Intraclass Similarity Structure Representation-Based Hyperspectral Imagery Classification With Few Samples

Wei Wei, *Senior Member, IEEE*, Lei Zhang, *Student Member, IEEE*, Yu Li, Cong Wang, and Yanning Zhang, *Senior Member, IEEE*

***Abstract*—Hyperspectral imagery (HSI) classification is one of the fundamental applications in remote sensing domain, which aims at predicting the labels of unlabeled pixels in an image with a classifier trained on a certain amount of labeled pixels. However, due to the expensive cost on manual labeling, only limited labeled pixels can be obtained in real applications, which is prone to result in the training of classifier to be overfitting. To address this problem, we present an intraclass similarity structure representation-based HSI classification method. First, according to the intraclass spectrum similarity of pixels, we establish a mixed labels-based annotation model. Given some randomly selected unlabeled pixels, we employ the proposed annotation model to assign each pixel a mixed label from the top-two possible classes, and then augment the original training set with those labeled pixels. On the augmented training set, we train a deep convolutional neural network-based classification model. With several individual rounds of the annotation and classifier training procedures, we obtain several independent classification models and predict the final labels as their fusion results with a voting strategy. Experimental results demonstrate the effectiveness of the proposed method in terms of HSI classification with few training samples.**

***Index Terms*—Classifier fusion, few samples learning, hyperspectral imagery classification, intra-class similarity structure representation.**

I. Introduction

HYPERSPECTRAL imagery (HSI) is a digital image that records the reflectivity of natural scene under different spectral irradiation frequency with high spectral resolution [1], [2]. Unlike color or grayscale images, HSI usually contains hundreds or even thousands of bands, and each pixel records the spectrum curve of the corresponding object. Since the spectrum is varied with different substance, it can be exploited to distinguish the substances in the imaging scene [3]. Thus, HSI has been widely used in lots of fields including mineral mining [4], environmental detection [5], and military defense [6], etc.

As one of the fundamental applications in remote sensing domain [7], [8], HSI classification aims to train the classifier with small amount of labeled pixels, and then predict the labels of unlabeled pixels, with which the distribution of different land covers in HSI can be obtained [8], [9]. At present, a large number of HSI classification methods [8] have been proposed successively. Based on the structure of features, the existing HSI classification methods can be divided into shallow feature-based method and deep neural network based one. In the shallow feature-based method, feature extraction is independent with the classifier design. The features can be the spectral information themselves [8], [9], or acquired from the spectral information via unsupervised learning methods such as band selection [10], [11], feature transformation [12], dictionary learning [2], [13], and image decomposition [14]. Then based on the obtained features, commonly used classifiers such as support vector machine (SVM) [9], nearest neighbor classifier (NNC) [15], and logistic regression classifier (LRC) [16] are then used to predict the labels. The deep neural network-based method is different from the shallow feature-based method which divides feature extraction and classifier learning into two independent modules in that it implicitly embeds these two modules into a unified framework using network structure [17]. Due to the powerful representation ability as well as jointly optimizing feature extraction and classifier for training, deep neural network-based method can learn more discriminative features and classifiers when the number of labeled training samples are sufficient, with which the accuracy of the classification task [18], [19], can be significantly improved. Thus, deep neural network-based HSI classification method has become a hot topic to date [20]. However, in the practical application, the assumption of providing sufficient training samples is always infeasible, i.e., only few labeled training samples can be obtained. As a result, the method based on deep neural network is prone to be overfitting when giving only few labeled

Manuscript received December 10, 2019; revised February 13, 2020; accepted February 18, 2020. Date of publication March 10, 2020; date of current version March 25, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61671385 and in part by the Science, Technology and Innovation Commission of Shenzhen Manicipality under Grant JCYJ20190806160210899. *(Corresponding author: Lei Zhang.)*

Wei Wei is with the Research and Development Institute of Northwestern Polytechnical University in Shenzhen, Northwestern Polytechnical University, Shenzhen 518057, China, and also with the School of Computer Science, National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, the Shaanxi Provincial Key Laboratory of Speech and Image Information Processing, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: weiwin1979@hotmail.com).

Lei Zhang, Yu Li, Cong Wang, and Yanning Zhang are with the School of Computer Science, National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, the Shaanxi Provincial Key Laboratory of Speech and Image Information Processing, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: zhanglei211@mail.nwpu.edu.cn; liyunwpu@gmail.com; wangcong12@mail.nwpu.edu.cn; ynzhang@nwpu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.2977655

training data, and the classification accuracy thus decreases greatly.

To address this problem, we propose an intraclass similarity structure representation-based HSI classification method with few labeled pixels (i.e., samples) in this article. First, a mixed labels-based annotation model is established based on the intraclass spectrum similarity of pixels. Given some randomly selected unlabeled pixels, we employ the proposed annotation model to assign each pixel a mixed label from the top-two possible classes, and then augment the original training set with those labeled pixels. On the augmented training set, we then train a deep convolutional neural network-based classification model. With several individual rounds of the annotation and classifier training procedures, we can obtain several independent classification models and predict the final labels as their fusion results with a voting strategy. The experimental results on different datasets demonstrate that the proposed method can obtain higher classification accuracy compared with the competing HSI classification methods given small amount of training samples.

In general, the main contributions of this article are summarized as follows.

1) A novel mixed labels-based annotation model is proposed based on the intraclass spectrum similarity of pixels, with which more samples with soft labels can be obtained from the unlabeled data resulting in the augmentation of training dataset.
2) A classifier fusion model is proposed by fusing multiple convolutional neural networks trained separately on several individual rounds of augmented datasets, with which the diversity and complementary characteristic of labeled samples among different augmented training datasets are exploited better for HSI classification.
3) With the mixed labels-based annotation model and classifier fusion, the proposed HSI classification method can effectively reduce the overfitting problem and obtain better HSI classification results when exploiting together with deep neural networks.

