Automatic Clustering-Based Two-Branch CNN for Hyperspectral Image Classification

Yuan Li, Qizhi Xu, *Member, IEEE*, Wei Li, *Senior Member, IEEE*, and Jinyan Nie

***Abstract*— It is observed that the great spectral variation in the same hyperspectral image (HSI) pixel class often leads to misclassification. To solve this problem, we have proposed an automatic clustering-based two-branch convolutional neural network (CNN): first, to reduce the intraclass spectral variation, the HSI pixels are automatically subdivided into smaller classes by clustering; second, in order to suppress the interference of spectral amplitude variation, the SincNet is introduced to capture the spectral pattern by giving more weight to the spectral shape; third, the DS-CNN with double directional strip convolution kernel is designed to extract spatial feature, so that specific contextual interactional features can be collected, especially in strip-shaped field-like roads and farmlands; finally, the spectral and spatial features extracted by the two branches are fused at fully connected layer to obtain an accurate classification. Extensive experiments demonstrated that the proposed method can obtain better classification performance than the state-of-the-art methods.**

***Index Terms*— Automatic clustering, convolutional neural network (CNN), deep learning, hyperspectral image (HSI).**

I. Introduction

HYPERSPECTRAL image (HSI) has received considerable interest in recent years for high spectral and spatial resolution [1]–[3]. Especially for classification, it has unique advantages because it can uncover subtle differences in spectral features of different materials [4]–[6]. The HSI classification has numerous useful applications, to name a few, urban planning, disaster monitoring, and scene recognition. However, it can be observed through a large number of data sets that the spectra of most materials varied widely (see Fig. 1). The spectral variability fundamentally affects the accuracy of HSI classification [7], [8].

Assuming that the spectral curves of all pixels in each class are the same (e.g., they are all average spectral vectors, as shown by the black line in Fig. 1), each class can be easily distinguished. Nevertheless, in the same ground classes, the variation range of the spectral amplitude or shape of pixels may be different. So that the foreign matter may have similar spectra, and the objects within a class may have different spectra. This is the bottleneck hindering the improvement of classification accuracy. Furthermore, spectral variation complicates the statistical distribution of the sample points and aggravates the problems caused by a small number of samples. As a consequence, reducing the impact of spectral variation is one of the key problems of HSI classification. To tackle this problem, a wide variety of HSI classification methods were proposed as follows.

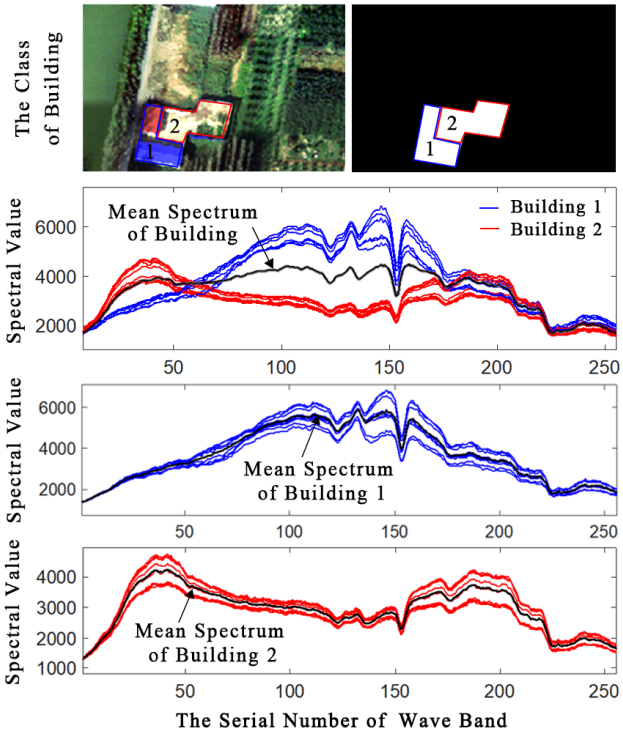

Fig. 1. Spectral variation of building pixels. The images in the first column are the original image and ground truth map of the Xiong-An data set. The white area in ground truth is the hand-craft labels of the building and the spectral curves of this class are illustrated in the second column. The latter two figures represent the spectral curves of different materials in building class, respectively.

A. HSI Classification Using Feature Engineering

Effective feature representation is a very critical part of the HSI classification system [9]–[11]. Traditionally, in the

Manuscript received August 16, 2020; revised November 5, 2020; accepted November 10, 2020. Date of publication December 3, 2020; date of current version August 30, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61972021 and Grant 61672076. *(Corresponding author: Qizhi Xu.)*

Yuan Li and Qizhi Xu are with the School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China, and also with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: qizhi@bit.edu.cn).

Wei Li is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: liwei089@ieee.org).

Jinyan Nie is with State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China.

Digital Object Identifier 10.1109/TGRS.2020.3038425

early development stage, the HSI classification was dominated by the works that take advantage of spectral features [12]. Xue *et al.* [13] presented a completely different approach from a subpixel target detection viewpoint. They utilized band selection then nonlinear expansion (BSNE) and iterative constrained energy minimization to classify the HSI. It can be easily implemented and has advantages over other methods. At the same time, with the development of remote sensing imaging technology, the spatial resolution of HSI has been increased, and the spectral–spatial joint features have also attracted more attention [14]. For example, in [15], many features, such as multiple features, textural features, Gray Level Cooccurrence features, and statistical features, are incorporated as the calculation parameters to get better classification results. Moreover, inspired by the tensor learning methods, spatial and spectral features can also be fused into 3-D tensors [16]. It can reduce the loss of intrinsic structure information for HSI. Guo *et al.* [17] proposed a tensor-based technique for HSI classification and use a multilinear principal component analysis to preprocess tensors. This method shows an interesting result. However, most of these tensorial methods build tensors physically and neglect the features that follow a predefined logical arrangement. Based on this work, in [18], a novel generalized tensor regression (GTR) approach, extended from a simple and efficient classifier, is utilized for HSI classification.

On the other hand, different methods have different benefits and challenges. Support vector machines (SVMs) have shown excellent performances for classifying hyperspectral remote sensing images. Xia *et al.* [19] devised a novel ensemble rotation-based SVM (RoSVM), which can effectively enhance classification accuracy by combining with multiple classifier systems (MCSs). And in [20], an adaptive kernel sparse representation classifier (AK-SRC) is utilized to classify these interest points into different classes. It can take the similarity and diversity of different types of feature descriptors into full consideration and show excellent classification performance. Simultaneously, the problem of classifying objects can also be formulated as a clustering procedure [21]. The cluster-based techniques can get good classification results by finding distinct structures in the spectral feature space. In [22], a novel spectral–spatial sparse subspace clustering (SSC) algorithm is proposed for hyperspectral remote sensing images. They introduce the SSC algorithm to HSIs, and results show that the method has excellent performance.

B. HSI Classification Using Convolutional Neural Networks

Currently, convolutional neural networks (CNNs) have demonstrated excellent performance for HSI classification [23], [24]. Because it can be naturally adapted to deal with the problem that HSI often lies in a nonlinear and complex feature space [25]. We describe these CNN methods from two aspects: one-stream and multistream. First, CNN whose architecture with only one branch is considered one-stream CNN. In [26], the hybrid of one-stream deep CNN and logistic regression is first introduced into HSI classification. It is a unique idea. Chen *et al.* [27] proposed a 3-D CNN with combined regularization to extract effective spectral–spatial features of hyperspectral imagery. The experiments reveal that the proposed approach can provide competitive results. In 2015, a deep learning-based classification method is designed in [28]. The method can hierarchically construct high-level features in an automated way, and the results showcase the potential of the developed approach for accurate hyperspectral data classification. Besides, Hang *et al.* [29] proposed a new cascaded recurrent neural network, which was integrated by two cascaded CNN. The first one is to reduce redundancy and the second one is to learn complementarity. One-stream CNN has a simpler connection structure and can achieve better classification results than traditional methods.

Second, unlike these methods, other researchers proposed multistream CNN-based HSI classification methods from another aspect [30], [31]. In most CNNs, the feature extracting and classifier training are separated. To overcome this drawback [32], a spectral–spatial unified network (SSUN) is designed and combines both shallow and deep convolutional layers to deal with the information loss. In [33], a two-channel deep CNN is proposed. Discriminant information is captured separately from the spectral domain and the spatial domain and can be effectively exploited and fused. Xu *et al.* [4] proposed a novel two branch CNN based on multisource data (MS-CNN) for classification fusion of HSI and data from other multiple sensors, such as light detection and ranging (LiDAR) data. They aid two networks to focus on different features separately and obtained an excellent classification performance. In 2018, a diverse region-based CNN is proposed in [34] for HSI classification. It extracts multiscale spatial–spectral features by exploiting diverse region-based inputs, to get more discriminative power. The multistream architecture helps extract more effective features and all these approaches can yield competitive performance. However, as far as HSI classification is concerned, they do not further subdivide the classes with large spectral differences, which easily promotes network overfitting.

