supervised spectral–spatial hyperspectral image classification with weighted markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1490–1503, Mar. 2015.

[7] M. Hasanlou, F. Samadzadegan, and S. Homayouni, "SVM-based hyperspectral image classification using intrinsic dimension," *Arabian J. Geosci.*, vol. 8, no. 1, pp. 477–487, Jan. 2015.

[8] L. Gao *et al.*, "Subspace-based support vector machines for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 349–353, Feb. 2015.

[9] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogram. Remote Sens.*, vol. 158, pp. 279–317, Dec. 2019.

[10] B. Rasti *et al.*, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep (overview and toolbox)," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Dec. 2020.

[11] K. Tan, E. Li, Q. Du, and P. Du, "Hyperspectral image classification using band selection and morphological profile," in *Proc. 4th Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, Shanghai, 2012, pp. 1–4.

[12] A. Santara *et al.*, "BASS net: Band-adaptive spectral-spatial feature learning neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5293–5301, Sep. 2017.

[13] C. Yu, Y. Wang, M. Song, and C. Chang, "Class signature-constrained background-suppressed approach to band selection for classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 14–31, Jan. 2019.

[14] J. Zou, W. Li, and Q. Du, "Sparse representation-based nearest neighbor classifiers for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2418–2422, Dec. 2015.

[15] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.
[16] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.
[17] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 16–28, Jan. 2018.
[18] Q. Liu, Z.Wu, Y. Xu, L. Du, and Z. Wei, "Kernel low-rank representation based on local similarity for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1920–1932, Jun. 2019.
[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[21] 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. Artif. Intell. Statist.*, 2012, pp. 127–135.
[22] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, 2014, pp. 580–587.
[23] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.*, vol. 44, no. 10, 2016, Art. no. e360.
[24] 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.
[25] 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.
[26] H. Sun, X. Zheng, X. Lu, and S. Wu, "Spectral–Spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3232–3245, May 2020.
[27] 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.
[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, 2016, pp. 770–778.
[29] C. Zhang, G. Li, and S. Du, "Multi-scale dense networks for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9201–9222, Nov. 2019.
[30] G. Yang, U. B. Gewali, E. Ientilucci, M. Gartley, and S. T. Monteiro, "Dual-channel densenet for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Valencia, 2018, pp. 2595–2598.
[31] 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.
[32] X. Li, M. Ding, and A. Pižurica, "Deep feature fusion via two-stream convolutional neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2615–2629, Apr. 2020.
[33] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, 2020, to be published.
[34] H. Yu *et al.*, "Global spatial and local spectral similarity-based manifold learning group sparse representation for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3043–3056, May 2020.
[35] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791–3808, Jun. 2020.
[36] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3016820.
[37] J. Liang, J. Zhou, Y. Qian, L. Wen, X. Bai, and Y. Gao, "On the sampling strategy for evaluation of spectral-spatial methods in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 862–880, Feb. 2017.
[38] H. M. Gao *et al.*, "Multi-branch fusion network for hyperspectral image classification," *Knowl. Based Syst.*, vol. 167, pp. 11–25, 2019.

Hongmin Gao received the Ph.D. degree in computer application technology from Hohai University, Nanjing, China, in 2014.

He is a Professor with the College of Computer and Information, Hohai University. His research interests include deep learning, information fusion, and image processing in remote sensing.

Zhonghao Chen received the B.S. degree in electronics and information engineering from West Anhui University, Luan, China, in 2019.

He is a Graduate Student with the College of Computer and Information, Hohai University. His research interests include deep learning and image processing.

Chenming Li received the B.S., M.S., and Ph.D. degrees in computer application technology from Hohai University, Nanjing, China, in 1993, 2003, and 2010, respectively.

He is a Professor and the Deputy Dean at the College of Computer and Information, Hohai University. His research interests include information processing systems and applications, system modeling and simulation, multisensor systems, and information processing.

Dr. Li is a Senior Member of the China Computer Federation and the Chinese Institute of Electronics.