II. Related Work

For shallow feature-based method, a commonly used strategy is to manually design the shallow features for each pixel, and then train the linear or nonlinear classifier with labeled samples. For example, Demir [8] and Melgani [9] used the original spectral curve to represent each pixel directly, and then trained a relevance vector machine (RVM) classifier or SVM classifier based on the spectra. In order to eliminate the redundant information in the spectra and thus obtain more discriminative representations, Yuan *et al.* [10] proposed a spectral band selection model based on the multigraph determinantal point process. The proposed method used the graph model to mine the structure between bands, from which the redundant information was eliminated and the optimal subset of bands was obtained. Wang *et al.* [11] eliminated the redundant information in the spectra via band clustering strategy, and proposed an optimal clustering optimization framework. Different from the band selection methods mentioned above, Bruce *et al.* [12] proposed a binary discrete wavelet transform method to map high-dimensional spectral data into low-dimensional feature space for classification. However, discrete wavelet transform uses fixed basis functions to represent the spectra, which can not adaptively mine the specific structural relationships among the data. To address this problem, Chen *et al.* [2] learned redundant spectral dictionaries from HSI based on the sparse representation framework [2], with which the original continuous spectra are represented by sparse feature vectors. Then based on the sparse representation model, the reconstruction error of the pixel is exploited to construct the classification model. Zhang *et al.* [13] further introduced the spatial nonlocal similarity of pixels into the sparse representation model, thus a classification model with stronger discriminability can be obtained. In recent studies, Li *et al.* [21] proposed a structure-aware collaborative representation method with Tikhonov regularization for HSI classification. Zhang *et al.* [22] proposed a discriminative marginalized least square regression method for HSI classification. These methods address HSI classification from novel and effective perspective, and result in better classification results [21], [22]. Although the above methods promote the development of HSI classification research, they have limitations on classification problems especially in complex classification tasks due to the shallow structure feature as well as the feature needs to be modeled with stronger professional background knowledge.

Instead, deep neural network-based method can learn a deep yet complex representation model automatically from the labeled training data, with which the feature expression ability can be significantly improved [23]–[28]. More importantly, feature extraction and classifier are naturally integrated into the deep neural network framework, which avoids human interventions and thus fully exploits the characteristic of HSI. For example, Chen *et al.* [20] constructed a deep convolutional network for HSI classification by cascading multiple convolutional layers and pooling layers. The constructed deep model significantly improved the classification accuracy of HSI by extracting deep and nonlinear features of the spectrum of each pixel. Zhao *et al.* [23] proposed a HSI classification model based on spatial–spectral features. First, a local discriminant embedding model is proposed to map high-dimensional spectrum into low-dimensional feature, and a deep convolutional network is used to extract the spatial features of the image. Then, these two kinds of features are combined to train the classifier. Although the existing deep neural network-based methods significantly improve the classification accuracy, the prerequisite of obtaining such a performance is that sufficient labeled training samples are provided. However, labeling HSI is tedious, expensive, and only can be accomplished by experts. As a result, only few labeled samples are accompanied with a HSI. In this case, majority of deep neural networks are prone to be overfitting [29], i.e., the classification accuracy drops significantly. In this article, we mainly focus on how to mine the internal structural information of HSI to solve the problem when deep neural network meets few labeled training samples. Specifically, we emphasize on how to utilize few labeled data to boost the performance of existing deep learning-based method. First, a new mixed labels-based annotation model is proposed based on the intraclass spectrum similarity of pixels, with which more samples with soft labels can be obtained from the unlabeled data resulting in the

augmentation of training dataset. Then, a classifier fusion model is proposed by fusing multiple convolutional neural networks trained separately on several individual rounds of augmented datasets, with which the diversity and complementary characteristic of labeled samples among different augmented training datasets are exploited better for HSI classification.

III. Proposed Method

In this section, we will introduce the proposed method in details. The proposed mixed labels-based annotation model is described in Section III-A. The constructed deep convolutional network is presented in Section III-B, and the multiclassifier fusion strategy is described in Section III-C.

A. Mixed Labels-Based Annotation Model

Denote a d dimensional spectrum in the HSI as $\boldsymbol{x}_i \in \mathbb{R}^{d\times 1}$, and its class label as $\boldsymbol{y}_i \in \{1, 2, \ldots, L\}$. HSI classification task aims to classify the spectra vectors $\boldsymbol{\mathcal{S}} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ into L categories. In general, HSI classification task is composed of the following two steps. 1) Train a classifier with the labeled training dataset $\boldsymbol{\mathcal{T}} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{m}$; 2) predict the labels $\{\boldsymbol{y}_i\}_{i=m+1}^{n}$ of other unlabeled samples $\boldsymbol{\mathcal{U}} = \{\boldsymbol{x}_i\}_{i=m+1}^{n}$ based on the learned classifier. For lots of classifiers including deep neural network based one, the prerequisite of obtaining a good classification result is that plenty of labeled training samples are provided, i.e., the number of m is large. However, in practical applications, only few labeled training samples can be obtained, namely, m is small and $m \ll n$. In this case, those classifiers are prone to be overfitting and the generalization ability of those classifiers is degenerated, i.e., the prediction ability of the classifier drops greatly.

To address this problem, we propose a new mixed labels-based annotation model, which uses the existing training data $\boldsymbol{\mathcal{T}}$ to further label some unlabeled pixels in $\boldsymbol{\mathcal{U}}$, with which the number of labeled training pixels can be augmented and the overfitting problem can be reduced. The specific ideas are given as follows. Since the spectra curve in HSI reflect the unique physical characteristics of objects in the scene, the spectra from the same object are likely to be similar. On the contrary, pixels with similar spectra are likely to be chosen from the same object, i.e., have the same class label [30]. Thus, by mining the spectrum similarity among different pixels in $\boldsymbol{\mathcal{T}}$ and $\boldsymbol{\mathcal{U}}$, we can transfer the labels in $\boldsymbol{\mathcal{T}}$ to the spectra similar pixels in $\boldsymbol{\mathcal{U}}$.

Following this idea, we construct spectrum similarity model first. Suppose there are k samples $\{\boldsymbol{x}_j^i\}_{j=1}^{k}$ whose spectra are similar to $\boldsymbol{x}_i$, each $\boldsymbol{x}_j^i$ thus can be approximated by $\boldsymbol{x}_i$ together with a certain (zero mean) gaussian white noise due to the spectrum similarity [31], [32]. Thus, $\boldsymbol{x}_i$ can be reconstructed by a linear combination of all spectral similar samples $\boldsymbol{x}_j^i$ as

$$\boldsymbol{x}_i = \sum_{j=1}^{k} w_j \boldsymbol{x}_j^i, \ \text{s.t.,}\ w_j \geq 0, \ \sum_{j=1}^{k} w_j = 1. \tag{1}$$

w_j denotes the weight using the jth similar sample to reconstruct $\boldsymbol{x}_i$. To make the weights with physical meanings, the weight w_j is required to be nonnegative and the summation of all weights from different similar samples equals to 1 [31], [32], as shown in (1). In order to guarantee the samples closer to $\boldsymbol{x}_i$ have higher influence on reconstructing $\boldsymbol{x}_i$ (i.e., have a larger weight), we use the following weighting equation shown as

$$w_j = \frac{\exp\left(-\frac{1}{\sigma}\left|\left|\boldsymbol{x}_i - \boldsymbol{x}_j^i\right|\right|_2^2\right)}{\sum_j \exp\left(-\frac{1}{\sigma}\left|\left|\boldsymbol{x}_i - \boldsymbol{x}_j^i\right|\right|_2^2\right)} \tag{2}$$

where σ is a predefined scalar.