C. Classification Methods Aiming at Spectral Variation

From a large amount of empirical knowledge, the notion of a single predetermined spectral signature for each material (or class) is an ideal concept not observed in real-world applications [35], [36]. First, in the satellite sensor imaging process, uncompensated errors in the sensor, uncompensated atmospheric, and the sun angle relative to the zenith, these objective factors may all cause the variation of spectral in HSI [37]. Second, biological activity is also an important reason. For instance, different benthic animals may cause differences in the spectra reflected by the water [38]. Furthermore, whether or not to be affected by pests and diseases will also make crop areas produce spectral differences. In these cases, even if the sensor images the same object, it may produce different spectral information. Third, semantically called objects of a certain class may be different in practice. It is mentioned in [39] that the spectral character of an object (e.g., mineral actinolite) even changes with particle sizes. This means that the mixing of different materials within a class will inevitably

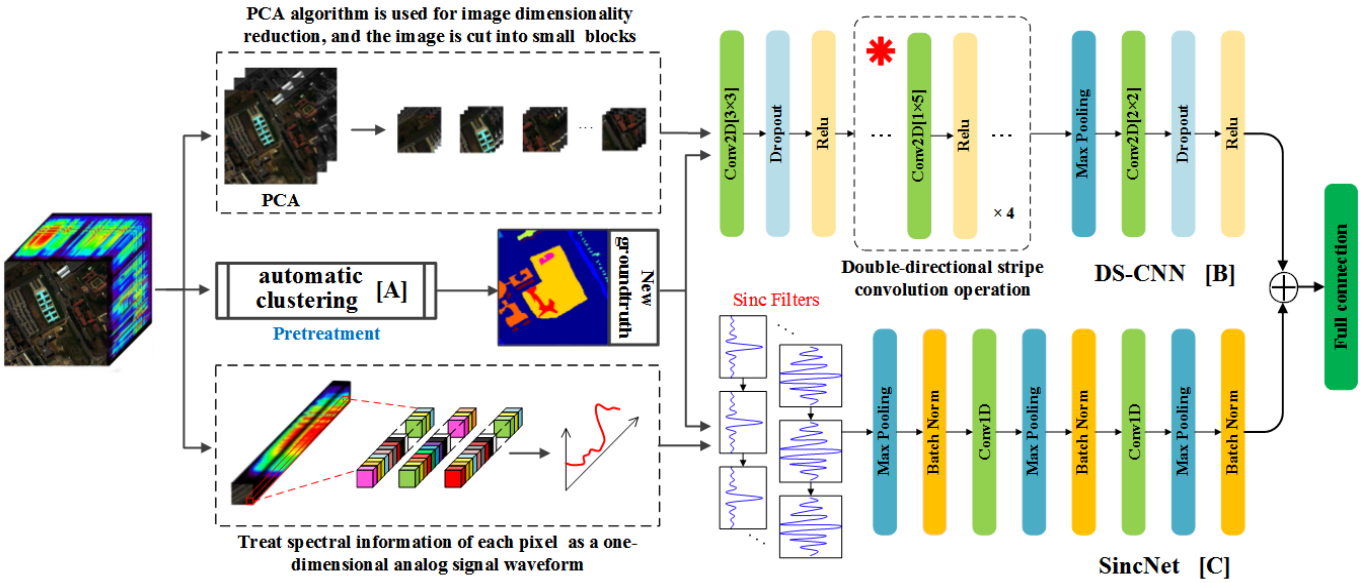

Fig. 2. Overall flowchart of proposed two-branch CNN for HSI classification. Pretreatment. (a) Clustering module. (b) DS-CNN module for spatial feature extraction. (c) SincNet module for 1-D spectral information extraction.

lead to the difference in the spectra. In another case, the hand-crafted label bundle different classes together. Fig. 1 shows several instances of reflectance spectrum derived for pixels in the buildings class in a scene, some of which are courtyard pixels (called building 2 in the figure) instead of real building pixels (called building 1 in the figure). We can note that the shape of the spectra in different subclasses varies widely.

Consequently, spectrum variability increases the difference within a class and reduces the discrimination between different classes. It brings great difficulties to the fine classification. To reduce the intraclass spectral variation, many research scholars have made efforts. In [40], a method is proposed to extract features by employing graph embedding and deep learning models. The supervised within-class/between-class hypergraph is constructed to reduce intraclass variation as well as the similarity between different classes in the spectral domain. The model is very interesting. Intra clustering minimization can also be achieved by graph partitioning [41], [42]. Abou-Rjeili and Karypis present a new multilevel graph partitioning algorithm in [43]. The new clustering-based schemes can identify and collapse together groups of vertices that are highly connected, which significantly outperform existing approaches. Moreover, the methods based on clustering are also available. In [44], a technique is devised to reduce the degree of intraclass spectral variation by defining spectral subclasses. It is proven that the technique is very effective in soft classification. In 2014, a density-peak clustering algorithm was designed by Rodriguez *et al.* [45]. This method can automatically detect nonspherical clusters and find the number of clusters without manual operation. What is more, the initialization parameters of the classification system are very important [46], [47]. So, we applied the density-peak cluster algorithm in our work as a preprocess module to get a better classification result.

From another perspective, spectral variation can result in the problem with small sample size, because it complicates the distribution of samples. In other words, if the distribution of the samples is very uniform, a small number of samples can accurately describe its statistics characteristics. In recent years, tensor-based methods have shown excellent performance against the problem of a small number of samples [48], [49]. A tensor-based method classification models are proposed in [50]. The model constructs a classifier, whose network weights are constrained to satisfy a rank-R Canonical Polyadic Decomposition. So that it can significantly reduce the number of weight parameters required to train the model (and thus the respective number of training samples). And, the Rank-R feedforward neural network (FNN) classifier shows better classification performance than the rank-1 FNN classifier [51]. However, from a statistical point of view, clustering can reduce the difference within a class and make the distribution of samples more uniform. Therefore, it is also a very good method to solve the problem with a small samples size for HSI classification.

Based on the above analysis, we borrowed the density-peak clustering algorithm [45] as a preprocessing module to reduce the large within-class differences. The clustering result was regarded as a new ground truth map to supervise the two-branch CNN. On the one hand, it is observed that square or strip-shaped objects exist in most scenes, such as roads, farmland, and buildings. Then, CNN with double directional stripe convolution kernels (DS-CNN) is designed to extract spatial features of HSI, ensuring the architecture with a strip receptive field can excavating more meaningful features. On the other hand, we can see that the spectral curve is much like a 1-D speech signal. Mirco Ravanelli *et al.* [52] proposed a method named SincNet for directly recognizing speech from the raw waveform. The novel CNN encourages the

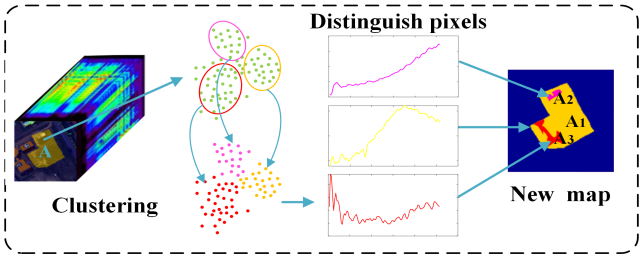

Fig. 3. Pretreatment: New ground truth map generated by Automatic Clustering. The A class in original ground truth is subdivided into three smaller classes: A1, A2, and A3.

first layer to discover more meaningful features by exploiting parametrized sinc functions. It makes the network converge faster and perform better. Inspired by this, we employed the SincNet to capture the spectral information of the pixels in HSI. It can fit more precise spectral details and give more weight to spectral shape, which suppresses the interference of spectral amplitude variation. Finally, a softmax classifier is used to fuse all features obtained from SincNet and DS-CNN and to classify pixels of HSI. Accordingly, we proposed to construct an automatic clustering-based two-branch CNN for HSI classification.

The main contributions of this article are as follows: First, in order to reduce the within-class difference caused by spectral variation, the HSI pixels are automatically subdivided into smaller classes by exploiting the density peak clustering algorithm. Second, the "double stripe" convolution module is designed to extract spatial information, and the collected specific contextual interactional features enhance the sensitivity of the CNN to objects in specific scenes. Finally, the SincNet is introduced to capture the details of the spectral pattern by giving more weight to the spectral shape, so that the interference of spectral amplitude variation can also be suppressed.

The remainder of this article is organized as follows. The proposed classification framework is described in Section II. The experiments and analysis are discussed in Section III. The conclusion is drawn in Section IV.

II. Methodology

Our approach for HSI classification utilizes an end-to-end framework including a clustering module, a DS-CNN, and a SincNet. As shown in Fig. 2, a new ground truth map is obtained by clustering HSI pixels. Then we utilize a two-branch network, constructed by DS-CNN and SincNet, to extract spatial and spectral features separately. The end of our framework is a classifier that consists of fully connected layers with log-softmax loss. And the details are elaborated in Sections II-A–II-C.

A. Clustering Module

Based on the aforementioned analysis, clustering is an effective way to reduce the intraclass spectral variation of HSI. Before being input the network for training, the HSI is preprocessed by the density peak clustering algorithm, which clusters the classes with large intraclass differences in HSI into some smaller subclasses. These new subclass labels are treated as a new ground-truth map to supervise the training of CNN, as shown in Fig. 3. The density peak clustering approach is based on the idea that cluster center pixels are characterized by a higher density than their neighbors and a large distance from points with higher densities.

For each pixel spectral vector x of HSI, we assume $x = [x_0, x_1, x_2, \ldots x_{n-1}]^T$, where n is the number of bands. And two quantities need to be computed: its local density ρ and its distance δ from pixels of higher density. Both two quantities depend on the similarity s_{ij} between the pixel vectors x_i and x_j. It can be calculated by the distance formulas such as Euclidean distance [53] and Manhattan distance [54]. The larger the s_{ij}, the smaller the similarity between the two vectors. Compared with other distance calculation methods, Manhattan distance has higher stability, and more intuitive representation for vectors with large feature differences in the data set. Therefore, we choose the Manhattan distance d_{ij} in our approach to calculating the similarity between pixel vectors

$$d_{ij} = \sqrt{|x_i - x_j|' \sum^{-1} |x_i - x_j|} \tag{1}$$

$$s_{ij} = d_{ij}^{-1}. \tag{2}$$

Based on the above calculation results, the local density ρ_i of ith data point x_i is defined as

$$\rho_i = \sum_j \varphi\left(s_{ij}^{-1} - c\right) \tag{3}$$

where $\varphi(m) = 1$ if $m < 0$ and $\varphi(m) = 0$ if $m = 0$, otherwise, and c is a similarity cutoff value. The ρ_i can be treated as the number of points that are similar to point x_i in the range of similarity cut-off value c. δ_i is expressed by computing the maximum similarity (or the minimum distance) between the point x_i and any other point x_j with higher density

$$\delta_i = \min_{j:\rho_i > \rho_j} \left(s_{ij}^{-1}\right). \tag{4}$$

For the point with highest density, we conventionally take $\delta_i = \max_j(s_{ij}^{-1}) = \max_j(d_{ij})$. Note that δ_i is much larger than the nearest neighbor distance only for points that are local or global maxima in the density. Thus, cluster centers are recognized as points for which the value of ρ_i and δ_i are anomalously large. If we construct a 2-D space with ρ_i and δ_i as the coordinate axes, then the cluster centers are the points near the upper right. After the cluster centers have been found, each remaining point is assigned to the same cluster as its nearest neighbor of higher density.

