

Adaptive Spatial Pyramid Constraint for Hyperspectral Image Classification With Limited Training Samples

Jun Yue^{ID}, Dingshun Zhu^{ID}, Leyuan Fang^{ID}, Senior Member, IEEE, Pedram Ghamisi^{ID}, Senior Member, IEEE, and Yaowei Wang^{ID}, Member, IEEE

Abstract—Deep learning-based methods have made significant progress in hyperspectral image (HSI) classification in recent years. However, deep learning-based methods usually rely on a large number of samples, and in many cases, it is difficult to label HSI and only limited training samples are available. To solve this problem, an HSI classification method based on adaptive spatial pyramid constraint (ASPC) is proposed to make full use of the global spatial neighborhood information of the labeled samples, which can improve the generalization ability of the classification model. The main steps of the proposed method are as follows. First, an HSI complexity evaluation method based on edge detection is proposed to assess the homogeneity of the objects in the HSI. Second, an HSI pyramid segmentation method based on spatial pyramid is proposed to generate multiscale subregions, where HSI complexity is used to adaptively determine the scale of the segmentation. Third, a spatial supervised constraint is proposed to generate the loss function of labeled subregions. Fourth, a spatial unsupervised constraint is proposed to generate the loss function of unlabeled subregions. The proposed method fully explores the spatial-spectral correlation between unlabeled samples and labeled samples, and add corresponding constraints to the training objective according to the correlation. By adding the ASPC, the trained model becomes more robust and can make full use of the limited training samples. To verify the effectiveness of the proposed method, three benchmark hyperspectral datasets are used to verify the performance of the proposed method. Experimental results show that the performance of this method is better than the existing state-of-the-art methods.

Manuscript received May 30, 2021; accepted June 27, 2021. Date of publication July 20, 2021; date of current version January 17, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFC1510905, in part by the National Natural Science Fund of China under Grant 61922029, in part by the Science and Technology Plan Project Fund of Hunan Province under Grant 2019RS2016, and in part by the Scientific Research Foundation of the Hunan Education Department under Grant 20B022. (Jun Yue, Dingshun Zhu, and Leyuan Fang contributed equally to this work.) (Corresponding author: Yaowei Wang.)

Jun Yue is with the Department of Geomatics Engineering, Changsha University of Science and Technology, Changsha 410114, China (e-mail: jyue@pku.edu.cn).

Dingshun Zhu is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: dszhu@hnu.edu.cn).

Leyuan Fang is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China, and also with the Peng Cheng Laboratory, Shenzhen 518000, China (e-mail: fangleyuan@gmail.com).

Pedram Ghamisi is with the Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Helmholtz Institute Freiberg for Resource Technology, 09599 Freiberg, Germany, and also with the Institute of Advanced Research in Artificial Intelligence (IARAI), 1030 Vienna, Austria (e-mail: p.ghamisi@gmail.com).

Yaowei Wang is with the Peng Cheng Laboratory, Shenzhen 518000, China (e-mail: wangyw@pcl.ac.cn).

Digital Object Identifier 10.1109/TGRS.2021.3095056

Index Terms—Deep neural network, hyperspectral image (HSI) classification, image complexity, over segmentation, spatial pyramid constraint.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) contain both spatial and spectral information, and each pixel vector in HSI shapes a spectral curve. The purpose of HSI image classification is to determine the corresponding category (such as forest, farmland, river, lake, grassland, building, mineral, rock, and road) for each pixel. As a key step in HSI applications, classification plays an important role in geology, ecology, environment, mining, agriculture, forestry, and other fields [1]–[3].

Due to the high spectral dimension of HSI, many statistical transformation methods have been proposed to transform HSI spectral vector from high-dimensional feature space to low-dimensional feature space for spectral feature extraction. A few examples of such methods are principal component analysis (PCA) [4], linear discriminant analysis (LDA) [5], independent component analysis (ICA) [6], and minimum noise fraction (MNF) [7]. MNF is used to separate noise from signals, which is realized by two cascading PCA. The principal components generated by MNF are sorted according to the signal-to-noise ratio (SNR). However, the aforementioned spectral feature extraction and statistical transformation methods are linear transformation methods. Because of the nonlinear characteristics of HSI, some spectral feature extraction methods based on nonlinear transformation have been proposed in recent years [8], [9], such as locality-preserving discriminant analysis (LFDA) [10], sparsity preserving projections (SPP) [11], manifold coordinate representations [12], and locality preserving projections (LPP) [13].