Based on the spectrum similarity model, we then construct mixed labels-based annotation model to label the samples in $\boldsymbol{\mathcal{U}}$. Since samples with similar spectrum are prone to be from the same label with high probability, we select the nearest k samples $\{\boldsymbol{x}_j^i\}_{j=1}^{k}$ from the training dataset $\boldsymbol{\mathcal{T}}$ for each sample in $\boldsymbol{\mathcal{U}}$. Suppose the label of $\boldsymbol{x}_j^i$ is $\boldsymbol{y}_j^i$, the label of $\boldsymbol{x}_i$ (namely, $\boldsymbol{y}_i$) can be predicted by the following mixed labels based-annotation model

$$\boldsymbol{y}_i = \sum_{j=1}^{k} w_j \boldsymbol{y}_j^i. \tag{3}$$

It is worth noting that all the labels of $\boldsymbol{y}_j^i$ in (3) use the one-hot representation, and thus the generated label of $\boldsymbol{y}_i$, according to (3), is a soft label.

According to above description, it can be seen that the proposed mixed labels-based annotation model is based on the premise that samples with similar spectra are from the same label. However, in reality, the premise might be slightly different. That is, in practice, more similar the spectra, higher probability of above premise is true. To make the proposed model can function well in reality, (2) is further refined with the following two steps.

1) Step 1: For each class, one sample whose spectrum closet to that of $\boldsymbol{x}_i$ is chosen from $\boldsymbol{\mathcal{T}}$ based on the Euclidean distance. By assembling the chosen samples from all L classes into a set, we obtain an alternative set $\boldsymbol{\mathcal{N}} = \{\hat{\boldsymbol{x}}_j^i\}_{j=1}^{L}$ in which $\hat{\boldsymbol{x}}_j^i$ indicates the sample whose spectrum closest to that of $\boldsymbol{x}_i$ in the jth class of training set.

2) Step 2: Top two samples whose spectra closest to that of $\boldsymbol{x}_i$ are chosen from $\boldsymbol{\mathcal{N}}$, which composes $\{\boldsymbol{x}_j^i\}_{j=1}^{k=2}$ and exploits to predict the label of $\boldsymbol{x}_i$ according to (3).

With above proposed annotation model, we then randomly select l unlabeled samples from $\boldsymbol{\mathcal{U}} = \{\boldsymbol{x}_i\}_{i=m+1}^{n}$ and predict their labels. Then, we add these data into the original training dataset $\boldsymbol{\mathcal{T}}$, from which we obtain the augmented training dataset $\boldsymbol{\mathcal{R}} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{m+l}$.

B. Deep Convolutional Network-Based Classification Model

Given the augmented training dataset $\boldsymbol{\mathcal{R}}$, we then design a specific classifier model. Since deep convolutional network can learn powerful features, it is adopted to construct the classification model in this article [17], [18].

Specifically, three convolutional layers together with a fully connected layer are used to map the input spectrum into complex nonlinear feature space. Each convolutional layer consists of three consecutive submodules including a convolution submodule, a nonlinear activation function submodule, and a batch normalization submodule [33]. In this article, rectified linear unit [34] is adopted as a nonlinear activation function module.

Batch normalization submodule normalizes the input data of any value into a real-number data ranging from 0 to 1. In the constructed structure, the convolutional layer focuses on extracting the local information within the spectrum, while the fully connected layer introduces the global information of the spectrum curve into the features. Given the generated features, we further introduce a classifier layer, which outputs the probability of each class. The classifier layer contains two cascading submodules including the fully connected submodule and an activation function submodule. The fully connected submodule is equivalent to a linear classifier, and the softmax activation function submodule is used to ensure that the output of the classifier meets the probability requirement, i.e., each element in the output is nonnegative and the summation of all elements in the output equals to one.

Assuming that the proposed deep convolutional network-based classification model can be expressed as f_θ, the classification model can be trained by solving the following optimization problem

$$\min_{\theta} \sum_{i=1}^{m+l} \text{Loss}\left(f_\theta\left(\boldsymbol{x}_i\right), \boldsymbol{y}_i\right). \tag{4}$$

θ represents the parameters to be estimated in the network and Loss represents the cross entropy loss function [35]. Similar as the commonly used deep learning method, equation (4) can also be solved by back propagation algorithm [36].

C. Classifier Fusion Model

With the proposed mixed labels-based annotation model, the training dataset can be augmented by predicting the labels of l unlabeled samples chosen randomly from $\boldsymbol{\mathcal{U}}$. It is easy to see that the samples randomly chosen are always complemented with different sampling (i.e., different samples exist among different sampling). Considering such complementary property, we propose to fuse multiple convolutional neural networks trained separately on several individual rounds of augmented datasets to further boost the classification results. That is, we first perform the sampling r times independently and train r independent classifiers, respectively, and then obtain the labels for the test data by fusing the results from all r classifiers.

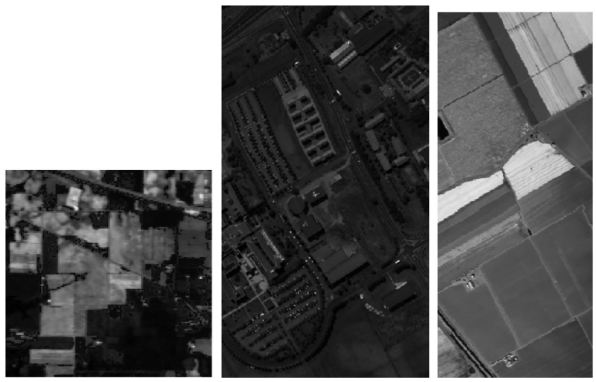
Specifically, suppose the trained r classifiers are denoted as $f_\theta^{(1)}, f_\theta^{(2)}, \ldots, f_\theta^{(r)}$, we obtain the class label $\boldsymbol{y}_i$ of the test sample $\boldsymbol{x}_i$ by the following three steps, in which we utilize c as an index to represent a possible class label among $\{1, 2, \ldots, L\}$.