After clustering, a new ground truth map with smaller subclass labels of HSI can be obtained. Assuming that the labels of original ground truth map classes are A, B, C. . ., then the new ground-truth map labels are expressed as A1, A2, B1, B2, B3, C. . . We utilize the new ground-truth map to supervise the network adjust parameters, but the original ground-truth map is applied to calculate the network classification accuracy. Assuming that the new label of one A pixel is A1/A2, and its prediction result is also A1/A2, it is considered to be correctly

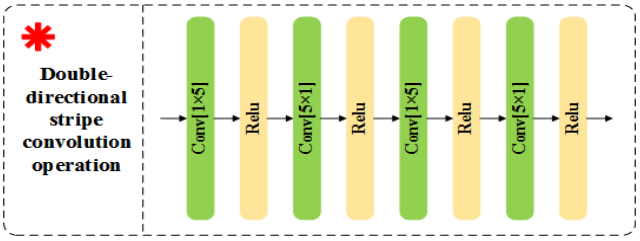

Fig. 4. Detailed framework of double-stripe convolution. The module is mainly composed of four convolutional layers (green blocks) and four rule activation function layers (yellow blocks). The size of each convolution kernel is 1×5 or 5×1, which means that it can extract spatial features of strip regions in different directions.

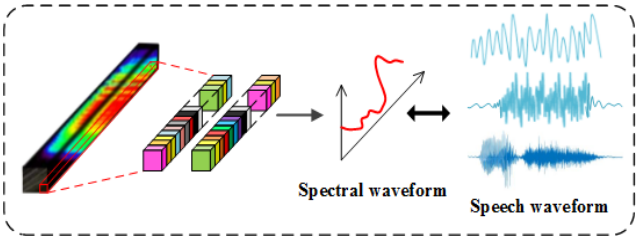

Fig. 5. Spectral waveforms are very similar to low-frequency speech waveforms. The rich spectral information of each pixel in the HSI can be regarded as a 1-D signal. According to its similarity with the speech signal, we can effectively learn from relevant processing methods to obtain more interesting results.

classified. Among them, if the pixel labeled A1 is predicted to be in the class A2, it is not regarded as a misclassified pixel.

B. DS-CNN Module

The DS-CNN is designed to extract spatial information of HSI, in order to better consider the spatial relationship of pixels. However, a lot of redundant information of HSI needs to be removed before it is input so that the computational complexity can be reduced. Here, the principal components analysis (PCA) [55] is used to reduce the dimension of HSI and only retain the largest three principal components. We cut the image into patches with the size of $13 \times 13 \times 3$ centered the labeled pixel and randomly select a fixed number of patches as network training samples.

The architecture of the proposed DS-CNN is illustrated in Fig. 2(b). It makes up of 2-D convolution layers, activation layers, max-pool layers, and dropout layers [56]. Among them, the first layer and the last layer are 3×3 and 2×2 square convolution kernels. The middle four layers are alternately arranged with 5×1 and 1×5 strip convolution kernels, as shown in Fig. 4. Specifically, before square convolution kernel operations, the surrounding of input data is padding with zero, and before stripe convolution operations, padding with one. At the same time, the convolution stride is set to 1. Besides, the number of convolution kernels used in each convolution operation is 256. The dropout layer is performed to prevent the network from overfitting, and the nonlinear transformation layer (Relu) [57] is chosen to compute the output activation value of each layer.

As discussed previously, most CNN commonly uses square convolution kernels to extract features. But for scenes with complex backgrounds, the square selection region may include between-class pixels. Especially in areas including narrow strip-shaped objects, it is difficult to select materials only from one class even for a small receptive field. It is obvious that the classification result will be disappointed. However, the multi-layer stripe convolution kernel can expand the receptive field in different directions, which can make the selected region more representative. Thus, DS-CNN has a greater advantage especially for scenes containing roads and houses.

C. SincNet Module

Through careful observation of the experimental data, we found that the waveform of the spectrum is very similar to the low-frequency speech waveform, as shown in Fig. 5. In [52], the sinc function is used as a filter to extract speech signal features. It achieves better performance and is more interpretable. Inspired by this method, we employed SincNet for spectral feature extraction of HSI images.

The flowchart of SincNet consists of four layers with different functions. As shown in Fig. 2(c), it includes sinc filters, convolution layer, batch norm layer, and max-pool layer [58], [59], which aim to extract spectral features in HSI. In this module, the preprocessed 1-D data (each pixel with all band information, that is pixel vector $x = [x_0, x_1, x_2, \ldots, x_{n-1}]^T$) are first fed into the sinc function filter. Then, we utilize convolution and max-pool operation to extract normalized features. Each convolution of a standard CNN is defined as follows:

$$y = x * h \tag{5}$$

$$y(k) = \sum_{i=0}^{n-1} x(i) \cdot h(k-i) \tag{6}$$

where x is input pixel spectral vector, $h = [h_0, h_1, h_2, \ldots, h_{m-1}]^T$ is the filter of length m, and $y = [y_0, y_1, y_2, \ldots, y_{m+n-1}]^T$ is the filtered output. All the m elements of each filter have to be adjustable through CNN and learned from data. Conversely, the SincNet performs the convolution with a predefined function G that depends on few learnable parameters f_1, f_2 only, as highlighted in the following equation:

$$y = x * g[f_1, f_2]. \tag{7}$$

The vector $g = [g_0, g_1, g_2, \ldots, g_{m-1}]^T$ calculated by the filter function G can be treated as a convolution kernel in network. But the adjustable network parameters is f_1 and f_2, instead of the elements in the vector (e.g., g_0). The function G is defined as a filter composed of rectangular bandpass filter [60], f_1, f_2 are the high and low cutoff frequencies of each filter. It can be written as the difference between two low-pass filters in the frequency domain. When converted to the time domain [61], the function G can be expressed as follows:

$$G[f_1, f_2] = 2f_2 \text{sinc}(2\pi f_2) - 2f_1 \text{sinc}(2\pi f_1) \tag{8}$$

Where the sinc function is defined as $\text{sinc}(a) = \sin(a)/a$. Due to the only two learnable parameters of each filter, the training process of SincNet is much faster than standard CNN. At the

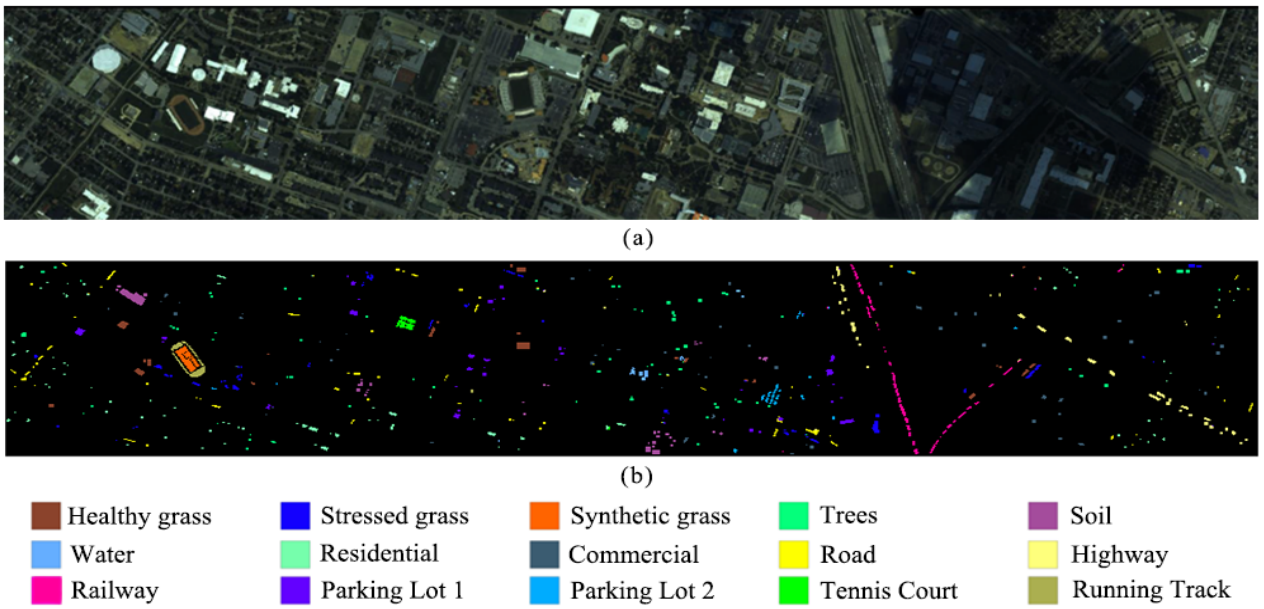

Fig. 6. Houston data set. (a) False color image. (b) Ground-truth map.

same time, the characteristics of the sinc function or the prior knowledge about filter shape enable the filter to fit the peak details of the spectral curve and force the network to focus on the spectral shape rather than amplitude.