HSI spectral feature extraction methods extract abstract features from spectral vectors to reduce the dimension of an HSI. However, due to the spatial homogeneity and heterogeneity of remote-sensing images, extracting joint spatial-spectral features is key to make full use of HSI features. In recent years, many joint spatial-spectral feature extraction methods have been proposed, including extended mathematical profile (EMP) [14], extended attribute profile (EAP) [15], and directional morphological profile (DMP) [16]. To prevent the excessive weight of spatial features, researchers have proposed

improved methods, such as morphological feature selection based on genetic algorithm and particle swarm optimization algorithm [17], [18]. In addition to the joint spatial-spectral feature extraction methods based on morphology, researchers also proposed spatial-spectral feature extraction methods based on tensor discriminative locality alignment [19], improved spatial-spectral similarity measure [20], and spatial-spectral regularized local discriminant embedding [21].

Deep learning is a breakthrough in the field of machine learning. With the rapid development of deep learning [22]–[24], this branch of machine learning approaches has achieved great success in image segmentation [25]–[28], image classification [29], [30], object detection [31]–[33], image super-resolution [34], [35], image restoration [36], [37], image denoising [38], and so on. Researchers have also proposed many HSI classification methods based on deep learning, which enhances the accuracy of HSI classification methods. To make full use of the spatial-spectral features of HSIs, a series of joint spatial-spectral HSI feature extraction methods based on deep neural network were proposed, including methods based on stacked autoencoder [39], deep fully convolutional network [40], deep deconvolution network [41], spatial pyramid pooling [42], deep belief network [43], [44], deep residual network [45], [46], deep recurrent neural network [47], and so on.

However, the existing HSI feature extraction and classification methods based on deep learning still need further exploration. One of the important reasons is that training deep models depends on a large number of labeled samples. For remote-sensing images, there are two ways to obtain labeled samples: 1) field investigation and 2) visual interpretation [48]. Compared with visual interpretation, the samples obtained by field surveys often lead to higher accuracy. However, obtaining labeled samples through field investigation leads to higher operational costs, and in some cases, field investigation cannot be carried out. Due to the limitation of access to a large number of samples, it is difficult for researchers to further improve the accuracies of HSI classification methods based on deep neural network [49]–[51].

To solve the problem of insufficient labeled samples, researchers began to explore HSI classification methods based on few-shot learning. Inspired by the fact that the human visual recognition system can still obtain very good recognition ability under limited given samples, few-shot learning has attracted the attention of researchers in the field of computer vision [52]–[54]. At present, few-shot learning has been widely used in image segmentation [55], image classification [56], [57], and object detection [58]. Metric learning or prototype learning is usually used to learn the features of HSI when the number of samples is limited [50], [59], [60]. The nearest sample or the nearest prototype of each unlabeled sample is determined by calculating the absolute or relative distance between features, and the object classes of unlabeled samples are determined according to the object classes of the nearest samples or prototypes [53]. However, these methods cannot effectively utilize the prior spatial constraint knowledge between unlabeled samples and labeled samples. Therefore, researchers have begun to explore solutions to

obtain supervision from the data itself, one of which is self-supervised learning. The existing self-supervised learning methods include methods based on image reconstruction [61], spatial relationship [62], [63], color transformation [64], super-resolution [65], super-pixel [66]–[69], and spatial rotation transformation [70], [71]. Since there is a natural correlation between unlabeled samples and labeled samples, corresponding constraints can be added to the training objective (i.e., loss function) of the model according to the correlation. By adding correlation constraints, the trained model will be more robust, and the labeled samples can be fully utilized when the samples are limited. However, the aforementioned methods do not make full use of the spatial constraints between labeled and unlabeled samples. Therefore, it is necessary to explore the spatial constraints to improve the classification accuracy of HSI.