1) Step 1: Initialize the scores of L classes as $p_1 = p_2 = \cdots = p_L = 0$, where p_c denotes the score from cth class.

2) Step 2: Update the scores based on the predicted results from each classifier. For example, if the pixel $\boldsymbol{x}_i$ is predicted from the cth class via a classifier, only the score belonging to the of cth class (i.e., p_c) will be updated by $p_c = p_c + 1$. The scores stop updating until all r classifiers are used for prediction the label of $\boldsymbol{x}_i$.

3) Step 3: Predict the label of the test sample $\boldsymbol{x}_i$ based on the obtained scores as

$$\boldsymbol{y}_i = arg \max_c p_c. \tag{5}$$



(a) (b) (c)

Fig. 1. Representative band images of the adopted three HSI datasets. (a) Indian Pines. (b) Pavia University. (c) Salinas.

IV. Experiments Results and Analysis

In this section, we conduct different experiments on three HSI classification datasets to demonstrate the effectiveness of the proposed method.

A. Datasets

Three benchmark HSI datasets including Indian Pines dataset, Pavia University dataset, and Salinas dataset are adopted for experiments.

Indian Pines dataset contains 16 categories, which are collected by airborne visible/infrared imaging spectrometer (AVIRIS) from northwestern Indiana. It captures 220 different spectral bands ranging from 0.38 μm to 2.5 μm, and has 145×145 pixels. The Indian Pines dataset has 16 classes, but the number of pixels varied dramatically between classes. To ensure the relative balance of pixel numbers from different classes, we adopt the strategy lots of HSI classification adopt on this dataset [37], [38], i.e., we choose nine classes which have more data than the remaining 7 classes for experiments.

Pavia University dataset contains nine land covers and 610×340 pixels, which is acquired by reflective optics system imaging spectrometer (ROSIS) from Pavia University, Italy. It captures 115 spectral bands ranging from 0.43 μm to 0.86 μm. In the experiment, we remove 12 water-absorption bands from the data and preserve the remaining 103 bands for experiments.

Salinas dataset is acquired by AVIRIS from California, which includes 16 classes, 224 bands, and 512×217 pixels. Similar as that does in Pavia University dataset, 20 water-absorption bands are removed and the remaining 204 bands are preserved for experiments.

The illustrations of the adopted datasets can be seen from Fig. 1. In addition, the name and the numbers of pixels per adopted class on above three datasets are summarized in Table I.

B. Comparison Methods

In order to verify the proposed method, we compare the proposed method with four different comparison methods including spectra-based SVM [9], spatio–spectra Laplacian support vector machine (SSL-SVM) [39], deep convolutional neural

TABLE I
NAME AND PIXEL NUMBERS PER ADOPTED CLASS ON THREE DATASETS

	Indian Pines		Pavia University		Salinas	
No.	Class Name	Number	Class Name	Number	Class Name	Number
1	Corn-notill	1428	Asphalt	6631	Brocoli_green_weeds_1	2009
2	Corn-mintill	830	Meadows	18649	Brocoli_green_weeds_2	3726
3	Grass-pasture	483	Gravel	2099	Fallow	1976
4	Hay-windrowed	478	Trees	3064	Fallow_rough_piow	1394
5	Soybean-notill	972	Metal Sheets	1345	Fallow_smooth	2678
6	Soy-mintill	2455	Bare Soil	5029	Stubble	3959
7	Soybean-clean	593	Bitumen	1330	Celery	3579
8	Woods	1265	Bricks	3682	Grapes_untrained	11271
9	Grass-tress	730	Shadows	947	Soil_vinyard_develop	6203
10					Corn_senesced_green_weeds	3278
11					Lecture_romaine_4wk	1068
12					Lecture_romaine_5wk	1927
13					Lecture_romaine_6wk	916
14					Lecture_romaine_7wk	1070
15					Vinyard_untrained	7268
16					Vinyard_vertical_trellis	1807
sum		9234		42776		54129

networks (DCNN) [20], and mixture annotation-based deep convolutional neural networks (MA-DCNN). Among them, spectra-based SVM method and SSL-SVM are few-sample oriented classification methods, which have been testified to function well when given only few training samples. For spectra-based SVM method, it uses the spectrum as the feature of each pixel and trains the SVM classifier. SSL-SVM method is a semisupervised classification method, which has similarities with the proposed mixed labels-based annotation model since it also incorporates the unlabeled samples in the training stage. It first mines the spatial similarity among pixels, then construct a manifold structure regularizer to constrain the training process of SVM classifier, with which robust HSI classification results can be obtained. DCNN is a deep convolutional neural network-based HSI classification method, which is adopted to testify the effectiveness of the deep learning-based method when confronting with few samples. For a fair comparison, we adopt the same structure as the classification model described in Section IV-B for DCNN. MA-DCNN is a variant of the proposed method. The only difference between the proposed method and MA-DCNN is that MA-DCNN uses one augmented training dataset to train the classifier without classifier fusion. For simplification, we term the proposed method as Ours in this article.

C. Evaluation Criterion

In order to compare the performance of different methods objectively, we use three different measures including overall accuracy (OA) [9], average accuracy (AA) [9], and Kappa coefficient [40] in the experiments. For these measures, their values range from 0 to 1, and higher value represent higher classification performance.

D. Setting

In the following experiments, to simulate the case only few labeled pixels are provided for training, we randomly select 20 labeled pixels per class for training and the remaining pixels for test. With the training data, we first train all methods, and then calculate three evaluation measures (i.e., OA, AA, and Kappa) on test data for each method. In order to reduce the randomness of pixels sampling on classification, we randomly sample the training pixels ten times independently, with which ten independently classifiers are trained for each method. We then calculate the average measures on test data obtained from those ten classifiers, which is recorded as the final classification results.

TABLE II
CLASSIFICATION RESULTS OF DIFFERENT METHODS ON INDIAN PINES DATASET (%)

Class	SVM	SSL-SVM	DCNN	Ours
1	53.38	47.07	61.30	**70.24**
2	60.90	62.54	57.43	**65.41**
3	89.38	**91.55**	85.18	90.08
4	96.15	**97.67**	94.23	96.48
5	99.56	**100.00**	97.66	99.60
6	63.37	70.49	72.80	**76.83**
7	49.05	**58.15**	49.88	54.93
8	59.70	69.60	70.30	**71.21**
9	91.62	**96.55**	92.62	92.89
OA	67.26 (1.74)	71.17 (2.71)	69.73 (2.26)	**74.23** (1.20)
AA	73.68 (1.03)	77.07 (1.44)	75.71 (1.85)	**79.74** (0.84)
Kappa	62.31 (1.85)	66.69 (2.87)	65.11 (2.50)	**70.24** (1.34)

The best results are in bold.
The Numbers in Parentheses are Standard Variances (%).