After that, the features derived from two branches (DS-CNN and SincNet) are concatenated together and fed into the fully connected layer. Finally, we utilize the log-softmax function as a classifier to complete the data classification.

III. Experiments

In this section, we show the efficacy of the proposed two-branch CNN on three data sets and compared it with the state-of-the-art methods. In order to further analyze the effectiveness of different parameters, we have conducted controlled experiments. It is a scientific test done under controlled conditions, to test a single variable at a time. And, all programs in experiments are implemented using Python language, and the network is constructed by PyTorch deep learning framework. Pytorch is an open-source Python machine learning library that can define deep learning models and can be flexibly trained and used.

A. Data Sets

To evaluate the performance of the proposed method, the Houston data set, University of Pavia data set, and Xiong-An data set are employed. As shown in Figs. 6 and 7, for each data set, we select a fixed number of labeled pixels per class for training and other pixels for testing.

The Houston data were acquired by a Compact Airborne Spectrographic Imager produced by ITRES company in Canada sensor over the area of the University of Houston campus and neighbor area and provided by the 2013 IEEE GRSS data fusion competition. The data size is 349 $\times$ 1905 pixels and contains 144 bands with a spectral range from 0.364 to 1.046 μm. Approximately 20 029 labeled pixels with

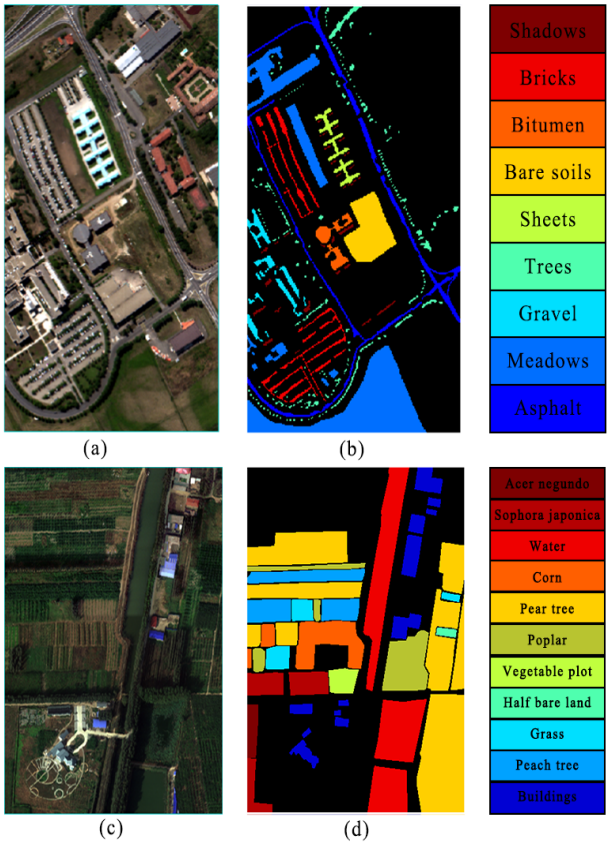

Fig. 7. Two data sets. (a) False color image of PaviaU data set. (b) Ground-truth of PaviaU data set. (c) False color image of Xiong-An data set. (d) Ground truth of Xiong-An data set.

15 classes are from the ground-truth map, and Table I shows the training and testing samples.

The University of Pavia data set was collected by the Reflective Optics System Imaging Spectrometer sensor covering

TABLE I
NUMBER OF TRAINING AND TESTING SAMPLES FOR THE HOUSTON DATA SET WITHOUT CLUSTERING

#	Class	Training	Test
1	Health grass	300	951
2	Stressed grass	300	965
3	Synthetic grass	300	398
4	Trees	300	920
5	Soil	300	966
6	Water	300	32
7	Residential	300	894
8	Commercial	300	858
9	Road	300	886
10	High Way	300	977
11	Rail Way	300	941
12	Parking lot 1	300	995
13	Parking lot 2	300	214
14	Tennis court	300	140
15	Running track	300	392
-	Total	4500	17029

TABLE II
NUMBER OF TRAINING AND TESTING SAMPLES FOR THE UNIVERSITY OF PAVIA DATA SET WITHOUT CLUSTERING

#	Class	Training	Test
1	Asphalt	800	5831
2	Meadows	800	17849
3	Gravel	800	1299
4	Trees	800	2264
5	Painted metal sheets	800	545
6	Bare Soil	800	4229
7	Bitumen	800	530
8	Self-Blocking Bricks	800	2882
9	Shadows	800	147
-	Total	7200	35576

the city of Pavia, Italy. The size of the data is 610 × 340 pixels with 1.3 m spatial resolution. And the scene comprises 103 spectral bands covering the range from 0.43 to 0.86 μm including nine classes. The numbers of training and testing samples are listed in Table II.

The Xiong-An data set is collected by a full-spectrum multimodal imaging spectrometer for a high-resolution special aviation system developed by the Shanghai Institute of Technical Physics in China. The scene has 1580 × 3750 pixels, including 19 classes. It consists of 250 bands ranging from 400 to 1000 nm, and the spatial resolution is 0.5 m. Among them, we cropped 1003×703 size images including 11 classes as experimental data. For each class, 1000 sample points were selected for training, and some sample points in the rest are used for testing, as shown in Table IV.

Also, our approach without clustering is based on the samples listed in Tables I, II, and IV. However, the clustering-based method requires twofold samples to ensure that there are enough samples for each subclass. Taking the PaviaU data set as an example, as shown in Table III, the total number of samples is twice the original data, but the number of training samples for each class is the same as the original data. The rest of the data are used as test samples.

TABLE III
NUMBER OF TRAINING AND TESTING SAMPLES FOR THE UNIVERSITY OF PAVIA DATA SET WITH CLUSTERING

#	Class	Training	Test
1	Asphalt	800	5831
2	Meadows	800	17849
3	Gravel	800	1299
4	Trees	800	2264
5	Painted metal sheets	800	950
6	Painted metal sheets2	800	140
7	Bare Soil	800	1040
8	Bare Soil2	800	7418
9	Bitumen	800	530
10	Self-Blocking Bricks	800	2882
11	Shadows	800	147
-	Total	8800	40350

TABLE IV
NUMBER OF TRAINING AND TESTING SAMPLES FOR THE XIONG-AN DATA SET WITHOUT CLUSTERING

#	Class	Training	Test
1	Acer negundo	1000	4120
2	Sophora japonica	1000	1223
3	water	1000	1975
4	corn	1000	3121
5	pear tree	1000	2026
6	poplar	1000	8107
7	vegetable plot	1000	5043
8	half bare land	1000	496
9	grass	1000	9487
10	peach tree	1000	7786
11	buildings	1000	5803
-	Total	11000	49187

B. Experimental Analysis

HSIs are difficult to be available due to the imaging conditions, and only a few labeled data can be used in experiments. Especially for the samples after automatic clustering, each subclass contains less data than the original classes. However, deep networks usually require a large number of samples for training. To tackle this issue, we use an effective method to achieve data augmentation. For the pixels of each subclass with a small sample size, we first copy it and then add random noise (± 3) to the original data, so that the number of samples can be increased by a factor of two. Apart from this, if the spectral variation within the class can be reduced effectively by the clustering module, the impact of the small sample size problem can also be reduced. In this way, we can ensure that there are enough samples to learn and accurately estimate a large number of parameters in the network.

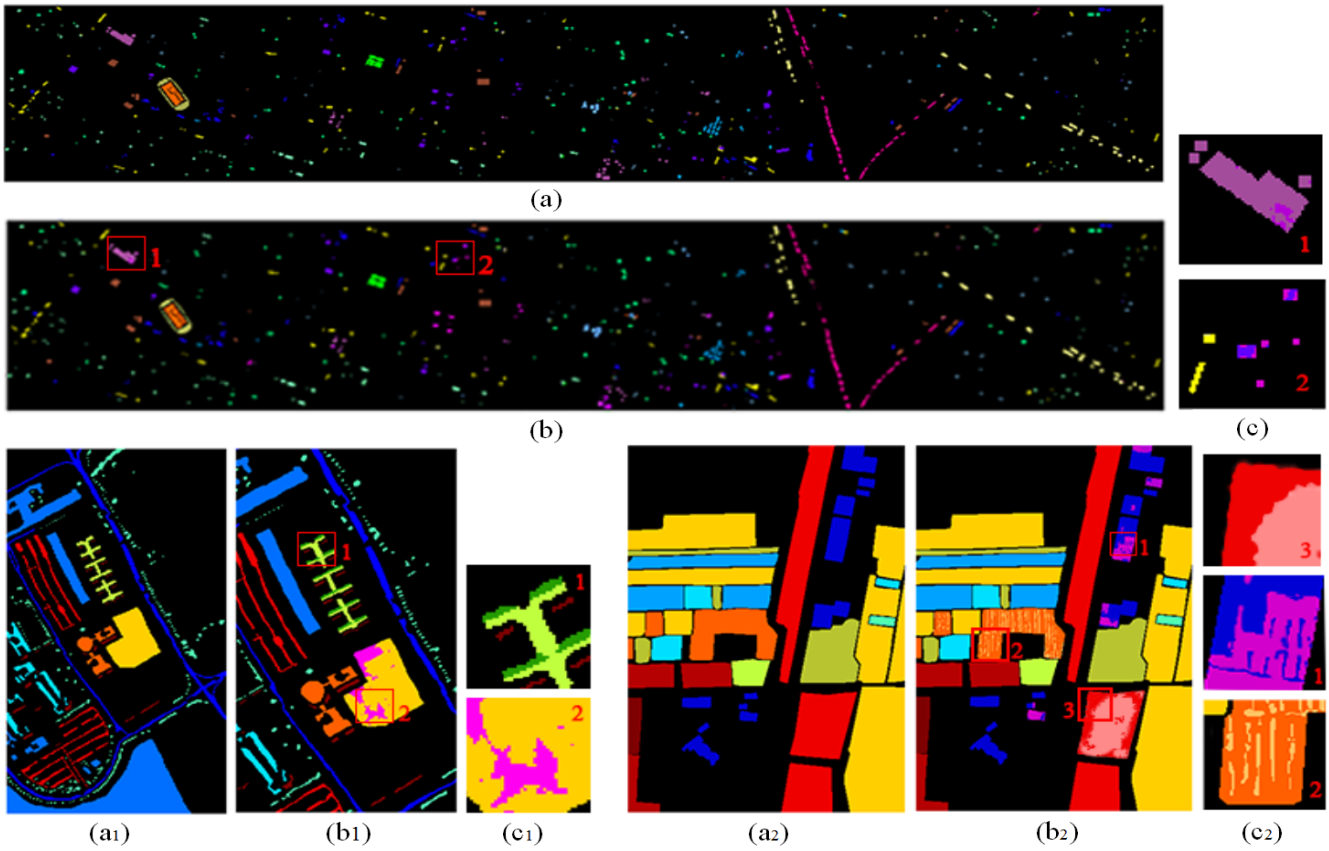

Fig. 8. Clustering result is displayed. (a), (a1), (a2) represent the original ground truth maps of the data sets in experiment. (b), (b1), (b2) are the new ground truth maps after clustering. (c), (c1), (c2) is the detailed images of clustering.