In this article, we propose an HSI classification method based on adaptive spatial pyramid constraint (ASPC). Using the multiscale spatial constraint between labeled and unlabeled samples, the proposed method can improve the generalization ability of the model and reduce the dependence on labeled samples. To make full use of the multiscale spatial constraint in HSI, an image complexity evaluation method based on an edge detection algorithm is proposed, which is used to evaluate the complexity of ground objects in HSI and its subregions. In this method, Canny edge detection is used to obtain the total number of edge pixels in the HSI subregions, and the edge samples to total samples ratios are further calculated to generate the image complexity value. For different image complexity, different segmentation parameters are given. In other words, the higher the complexity, the smaller the adaptive segmented subregion. By calculating the image complexity, we can adaptively generate over-segmented regions with different scales. Due to the uneven distribution of image complexity in the whole HSI, an image complexity evaluation method based on multiscale spatial pyramid is proposed. After the over-segmentation of multiscale regions based on a spatial pyramid, we use the spatial supervised constraint to generate the loss function of labeled over-segmented regions with labeled samples. At the same time, for unlabeled over-segmented regions, we use the spatial unsupervised constraint to generate the loss function of unlabeled over-segmented regions.

The main contributions of this article are as follows.

- 1) An image complexity evaluation method based on an edge detection algorithm is proposed, which is used to evaluate the complexity of ground objects in HSI and its subregions. According to the complexity, we can judge the homogeneity degree of ground objects in HSI and its subregions, and segment HSI in different scales according to the image complexity.
- 2) An HSI pyramid segmentation method based on spatial pyramid and image complexity is proposed. For different image complexity, the segmentation parameters are determined adaptively. In other words, the lower the complexity, the larger the adaptive segmentation area. Given the uneven distribution of image complexity in the whole image, different subregions are segmented

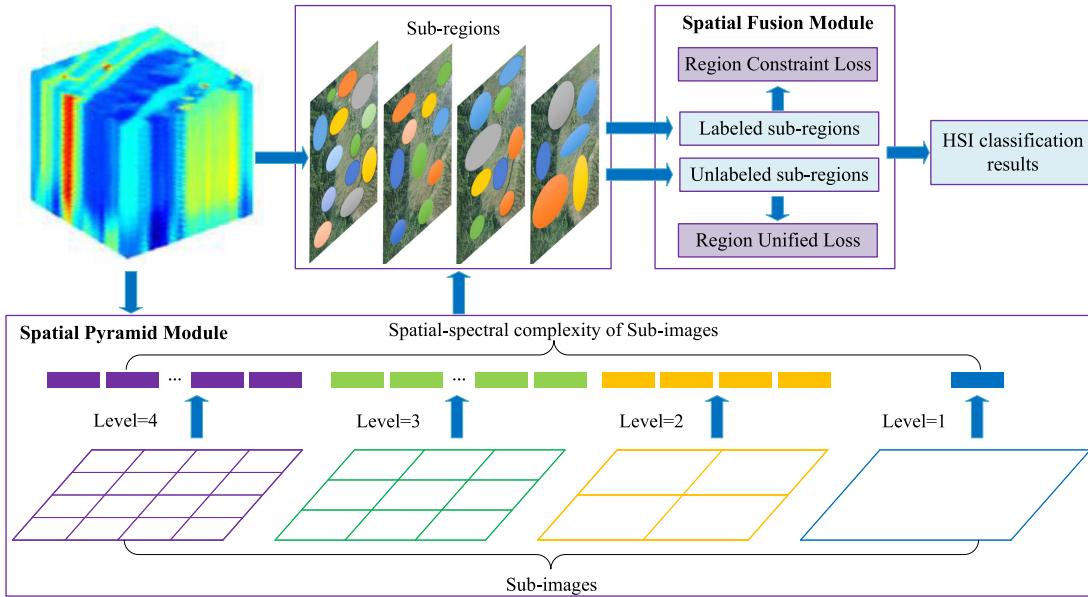


Fig. 1. Framework of the proposed method with two modules: spatial pyramid module and spatial fusion module.

according to the image complexity based on spatial pyramid.

- 3) A spatial supervised constraint and a spatial unsupervised constraint are proposed to generate the loss functions of unlabeled subregions and labeled over subregions. The ablation experiments show that these two losses can improve the performance of HSI classification with limited samples.

This article is organized as follows. In Section II, the proposed method is described in detail, including the spatial pyramid module and the spatial fusion module. In Section III, the experimental datasets and training details are described. In Section IV, the HSI classification results are shown and discussed, and an ablation study is implemented and analyzed. In Section V, the conclusions of this article are summarized.