The proposed method was implemented by Tensorflow [41]. During the training stage, Adam [42] optimizer was used to train the network, among which the initial learning rate, batch size, and the total epoch were set as 0.001, 100, and 50 000, respectively. DCNN, MA-DCNN, and Ours are adopted the same one-dimentional convolutional neural network architecture, which includes four convolutional layers. For the proposed mixed labels-based annotation model, parameters σ and l were set as 0.01 and 20, respectively. We set the sampling times r and the number of nearest sample k as 5 and 2 in this article, respectively.

E. Experiments Results

With the same experimental settings, the classification experiments are conducted for different methods on three hyperspectral datasets, and the results are summarized in Tables II to IV. In these tables, AA, OA, Kappa coefficients, as well as the classification accuracy of each class are provided.

TABLE III
CLASSIFICATION RESULTS OF DIFFERENT METHODS ON PAVIA UNIVERSITY DATASET (%)

Class	SVM	SSL-SVM	DCNN	Ours
1	75.00	74.75	73.25	**76.95**
2	74.22	71.82	71.80	**80.97**
3	76.69	**84.70**	77.62	81.34
4	91.78	**96.92**	87.98	86.43
5	99.24	99.39	99.12	**99.52**
6	76.23	**83.48**	69.79	70.90
7	90.56	**96.77**	89.77	90.83
8	75.56	**83.35**	70.24	71.12
9	99.82	**99.89**	99.59	99.77
OA	77.93 (2.61)	79.33 (3.01)	75.13 (3.65)	**80.03** (1.92)
AA	84.34 (1.07)	**87.90** (1.39)	82.13 (1.41)	84.20 (0.67)
Kappa	71.98 (2.99)	74.03 (3.46)	68.47 (4.14)	**74.17** (2.30)

The Numbers in Parentheses are Standard Variances (%).

TABLE IV
CLASSIFICATION RESULTS OF DIFFERENT METHODS ON SALINAS DATASET (%)

Class	SVM	SSL-SVM	DCNN	Ours
1	97.77	98.82	97.99	**98.20**
2	97.07	96.74	**99.60**	99.57
3	94.99	93.31	95.53	**98.34**
4	99.38	99.23	99.21	**99.40**
5	94.10	95.26	96.54	**96.73**
6	99.25	99.00	99.28	**99.40**
7	99.27	99.44	99.50	**99.56**
8	67.55	69.20	55.70	**72.01**
9	96.64	97.32	97.21	**97.56**
10	86.44	86.06	90.19	**90.21**
11	94.29	94.94	96.31	**96.53**
12	99.71	**99.98**	99.51	99.76
13	97.50	97.91	97.96	**98.16**
14	93.54	92.49	93.10	**93.93**
15	65.95	68.44	**77.99**	67.99
16	95.49	**95.98**	95.20	95.90
OA	86.14 (1.61)	86.90 (1.53)	85.94 (2.70)	**88.23** (1.12)
AA	92.43 (0.67)	92.76 (0.72)	93.18 (0.60)	**93.95** (0.40)
Kappa	84.62 (1.75)	85.46 (1.67)	84.46 (2.92)	**86.93** (1.23)

The Numbers in Parentheses are Standard Variances (%).

Specifically, Table II shows the results of different methods on Indian Pines dataset. It can be seen that the advantage of DCNN-based method over SVM is not obvious for classification when few labeled samples (i.e., 20 per class) are provided. For example, compared with SVM, OA of DCNN is only improved by 2.4%. The reason is that DCNN is prone to be overfitting when few training samples are provided, which will influence the performance of the deep structured network. On the contrary, the proposed method has better performance, though it adopts the same network structure of DCNN. For example, the proposed method improves OA about 7% over SVM. We attribute the improvement to the proposed mixed labels-based annotation model and the classifier fusion strategy, which is beneficial to reduce the overfitting problem.

SSL-SVM is a semisupervised classification method. It exploits the unlabeled samples in its training stage, which has similarities with the proposed method and thus has better performance over SVM. Even though, we can find that the proposed method has better performance over SSL-SVM in the measures of AA, OA, and Kappa coefficients. For example, the OA measure is improved by 3% for the proposed method over SSL-SVM.

Table III gives the classification results of different methods on Paiva University dataset. We can find that the classification accuracy of DCNN is even lower than that of the SVM. For example, the OA of DCNN is lower than that of SVM by 2.8%. This phenomenon indicates that DCNN has a more serious overfitting problem for Pavia University dataset. With the same network structure, the proposed method improves the classification results clearly. For example, the proposed method improves kappa coefficient by 5.7%, compared with DCNN, which demonstrates the proposed method can effectively reduce the overfitting problem of deep neural network meets when given only few samples.

Table IV shows the classification results of different methods on Salinas dataset. It can be found that the proposed method achieves the highest classification accuracy among all the methods, similar as the results obtained from Indian Pines dataset and Pavia University dataset. For example, the Kappa coefficient of the proposed method is improved by 2.4%, compared with DCNN, which demonstrates the overfitting problem of the deep neural network can be reduced with the proposed method. In addition, we report the standard variance of OA, AA, and Kappa coefficient on these three datasets in Tables II–IV, from which we can see that the proposed method is more stable than other competing methods. To further illustrate the classification results, the classification maps are given in Figs. 2–4. It can be seen that, compared with other methods, the classification map of the proposed method is closer to the ground truth in most cases, which further validates the effectiveness of the proposed method.

F. Ablation Study

Since both the proposed mixed labels-based annotation model and the fusion strategy contribute to the improvement of HSI classification over DCNN, we verify how those two parts influence the classification results independently in this subsection. In addition, the classification performance varied with parameters l as well as the numbers of training samples are also testified in this subsection.