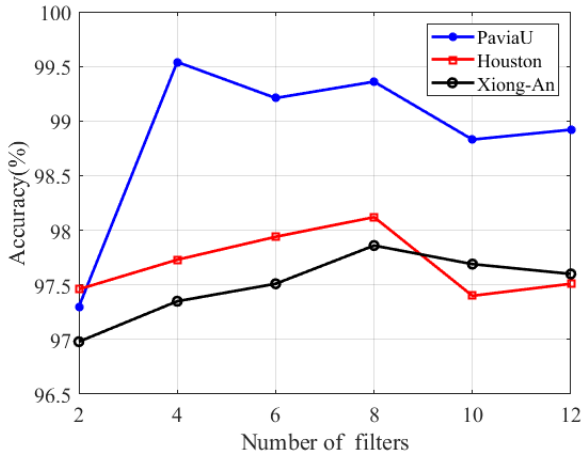

Fig. 9. Overall classification accuracy (%) versus different number of sinc function filters for three data sets.

For each pixel, SincNet treats each pixel spectral vector as a 1-D input signal, and DS-CNN selects 13×13 pixels surrounding it as the input data patch. The image patch size is set based on experience. Although the size 13×13 may not be the best window size for all the training data set, it also makes the classification result express the higher accuracy.

Besides, the learning rate is also a key factor to improve training efficiency and classification accuracy. In our approach, the learning rate is set to be dynamic, ranging from 0.1 to 0.001. We set up 300 training epochs for each sample data set. For every 100 epochs, the learning rate drops by an order of magnitude. For example, in the first 100 epochs, the learning rate is set to 0.1 and in the range of 200–300 epochs, the learning rate is set to 0.01. The larger the learning rate, the greater the parameters change in each training. In the early epochs, a large learning rate can speed up training efficiency. But in the later epochs, a small learning rate can prevent the network from overfitting and fit the image more accurately.

As one of the important parameters of the whole classification framework, the number of sinc filters always influences the classification accuracy and computational complexity. However, as the number of sinc filters increases, the classification accuracy cannot get effectively increase, but the computational complexity has increased exponentially. As shown in Fig. 9, we conducted ablation experiments on the influence of the number of sinc function filters on the classification effect. The number of sinc filters is set in the range of $\{2, 4, 6, 8, 10, 12\}$. We can see that for Houston and Xiong-An data sets, when the number of sinc filters is 8, the accuracy of classification is highest. Although the accuracy of the PaviaU data set based on eight sinc filters is not the highest, it can still get a good classification result. Therefore, we use 8 as the sinc function filters in our following experiments.

C. Classification Performance

To illustrate the performance of the proposed automatic clustering-based two branch CNN method, we conducted

TABLE V
CLASS SPECIFIC AND OVERALL CLASSIFICATION ACCURACY (%) OF DIFFERENT METHODS FOR THE HOUSTON DATA

No.	Class	CNN-PPF	CD-CNN	DHCNet	R-PCA	MS-CNN	SSUN	Proposed(nc)	Proposed
1	Water	95.31	95.97	98.49	93.58	99.30	97.84	98.96(±0.37)	**100(±0)**
2	Health grass	92.89	96.14	92.41	83.60	94.22	**98.75**	97.32(±0.66)	97.59(±0.33)
3	Railway	87.04	82.41	97.34	80.72	**100**	96.71	97.15(±0.25)	99.05(±0.17)
4	Commercial	94.98	90.27	93.89	82.46	93.09	97.09	96.18(±0.83)	**98.58(±0.16)**
5	Parking lot 2	85.32	86.38	90.86	85.63	**98.17**	96.73	97.76(±0.29)	96.21(±0.67)
6	Stressed grass	98.67	98.01	97.55	89.84	93.12	98.34	96.58(±0.16)	**99.28(±0.10)**
7	Highway	90.92	87.46	89.16	85.95	82.44	95.50	96.34(±0.41)	**97.63(±0.40)**
8	Parking lot 1	89.64	82.43	91.66	84.79	89.34	95.19	95.55(±0.34)	**98.40(±0.15)**
9	Synthetic grass	86.57	99.12	**99.71**	84.79	89.67	98.99	99.30(±0.07)	99.21(±0.12)
10	Residential	86.91	89.23	93.47	84.55	92.75	96.09	95.22(±0.23)	**96.12(±0.53)**
11	Soil	98.10	97.82	99.92	90.36	95.24	97.86	99.03(±0.22)	**100(±0)**
12	Road	94.70	94.75	90.13	82.11	89.53	**95.28**	93.95(±0.54)	94.33(±0.32)
13	Trees	97.73	96.74	95.49	84.08	92.28	97.50	96.38(±0.33)	**97.98(±0.38)**
14	Running track	99.29	99.42	99.57	86.35	99.36	99.23	99.16(±0.18)	**99.59(±0.13)**
15	Tennis court	96.76	96.38	98.41	85.64	96.76	97.29	**99.49(±0.05)**	97.34(±0.59)
AA		92.99	92.84	95.20	85.95	93.36	97.29	97.22(±0.32)	**98.09(±0.42)**
OA		94.18	90.69	97.15	85.34	94.17	97.21	96.96(±0.45)	**98.23(±0.34)**

TABLE VI
CLASS SPECIFIC AND OVERALL CLASSIFICATION ACCURACY (%) OF DIFFERENT METHODS FOR THE PAVIAU DATA

No.	Class	CNN-PPF	CD-CNN	DHCNet	R-PCA	MS-CNN	SSUN	Proposed(nc)	Proposed
1	Asphalt	97.48	93.60	98.36	92.88	98.69	99.64	99.12(±0.32)	**99.88(±0.07)**
2	Meadows	95.35	97.46	98.47	95.61	**99.21**	98.80	98.77(±0.24)	99.12(±0.43)
3	Gravel	89.52	81.83	99.68	91.97	99.45	99.71	**100(±0)**	99.44(±0.25)
4	Trees	97.03	96.72	93.92	94.55	98.75	**99.88**	**99.88(±0.05)**	99.72(±0.25)
5	sheets	94.48	99.55	99.82	**100**	98.39	99.82	99.81(±0.12)	**100(±0)**
6	Bare Soil	92.65	94.84	98.93	93.18	99.83	99.74	99.33(±0.14)	**99.94(±0.06)**
7	Bitumen	96.42	91.98	99.99	97.32	99.93	99.96	99.98(±0.02)	**100(±0)**
8	Bricks	93.40	87.44	99.51	94.43	96.99	98.96	**99.77(±0.08)**	99.76(±0.13)
9	Shadows	99.87	99.82	99.89	99.86	**100**	99.68	99.38(±0.25)	**100(±0)**
AA		95.33	93.69	98.73	95.42	99.26	99.53	99.56(±0.19)	**99.74(±0.24)**
OA		96.27	94.55	99.05	96.95	99.13	99.45	99.49(±0.21)	**99.68(±0.31)**

TABLE VII
CLASS SPECIFIC AND OVERALL CLASSIFICATION ACCURACY (%) OF DIFFERENT METHODS FOR THE XIONG-AN DATA

No.	Class	CNN-PPF	CD-CNN	DHCNet	R-PCA	MS-CNN	SSUN	Proposed(nc)	Proposed
1	Acer negundo	99.60	94.82	98.48	92.82	97.66	98.90	99.22(±0.15)	**99.93(±0.02)**
2	Locust tree	97.88	96.15	96.72	88.31	95.34	**98.61**	95.73(±0.21)	96.35(±0.31)
3	water	95.97	99.32	97.55	91.53	98.27	98.65	95.52(±0.38)	**99.63(±0.17)**
4	corn	96.34	92.09	97.97	83.02	93.20	96.43	96.21(±0.22)	**98.74(±0.23)**
5	pear tree	95.45	96.64	95.86	85.44	89.93	96.03	95.16(±0.46)	**96.69(±0.52)**
6	poplar	96.51	96.80	96.87	89.63	97.25	97.06	97.49(±0.24)	**98.08(±0.36)**
7	vegetable	98.72	87.53	97.98	90.58	98.74	98.52	99.53(±0.07)	**99.89(±0.03)**
8	bare land	94.25	73.50	99.32	89.62	99.50	96.34	**99.82(±0.18)**	99.25(±0.17)
9	grass	**98.96**	87.25	97.87	88.37	93.12	96.55	96.62(±0.33)	98.16(±0.25)
10	peach tree	94.15	92.74	96.68	88.79	92.35	**97.05**	96.23(±0.51)	96.74(±0.83)
11	buildings	96.39	99.03	97.64	91.65	96.79	98.17	98.06(±0.37)	**100(±0)**
AA		96.75	92.35	97.63	89.07	95.65	97.17	97.23(±0.42)	**98.38(±0.22)**
OA		95.82	96.04	96.48	89.85	96.36	96.55	96.72(±0.37)	**97.86(±0.34)**

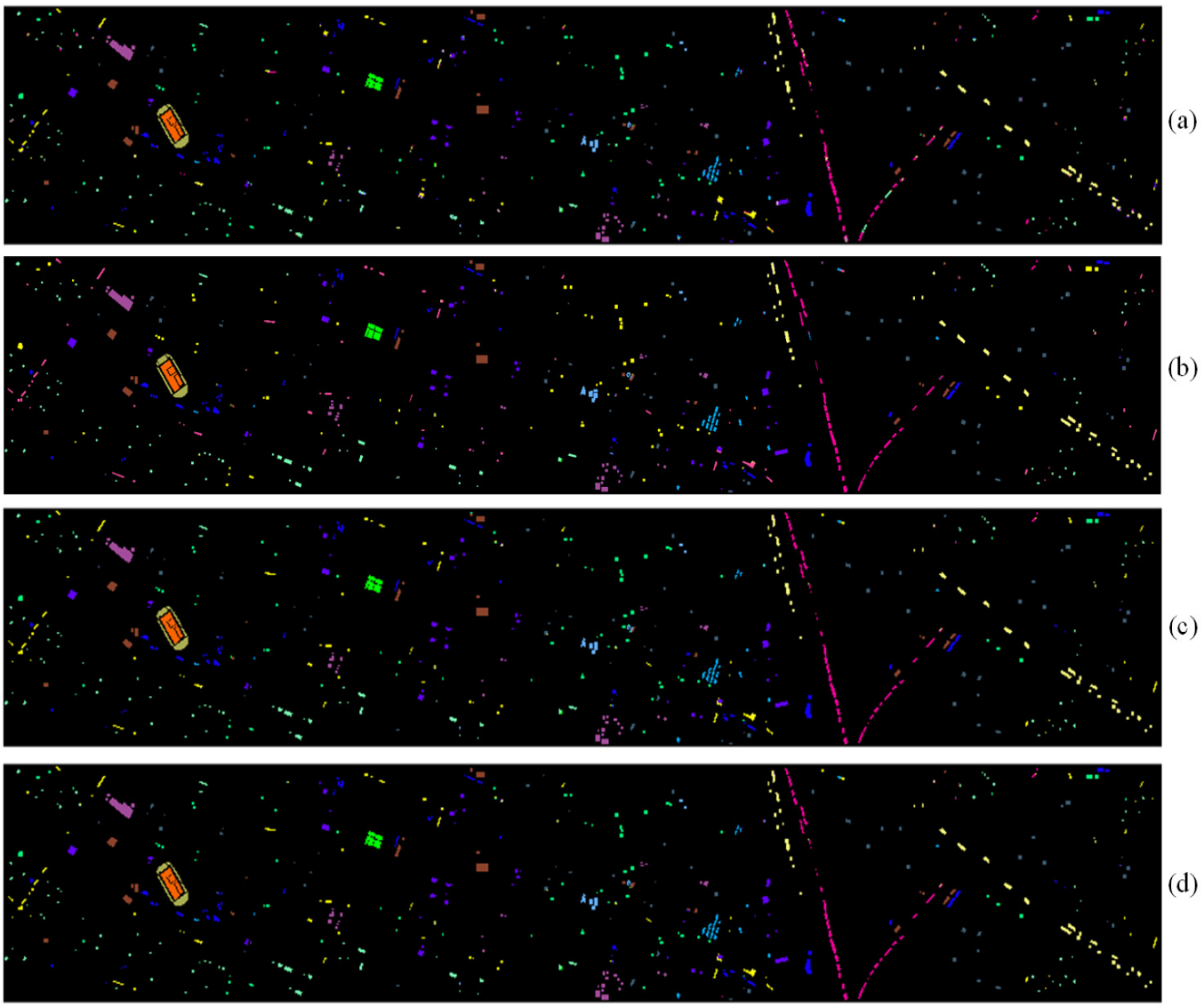

Fig. 10. Classification maps for the University of Houston data based on different methods. (a) CD-CNN:90.69%. (b) MS-CNN:94.17%. (c) DHCNet:97.15%. (d) Proposed:98.12%.

comparative experiments and compared them with some state-of-the-art HSI classification approaches such as CNN with pixel-pair features (CNN-PPF) [62], contextual deep CNN (CD-CNN) [63], deformable HSI classification networks (DHCNets) [64], R-PCA CNN [28], MS-CNN [4], and SSUN [32]. The experiments treat class-specific results accuracy, average accuracy (AA), and overall accuracy (OA) as evaluation indicators, which are listed in Tables V–VII. Among them, the "Proposed (NC)" represents that the data without clustering is input into two-branch CNN. And the "Proposed" represents that we cluster the data before giving it to the proposed two-branch CNN. For our approach, the experiments are repeated ten times by randomly selected training samples, and the average results with standard variations are reported.

The HSI classification CNN developed in the early stage can hierarchically construct high-level features in an automated way. But they lack the method of rational use of spatial information. And for the existing state-of-the-art methods, they have been able to extract enough spatial–spectral information through different technology, The DHCNet can earn 3% improvement from the perspective of OA compared with the classical CNN-PPF. They can even use images from other sensors to increase the required information (such as MS-CNN). However, since a one-to-one correspondence cannot be made between the handcraft labeled semantic class and the actual feature class, it becomes a key issue that reducing the spectral difference within a class. Consequently, clustering preprocessing is very critical.

The clustering results are shown in Fig. 8. For each image, some classes with greater variation were divided into smaller subclasses after clustering. Among them, (b), (b1), (b2) are the new ground-truth maps of each data obtained, and (c), (c1), (c2) are the detailed display. In the clustering process, due to hardware limitations and a large amount of calculation,

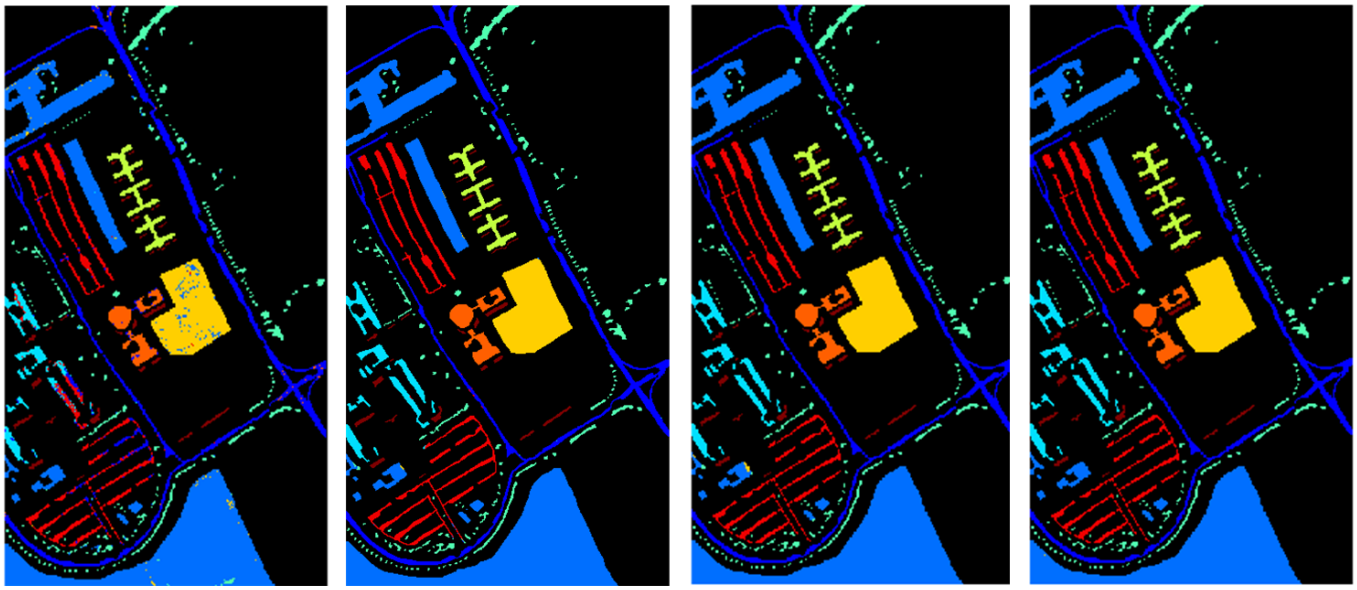

Fig. 11. Classification maps for the University of Pavia data based on different methods. (a) CD-CNN:94.55%. (b) MS-CNN:99.13%. (c) DHCNet:99.25%. (d) Proposed:99.82%.