1) Validity of the Model: In order to verify the effectiveness of the two different parts (i.e., mixed labels-based annotation model and the fusion strategy) on reducing the overfitting problem DCNN meets, the following experiments are designed and implemented on three HSI datasets. Specifically, we compare the proposed method with DCNN and MA-DCNN on three datasets. Those three methods share the same network structure. It is noticeable that DCNN and MA-DCNN can be regarded as special cases of the proposed method in this article. Specifically, DCNN does not use the mixed labels-based annotation model and the classifier fusion strategy. MA-DCNN only adopts mixed labels-based annotation model but without using the classifier fusion strategy. Table V gives the classification results of those methods. Compared with DCNN, MA-DCNN can steadily improve the performance, while the proposed method can improve the performance even more. Therefore, the improvement of MA-DCNN over DCNN in terms of HSI classification indicates

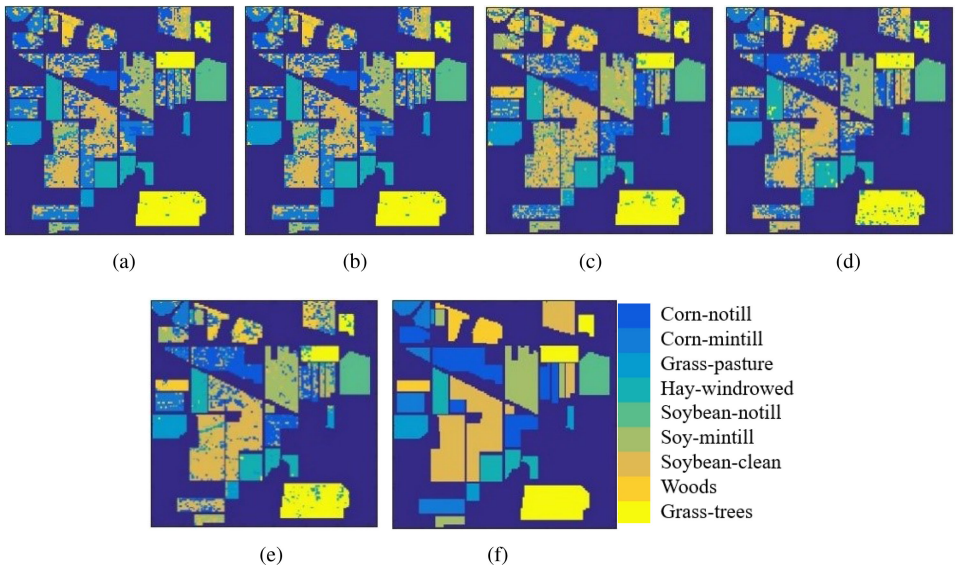


Fig. 2. Classification maps of different methods on Indian Pines dataset. (a) SVM. (b) SSL-SVM. (c) DCNN. (d) MA-DCNN. (e) Ours. (f) Ground Truth.

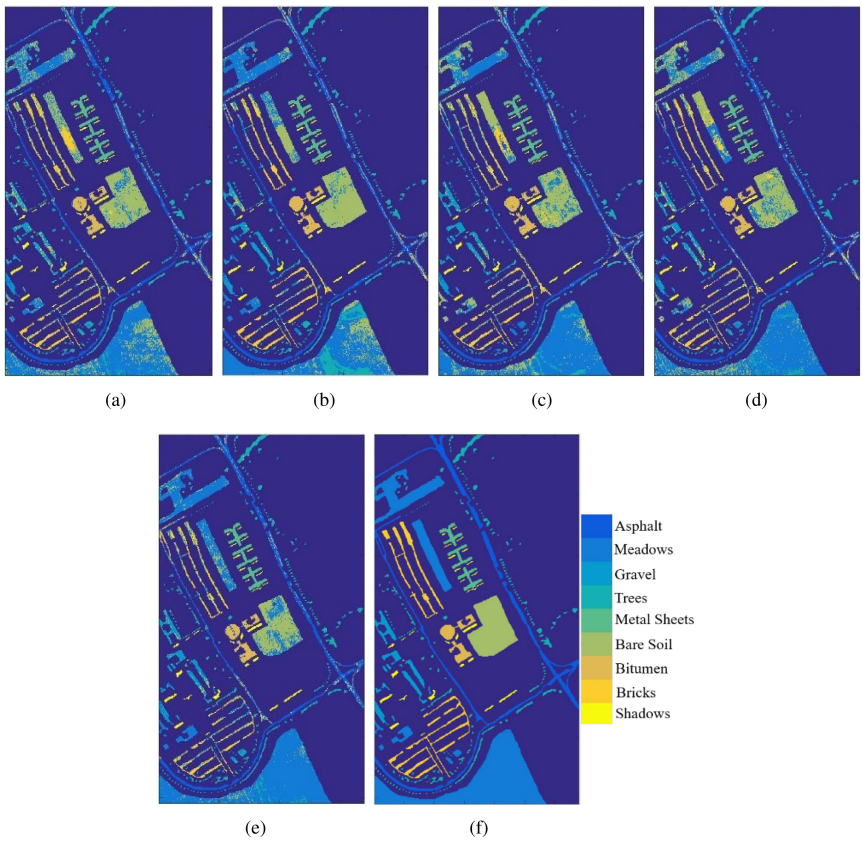


Fig. 3. Classification maps of different methods on Pavia University dataset. (a) SVM. (b) SSL-SVM. (c) DCNN. (d) MA-DCNN. (e) Ours. (f) Ground Truth.

that the proposed mixed labels-based annotation model can stably reduce the overfitting phenomenon. By comparing the results of the proposed method and MA-DCNN, it can be seen that the classifier fusion strategy can also boost the HSI classification performance. From above experiments, we can see that both the mixed labels-based annotation model and the fusion strategy can helpful to address overfitting problem.

2) Effect of Parameter l: l is the parameter related with the number of samples to be labeled via mixed labels-based annotation model. We testify the classification results varied with different l (i.e., l is chosen as 0, 10, 20, 30, and 50, respectively) on Indian Pines dataset, and the results are given in Table VI. It is worth noting that when $l = 0$, our proposed method directly degenerates into the original DCNN. From the

TABLE V
EFFECTIVENESS OF MIXTURE ANNOTATION MODEL AND CLASSIFIER FUSION MODEL ON DIFFERENT DATASETS (%)

Dataset	Indian Pines			Pavia University			Salinas		
Method	DCNN	MA-DCNN	Ours	DCNN	MA-DCNN	Ours	DCNN	MA-DCNN	Ours
OA	69.73	71.39	**74.23**	75.13	76.87	**80.03**	85.94	87.36	88.23
AA	75.71	76.44	**79.74**	82.13	83.16	**84.20**	93.18	93.39	**93.95**
Kappa	65.11	66.91	**70.24**	68.47	70.54	**74.17**	84.46	85.97	86.93