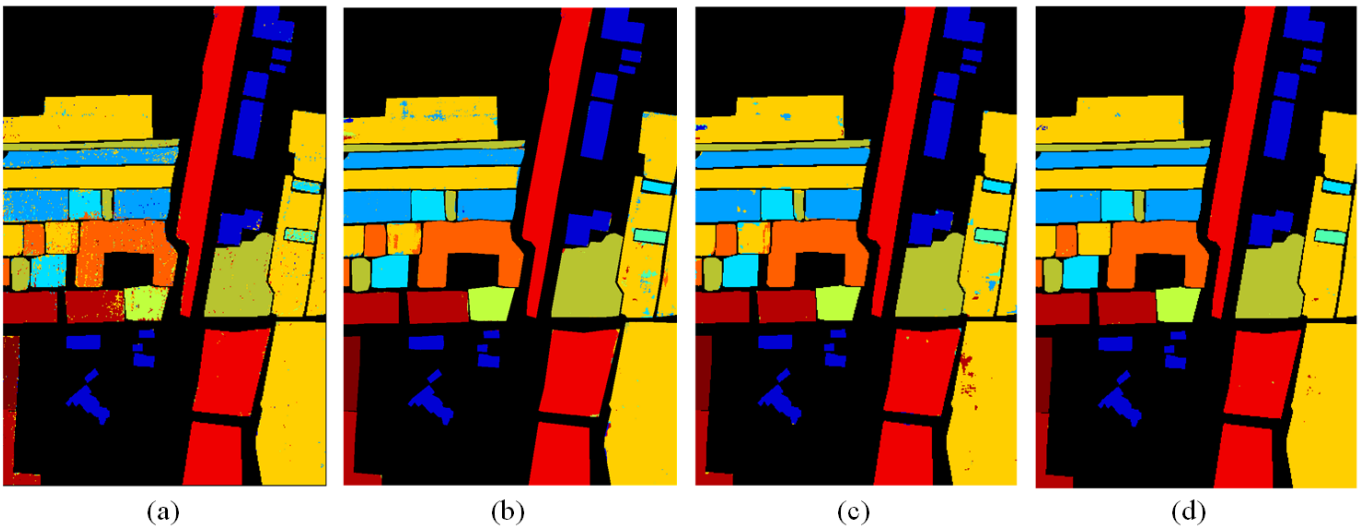

Fig. 12. Classification maps for the Xiong-An data based on different methods. (a) CD-CNN:96.04%. (b) MS-CNN:96.36%. (c) DHCNet:96.48%. (d) Proposed:97.86%.

we crop the image by classes and cluster them separately. The data set of Houston was divided into 17 classes from 15 classes. The classes of "Soil" and "Parking lot 1" were, respectively, clustered into two subclasses. The data set of PaviaU was divided into 9 classes from 11 classes. The classes of "Bare soils" and "Sheets" were, respectively, clustered into two subclasses. The data set of Xiong-An was divided into 11 classes from 14 classes. The classes of "Water," "Corn," and "Buildings" were, respectively, clustered into two subclasses. For the experimental results of each data, we can see the superiority of the clustering in the last two columns of Tables V–VII. For example, the accuracy of classes with a large difference (e.g., Soil and Parking lot 1 in Houston data) is higher and more stable than other methods. And the proposed method with clustering yields accuracy 98.23%, nearly 1.5% higher than that of the method without clustering (96.96%).

At the same time, it can be seen from the three tables that our approaches have better classification performance than other methods. Taking the Houston data as an example, the proposed method yields accuracy 98.23%, nearly 8% higher than CD-CNN (i.e., 90.69%), and approximately 4% higher than that of the CNN-PPF (i.e., 94.18%). Therefore, it can be concluded that the proposed two-branch method can result in higher classification accuracy and the clustering is

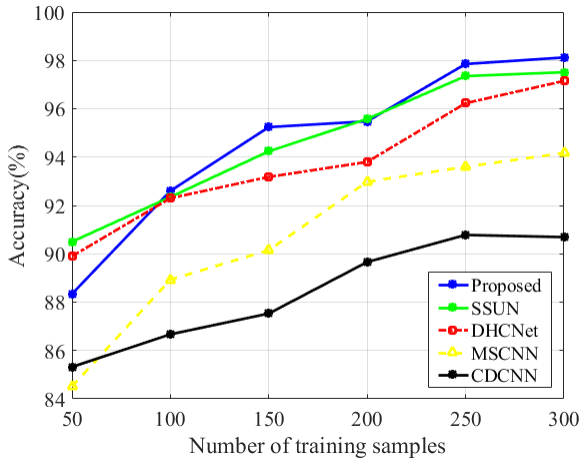

Fig. 13. Overall classification accuracy (%) versus different numbers of training samples per class for different methods in Houston data set.

a good strategy to reduce the within-class difference. We set the number of training samples of all methods to be the same, and the adjustment of parameters may cause the experimental results to be slightly different from those in other literature. Figs. 10–12 illustrate the classification map, and the visual result corresponds to the results in Tables V–VII. The ground cover maps provide test results overlaid on the original ground truth maps, and it clearly shows that the classification maps of our approach are less noisy than those of methods like CD-CNN and MS-CNN, e.g., taking the Xiong-An as an example, the areas of the pear tree have less mislabeled pixels.

Besides, we have explored the classification performances of our approach with different numbers of training samples. The experimental results of four methods based on the Houston data set are illustrated in Fig. 13. The number of training samples per class changed from 50 to 300 with an interval of 50. We can see that for all methods, the more the training samples, the higher the classification accuracy. And when the number of training samples is in the range of 150–200, the classification accuracy can maintain a relatively high steady state. When the number of training samples is more than 100, our approach can always maintain the best classification performance.

The computational complexity of the training and testing procedure of the proposed two-branch CNN is summarized in Table VIII. For the training procedure, all the data sets are trained in 300 epochs. Although the PCA preprocessing has relatively reduced the computational cost, the overall system is still very time-consuming. Because the clustering module increases the number of subclasses, and we also expand the data of some subclasses with a small sample size. At the same time, the two-branch complex network also brings a computational burden. Although the training procedure takes a much longer time, it can also be observed that the testing is relatively faster.

D. Discussion

With the help of the clustering module and SincNet, the intraclass difference brought by spectral variation can be reduced. And we can also observe from Tables V–VII that our method can indeed improve the classification accuracy for classes with large intraclass differences in the original image (such as the classes of Soil and Parking lot 1 in Houston Data). Unfortunately, although the intraclass difference is quite large for some classes, each subclass is also different from other classes. In view of this situation, our approach has no great advantage. In addition, our method does not make too much effort to reduce computational complexity, and we need further research.

TABLE VIII
DETAILED RUNNING TIME OF OUR APPROACH ON THREE DATA (m: MINUTES s: SECONDS)

	Houston	PaviaU	Xiong-An
Training(m)	6.85(±0.08)	7.28(±0.06)	8.39(±0.13)
Test(s)	23.14(±0.70)	47.55(±0.74)	64.46(±0.89)

IV. CONCLUSION

In this article, a novel automatic clustering-based two-branch CNN is proposed for HSI classification. First, a clustering module is utilized to reduce the intraclass variation. Second, DS-CNN is proposed to extract more meaningful spatial information. Here, the double-stripe convolution kernel can collect contextual interactional features in a specific direction. Third, we modify the SincNet for HSI classification to give more weight to the spectral shape, so that the finer spectral pattern can be captured. The advantages of our approach come from two aspects: the automatic clustering module can effectively reduce intraclass differences and the two-branch CNN can extract spectral and spatial features in a targeted manner. Experimental results show that the proposed clustering-based two-branch method outperforms other state-of-the-art methods on three data sets.

REFERENCES

[1] H. Su, Y. Yu, Z. Wu, and Q. Du, "Random subspace-based k-nearest class collaborative representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, early access, Oct. 22, 2020, doi: 10.1109/TGRS.2020.3029578.

[2] C.-I. Chang, Y.-M. Kuo, S. Chen, C.-C. Liang, K. Y. Ma, and P. F. Hu, "Self-mutual information-based band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, early access, Oct. 2, 2020, doi: 10.1109/TGRS.2020.3024602.

[3] J. Sun *et al.*, "Deep clustering with intraclass distance constraint for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, early access, Sep. 25, 2020, doi: 10.1109/TGRS.2020.3019313.

[4] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multisource remote sensing data classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 937–949, Feb. 2018.

[5] Q. Xie, M. Zhou, Q. Zhao, Z. Xu, and D. Meng, "MHF-net: An interpretable deep network for multispectral and hyperspectral image fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Aug. 11, 2020, doi: 10.1109/TPAMI.2020.3015691.

[6] B. Xi, J. Li, Y. Li, R. Song, W. Sun, and Q. Du, "Multiscale context-aware ensemble deep KELM for efficient hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, early acces, Sep. 22, 2020, doi: 10.1109/TGRS.2020.3022029.

[7] M. Wang, Q. Wang, J. Chanussot, and D. Li, "Hyperspectral image mixed noise removal based on multidirectional low-rank modeling and spatial–spectral total variation," *IEEE Trans. Geosci. Remote Sens.*, early access, May 27, 2020, doi: 10.1109/TGRS.2020.2993631.

[8] W. Li, J. Liu, and Q. Du, "Sparse and low-rank graph for discriminant analysis of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4094–4105, Jul. 2016.

[9] W. Lv and X. Wang, "Overview of hyperspectral image classification," *J. Sensors*, vol. 2020, no. 2, pp. 1–13, 2020.

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

[11] J. Liu, Z. Wu, J. Li, L. Xiao, A. Plaza, and J. A. Benediktsson, "Spatial–Spectral hyperspectral image classification using random multiscale representation," *IEEE J. Sel. Topics Appl. Earth Observatr. Remote Sens.*, vol. 9, no. 9, pp. 4129–4141, Sep. 2016, doi: 10.1109/JSTARS.2016.2587678.

[12] N. He *et al.*, "Feature extraction with multiscale covariance maps for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 755–769, Feb. 2019, doi: 10.1109/TGRS.2018.2860464.

[13] B. Xue *et al.*, "A subpixel target detection approach to hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5093–5114, Sep. 2017, doi: 10.1109/TGRS.2017.2702197.

[14] N. Falco, J. A. Benediktsson, and L. Bruzzone, "Spectral and spatial classification of hyperspectral images based on ICA and reduced morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6223–6240, Nov. 2015, doi: 10.1109/TGRS.2015.2436335.

[15] K. Kavitha and D. S. Arivazhagan, "A novel feature derivation technique for SVM based hyper spectral image classification," *Int. J. Comput. Appl.*, vol. 1, no. 15, pp. 27–34, Feb. 2010.