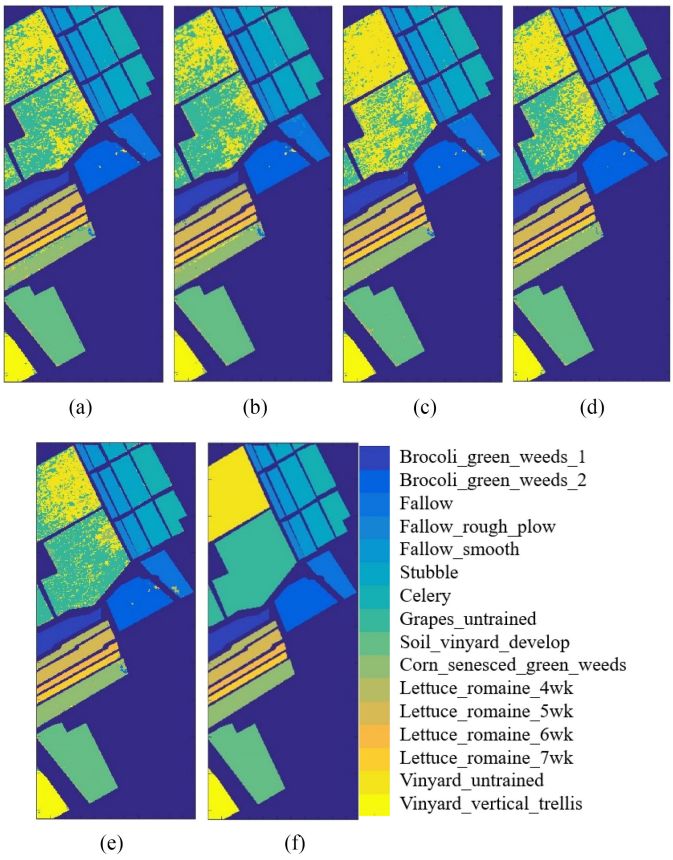


Fig. 4. Classification maps of different methods on Salinas dataset. (a) SVM. (b) SSL-SVM. (c) DCNN. (d) MA-DCNN. (e) Ours. (f) Ground Truth.

TABLE VI
CLASSIFICATION RESULTS WITH DIFFERENT l ON INDIAN PINES DATASET (%)

Number	0	10	20	30	50
OA	69.73	72.62	**74.23**	72.58	71.95
AA	75.71	78.22	**79.74**	78.09	77.75
Kappa	65.11	68.39	**70.24**	68.33	67.66

table we can see that, the classification performance of the proposed method first increases then decreases with the increase of l. When $l = 20$, the proposed method has best classification performance. The reason is analysed as follows. When l is small, the number of augmented samples is limited, which influences the classification performance. When l exceeds a certain value, the number of mistaken labeled samples increases, which leads to the degradation of classification performance.

3) Effect of Sampling Times r: r is the sampling times related with the number of classifiers used for fusion. We testify the classification results when r is set as 3, 5, 7, respectively, and summarize the experimental results in Table VII, from which we can see that the classification performance of the proposed method improves when r increases. Due to the randomly sampling, the obtained datasets from different sampling complement, and thus the classification results can be improved when using more classifiers trained on these datasets. Nevertheless, the training cost increases with r and thus we set the sampling times r as 5 in all experiments in this article.

TABLE VII
CLASSIFICATION RESULTS ON INDIAN PINES WITH DIFFERENT SAMPLING TIMES r

Sampling times	3	5	7
OA	73.17	74.23	**75.66**
AA	78.55	79.74	**81.01**
Kappa	69.01	70.24	**71.86**

TABLE VIII
CLASSIFICATION RESULTS ON INDIAN PINES DATASET WITH DIFFERENT NUMBER OF NEAREST SAMPLES k

k	1	2	3
OA	71.68	**74.23**	72.25
AA	77.52	**79.74**	78.09
Kappa	67.33	**70.24**	67.98

4) Effect of the Number of Nearest Neighbors k: k is the number of nearest neighbors. In above experiments, we fix $k = 2$. In this experiment, we compare the classification results when k is set as 1, 2, and 3, respectively, and summarize the experimental results in Table VIII. When k varies from 2 to 3, higher probability pixels from different classes are prone to be introduced for training, which thus degrades the performance of the proposed HSI classification method. It is noticeable that when $k = 1$, the proposed mixed labels-based annotation model degenerates to nearest neighbor method, i.e., we assign test sample a label of the most similar samples in the training dataset. However, utilizing only one pixel (i.e., $k = 1$) is easily influenced by the noise, which results in the degeneration of the classification performance, compared with the method we utilize $k = 2$ and $k = 3$ for labeling. In addition, the predicted label is a hard label when $k = 1$. In contrast, the predicted label is a soft label when $k = 2$ or $k = 3$, which indicates the mixed label annotation method is beneficial for HSI classification within the proposed method.

5) Effect of the Number of Training Samples: In the above-mentioned experiments, the number of labeled training samples in each class is fixed as 20. In this section, we test the performance of the proposed method and its most direct comparison,

TABLE IX
EFFECTIVENESS OF THE PROPOSED METHOD AND DCNN WITH DIFFERENT NUMBER OF TRAINING SAMPLES ON INDIAN PINES DATASET (%)

Number	Method	OA	AA	Kappa
1	DCNN	38.39	42.2	29.71
	Ours	**39.76**	**44.61**	**31.41**
5	DCNN	52.16	57.4	45.09
	Ours	**55.22**	**61.21**	**48.62**
10	DCNN	59.18	65.43	53.07
	Ours	**62.65**	**69.36**	**57.05**
20	DCNN	69.73	75.71	65.11
	Ours	**74.23**	**79.74**	**70.24**
50	DCNN	76.63	82.47	73.04
	Ours	**79.46**	**85.18**	**76.29**

i.e., DCNN, with different number of training samples. Specifically, we set the number of training samples for each class as 1, 5, 10, 20, and 50. Table IX shows the classification results of the proposed method and DCNN. It can be seen from the table that, with different number of training samples, the proposed method can reduce the overfitting problem and thus improve the classification accuracy.

V. CONCLUSION

To address the overfitting problem deep neural network confronts when providing few training samples, an intraclass similarity structure representation-based HSI classification method has been proposed in this article. First, according to the intraclass spectrum similarity of pixels, we establish a mixed labels-based annotation model. Given some randomly selected unlabeled pixels, we employ the proposed annotation model to assign them mixed labels from the top-two possible classes, and then augment the original training set with those labeled pixels. On the augmented training set, we train a deep convolutional neural network-based classification model. With several individual rounds of the annotation and classifier training procedures, we can obtain several independent classification models and predict the final labels as their fusion results with a voting strategy. Experimental results show that the proposed method can effectively deal with the problem few training samples encounters and thus obtain higher classification accuracy.

REFERENCES

[1] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, 2009.

[2] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[3] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[4] F. A. Kruse, J. W. Boardman, and J. F. Huntington, "Comparison of airborne hyperspectral data and EO-1 hyperion for mineral mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1388–1400, Jun. 2003.

[5] M. Gålfalk, G. Olofsson, and D. Bastviken, "Approaches for hyperspectral remote flux quantification and visualization of GHGs in the environment," *Remote Sens. Environ.*, vol. 191, pp. 81–94, 2017.

[6] Y. Zhang, B. Du, L. Zhang, and T. Liu, "Joint sparse representation and multitask learning for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 894–906, Feb. 2017.

[7] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, Jun. 2016.

[8] B. Demir and S. Erturk, "Hyperspectral image classification using relevance vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 586–590, Oct. 2007.

[9] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[10] Y. Yuan, X. Zheng, and X. Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 51–64, Jan. 2017.

[11] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018.

[12] L. M. Bruce, C. H. Koger, and J. Li, "Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2331–2338, Oct. 2002.

[13] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.

[14] Y. Gu, Q. Wang, H. Wang, D. You, and Y. Zhang, "Multiple kernel learning via low-rank nonnegative matrix factorization for classification of hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2739–2751, Jun. 2015.

[15] W. Li, Q. Du, F. Zhang, and W. Hu, "Collaborative-representation-based nearest neighbor classifier for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 389–393, Feb. 2015.

[16] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.

[17] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, 2015.

[18] 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.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[20] 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.

[21] W. Li, Y. Zhang, N. Liu, Q. Du, and R. Tao, "Structure-aware collaborative representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7246–7261, Sep. 2019.

[22] Y. Zhang, W. Li, H.-C. Li, R. Tao, and Q. Du, "Discriminative marginalized least-squares regression for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.,* to be published.

[23] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[24] L. Liu *et al.*, "Compositional model based fisher vector coding for image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2335–2348, Dec. 2017.

[25] X.-S. Wei, P. Wang, L. Liu, C. Shen, and J. Wu, "Piecewise classifier mappings: Learning fine-grained learners for novel categories with few examples," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 6116–6125, Dec. 2019.

[26] P. Zhou, J. Han, G. Cheng, and B. Zhang, "Learning compact and discriminative stacked autoencoder for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4823–4833, Jul. 2019.

[27] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

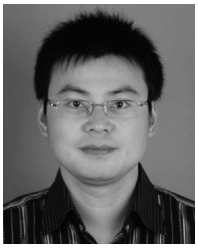
[28] J. Wang, J. Zhou, and W. Huang, "Attend in bands: Hyperspectral band weighting and selection for image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4712–4727, Dec. 2019.

[29] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 8609–8613.

[30] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York, NY, USA: Springer, 2003, vol. 1.

[31] L. Zhang, W. Wei, Y. Zhang, C. Shen, A. van den Hengel, and Q. Shi, "Cluster sparsity field: An internal hyperspectral imagery prior for reconstruction," *Int. J. Comput. Vision*, vol. 126, no. 8, pp. 797–821, 2018.

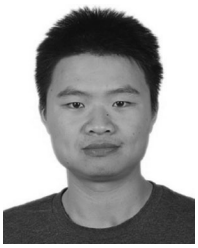
[32] L. Zhang, W. Wei, C. Bai, Y. Gao, and Y. Zhang, "Exploiting clustering manifold structure for hyperspectral imagery super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5969–5982, Dec. 2018.
[33] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.
[34] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
[35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, 2015, Art. no. 436.
[36] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural Networks for Perception*. New York, NY, USA: Elsevier, 1992, pp. 65–93.
[37] 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.
[38] W. Wei, J. Zhang, L. Zhang, C. Tian, and Y. Zhang, "Deep cube-pair network for hyperspectral imagery classification," *Remote Sens.*, vol. 10, no. 5, 2018, Art. no. 783.
[39] L. Yang, S. Yang, P. Jin, and R. Zhang, "Semi-supervised hyperspectral image classification using spatio-spectral Laplacian support vector machine," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 3, pp. 651–655, Mar. 2014.
[40] R. G. Pontius Jr and M. Millones, "Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment," *Int. J. Remote Sens.*, vol. 32, no. 15, pp. 4407–4429, 2011.
[41] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proc. 12th Symp. Operating Syst. Des. Implementation*, 2016, pp. 265–283.
[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.



Wei Wei (Senior Member, IEEE) received the Ph.D. degree from the Northwestern Polytechnical University, Xi'an, China, in 2012.

He is currently an Associate Professor with the School of Computer Science and Engineering, Northwestern Polytechnical University. He has been authored more than 40 papers including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, *Pattern Recognition*, Conference on Computer Vision and Pattern Recognition (CVPR), International Conference on Computer Vision (ICCV), European Conference on Computer Vision (ECCV), Association for the Advancement of Artificial Intelligence (AAAI), International Joint Conference on Artificial Intelligence (IJCAI), etc. His research interests include image processing, machine learning, and pattern recognition.

Dr. Wei is a reviewer for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE REMOTE SENSING LETTERS, and IEEE TRANSACTIONS ON NEURAL NETWORK AND LEARNING SYSTEM, etc. He has served as the PC member for around ten major international conference including CVPR, International Conference on Multimedia & Expo (ICME), etc.



Lei Zhang (Student Member, IEEE) received the Ph.D degree in computer science and technology from the Northwestern Polytechnical University, Xi'an, China, in 2018.

His research interests include hyperspectral image processing and machine learning.

Yu Li received the bachelor's degree from the Northwestern Polytechnical University, Xi'an, China, where she is currently working toward the M.S. degree.

Her research interests include hyperspectral image classification.

Cong Wang received the M.Sc. degree in 2016 from the Northwestern Polytechnical University, Xi'an, China, where she is currently working toward the Ph.D. degree.

Yanning Zhang (Senior Member, IEEE) received the B.S. degree from the Dalian University of Technology, Dalian, China, in 1988, the M.S. degree from the School of Electronic Engineering, Northwestern Polytechnical University, Xi'an, China, in 1993, and the Ph.D. degree from the School of Marine Engineering, Northwestern Polytechnical University, Xian, China, in 1996.

She is currently a Professor with the School of Computer Science, Northwestern Polytechnical University. She is also a Cheung Kong Professor of Ministry of Education, China. She has authored more than 200 papers. Her current research interests include remote sensing image analysis, computer vision and pattern recognition, etc.

Dr. Zhang is an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.