[16] H. Zhang, L. Liu, W. He, and L. Zhang, "Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3071–3084, May 2020, doi: 10.1109/TGRS.2019.2947333.

[17] X. Guo, X. Huang, L. Zhang, L. Zhang, A. Plaza, and J. A. Benediktsson, "Support tensor machines for classification of hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3248–3264, Jun. 2016.

[18] J. Liu, Z. Wu, L. Xiao, J. Sun, and H. Yan, "Generalized tensor regression for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1244–1258, Feb. 2020, doi: 10.1109/TGRS.2019.2944989.

[19] J. Xia, J. Chanussot, P. Du, and X. He, "Rotation-based support vector machine ensemble in classification of hyperspectral data with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1519–1531, Mar. 2016.

[20] D. Li, Q. Wang, and F. Kong, "Adaptive kernel sparse representation based on multiple feature learning for hyperspectral image classification," *Neurocomputing*, vol. 400, pp. 97–112, Aug. 2020.

[21] R. Wang, F. Nie, Z. Wang, F. He, and X. Li, "Scalable graph-based clustering with nonnegative relaxation for large hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7352–7364, Oct. 2019, doi: 10.1109/TGRS.2019.2913004.

[22] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3672–3684, Jun. 2016, doi: 10.1109/TGRS.2016.2524557.

[23] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.

[24] L. Lin and X. Song, "Using CNN to classify hyperspectral data based on spatial-spectral information," in *Advances in Intelligent Information Hiding and Multimedia Signal Processing*. Cham, Switzerland: Springer, 2017, pp. 61–68.

[25] V. Slavkovikj *et al.*, "Hyperspectral image classification with convolutional neural networks," in *Proc. 23rd ACM Int. Conf.*, 2015, pp. 1159–1162.

[26] J. Yue, W. Zhao, S. Mao, and H. Liu, "Spectral–spatial classification of hyperspectral images using deep convolutional neural networks," *Remote Sens. Lett.*, vol. 6, no. 6, pp. 468–477, 2015.

[27] 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.

[28] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 4959–4962.

[29] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, Aug. 2019.

[30] H. Gao, Y. Yang, S. Lei, C. Li, H. Zhou, and X. Qu, "Multi-branch fusion network for hyperspectral image classification," *Knowl. Based Syst.*, vol. 167, pp. 11–25, Mar. 2019.

[31] 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, 2019.

[32] Y. Xu, L. Zhang, B. Du, and F. Zhang, "Spectral–spatial unified networks for hyperspectral image classification," *IEEE Trans. Geoence Remote Sens.*, vol. 56, no. 10, pp. 5893–5909, May 2018.

[33] J. Yang, Y. Zhao, J. C.-W. Chan, and C. Yi, "Hyperspectral image classification using two-channel deep convolutional neural network," in *Proc. IEEE Int. Symp. Geosci. Remote Sens. (IGARSS)*, Jul. 2016, pp. 5079–5082.

[34] 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.

[35] S. Mei, Q. Bi, J. Ji, J. Hou, and Q. Du, "Spectral variation alleviation by low-rank matrix approximation for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 796–800, Jun. 2016.

[36] A. Zare and K. C. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2014.

[37] G. A. Shaw and H.-H. K. Burke, "Spectral imaging for remote sensing," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 3–28, 2003.

[38] S. Jay *et al.*, "Hyperspectral remote sensing of shallow waters: Considering environmental noise and bottom intra-class variability for modeling and inversion of water reflectance," *Remote Sens. Environ.*, vol. 200, pp. 352–367, Oct. 2017.

[39] J. Theiler, A. Ziemann, S. Matteoli, and M. Diani, "Spectral variability of remotely sensed target materials: Causes, models, and strategies for mitigation and robust exploitation," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 8–30, Jun. 2019.

[40] Y. Kong, X. Wang, and Y. Cheng, "Spectral–spatial feature extraction for HSI classification based on supervised hypergraph and sample expanded CNN," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4128–4140, Nov. 2018.

[41] G. Karypis, R. Aggarwal, V. Kumar, and S. Shekhar, "Multilevel hypergraph partitioning: Applications in VLSI domain," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 7, no. 1, pp. 69–79, Mar. 1999.

[42] N. Selvakkumaran, and G. Karypis, "Multi-objective hypergraph partitioning algorithms for cut and maximum subdomain degree minimization," in *Proc. Int. Conf. Comput. Aided Des. (ICCAD)*, Nov. 2003, pp. 726–733,

[43] A. Abou-Rjeili and G. Karypis, "Multilevel algorithms for partitioning power-law graphs," in *Proc. 20th IEEE Int. Parallel Distrib. Process. Symp.*, Apr. 2006, p. 10

[44] H. T. X. Doan and G. M. Foody, "Reducing the impacts of intra-class spectral variability on soft classification and its implications for super-resolution mapping," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2007, pp. 2585–2588.

[45] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.

[46] S. S. Sawant, M. Prabukumar, and S. Samiappan, "A band selection method for hyperspectral image classification based on cuckoo search algorithm with correlation based initialization," in *Proc. 10th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens. (WHISPERS)*, Amsterdam, The Netherlands, Sep. 2019, pp. 1–4, doi: 10.1109/WHISPERS.2019.8920950.

[47] L. Ji, X. Geng, Y. Zhao, and P. Gong, "Impact of initialization on nonnegative matrix fraction for endmember extraction for hyperspectral imagery," in *Proc. 8th Workshop Hyperspectral Image Signal Processing: Evol. Remote Sens. (WHISPERS)*, Los Angeles, CA, USA, Aug. 2016, pp. 1–5, doi: 10.1109/WHISPERS.2016.8071685.

[48] X. Tan, Y. Zhang, S. Tang, J. Shao, F. Wu, and Y. Zhuang, "Logistic tensor regression for classification," in *Proc. Int. Conf. Intell. Sci. Intell. Data Eng.*, 2012, pp. 573–581.

[49] Q. Li and D. Schonfeld, "Multilinear discriminant analysis for higher-order tensor data classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2524–2537, Dec. 2014.

[50] K. Makantasis, A. D. Doulamis, N. D. Doulamis, and A. Nikitakis, "Tensor-based classification models for hyperspectral data analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 6884–6898, Dec. 2018.

[51] K. Makantasis, A. Voulodimos, A. Doulamis, N. Doulamis, and I. Georgoulas, "Hyperspectral image classification with tensor-based rank-R learning models," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 3125–3148.
[52] M. Ravanelli and Y. Bengio, "Interpretable convolutional filters with SincNet," 2018 *arXiv:1811.09725*. [Online]. Available: https://arxiv.org/abs/1811.09725
[53] R. Parhizkar, "Euclidean distance matrices," in *Proc. EPFL*, 2013, pp. 108–117.
[54] H. Mittelmann and J. Peng, "Estimating bounds for quadratic assignment problems associated with Hamming and manhattan distance matrices based on semidefinite programming," *SIAM J. Optim.*, vol. 20, no. 6, pp. 3408–3426, 2010.
[55] T. Hsing and R. Eubank, "Sparse principal components analysis," in *Theoretical Foundations of Functional Data Analysis, With an Introduction to Linear Operators*. Hoboken, NJ, USA: Wiley, 2015.
[56] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[57] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, Haifa, Israel, Jun. 2010, pp. 21–24.
[58] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. 30th ICML*, 2013, vol. 30, no. 1, p. 3.
[59] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167
[60] M. Ravanelli and Y. Bengio, "Speaker recognition from raw waveform with SincNet," in *Proc. IEEE Spoken Lang. Technol. Workshop (SLT)*, Dec. 2018, pp. 1021–1028.
[61] L. R. Rabiner and R. W. Schafer, *Theory and Applications of Digital Speech Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2011.
[62] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, Feb. 2017.
[63] 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.
[64] J. Zhu, L. Fang, and P. Ghamisi, "Deformable convolutional neural networks for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 8, pp. 1254–1258, Aug. 2018.

Yuan Li received the B.S. degree from the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China, in 2018, where she is pursuing the M.S. degree.

Her research interests include remote sensing image process and pattern recognition.

Qizhi Xu (Member, IEEE) received the B.S. degree from Jiangxi Normal University, Nanchang, China, in 2005, and the Ph.D. degree from Beihang University, Beijing, China, in 2012.

He was a Post-Doctoral Fellow with the University of New Brunswick, Fredericton, NB, Canada. He is an Associate Professor with the School of Mechatronical Engineering, Beijing Institute of Technology, Beijing. His research interests include image fusion, image understanding, and big data analysis of remote sensing.

Dr. Xu was a recipient of the Technological Invention Award First Prize from the Chinese Institute of Electronics for his image fusion research in 2017.

Wei Li (Senior Member, IEEE) received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2007, the M.S. degree in information science and technology from Sun Yat-Sen University, Guangzhou, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2012.

Subsequently, he spent one year as a Post-Doctoral Researcher with the University of California, Davis, CA, USA. He was a Professor with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, from 2013 to 2019. He is a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing. His research interests include hyperspectral image analysis, pattern recognition, and data compression.

Dr. Li received the 2015 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS) for his service for IEEE JSTARS and the Outstanding Paper award at the IEEE International Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (Whispers), 2019. He is serving as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He has served as a Guest Editor for the Special Issue of *Journal of Real-Time Image Processing*, *Remote Sensing*, and IEEE JSTARS.

Jinyan Nie received the B.S. degree in electronic information engineering from the Wuhan University of Science of Technology, Wuhan, China, in 2015, and the M.S. degree in signal and information processing from Central China Normal University, Wuhan, China, in 2018. She is pursuing the Ph.D. degree with the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China.

Her research interests include hyperspectral remote sensing image analysis, image fusion, and deep learning.