II. METHOD

This article proposes an HSI classification method based on the ASPC, which consists of two modules: the spatial pyramid module and the spatial fusion module. First, the spatial pyramid module generates the spatial-spectral image complexity of HSI, and then performs pyramid segmentation on HSI. The spatial fusion module generates region constraint loss for labeled subregions and region unified loss for unlabeled subregions. The framework of the proposed method is shown in Fig. 1.

A. Spatial Pyramid Module

The module uses a spatial-spectral image complexity evaluation method to evaluate the complexity of HSI and its subregions, and adaptively segment HSI according to the image complexity.

1) Spatial-Spectral Image Complexity of HSI: Assume an HSI cube $\mathcal{H} \in \mathbb{R}^{X \times Y \times B}$, where X and Y are the spatial dimension and B is the number of spectral channels. For a spectral channel in HSI \mathcal{H}_i , we use Canny edge detection algorithm

to generate the number of edge pixels, which is denoted by $\text{Canny}(\mathcal{H}_i)$. Next, we calculate the ratio of the number of edge pixels of the HSI channel to the total number of pixels, which is the image complexity of each HSI channel. For a multichannel HSI, we calculate the image complexity of all HSI channels, and calculate the average value to obtain the HSI complexity value. The calculation method is defined as follows:

$$\begin{cases} C(\mathcal{H}_i) = \frac{\text{Canny}(\mathcal{H}_i)}{f_N(\mathcal{H}_i)} \quad \forall \mathcal{H}_i \in \mathcal{H} \\ C(\mathcal{H}) = \frac{1}{B} \sum_{i=1}^B C(\mathcal{H}_i) \end{cases} \quad (1)$$

where $C(\mathcal{H})$ denotes the image complexity of \mathcal{H} , $\text{Canny}(\mathcal{H}_i)$ denotes the number of edge pixels of the HSI cube \mathcal{H}_i after performing Canny edge detection and $f_N(\mathcal{H}_i)$ is the total number of pixels in \mathcal{H}_i . For a single channel image, the image complexity value is equal to the ratio of the number of edge pixels of the image to the total number of pixels. For an HSI, the image complexity value is the average of the image complexity values of all channels among the HSI, which also means that the spectral information is included. Therefore, here we also call the image complexity of HSI as the spatial-spectral image complexity. The proposed method uses edge detection to measure the complexity of ground objects in HSIs. However, the existing methods based on edge detection are used to segment HSIs as the prior knowledge of classification [68], [69], [72], so the proposed method is quite different from the existing methods based on edge detection.

The procedure of the HSI complexity evaluation method is given in Algorithm 1. By evaluating the spatial-spectral image complexity of HSI, the complexity of ground objects in HSI can be evaluated, which can be used as the basis for determining the scale of ground objects.

2) HSI Pyramid Segmentation Method Based on HSI Complexity: Because the ground coverage area of HSI is usually large, the complexity difference of local area may be

Algorithm 1 HSI Complexity Evaluation Method**Input:**An HSI cube $\mathcal{H} \in \mathbb{R}^{X \times Y \times B}$.**Output:**The spatial-spectral image complexity of HSI $C(\mathcal{H})$.

- 1: $r \leftarrow 0$
- 2: **for** $\mathcal{H}_i \in \mathcal{H}$ **do**
- 3: Extract the number of edge pixels of \mathcal{H}_i Canny(\mathcal{H}_i)
- 4: Compute the complexity value $C(\mathcal{H}_i)$ by Eq. (1)
- 5: $r \leftarrow r + C(\mathcal{H}_i)$
- 6: **end for**
- 7: $C(\mathcal{H}) \leftarrow r/B$

ignored when the complexity evaluation of the whole map is used. Therefore, we cut HSI based on the spatial pyramid to form a set of multiscale subimages. Then, the image complexity of each subimage is evaluated. We use the classical segmentation method, i.e., felzenszwalb [73] to segment each subimage, while image complexity is used to determine the scale parameters of the method. The important feature of the felzenszwalb algorithm is to retain the details in the low variation region, but ignore the details in the high variation region. According to the subimage complexity, the corresponding segmentation method is applied. That is to say, for the subimages with different image complexity values, their segmentation methods are different as well. Finally, the whole HSI is restored from the bottom to the top.

We define the total number of layers of a spatial pyramid as $N_{\mathcal{L}}$, where the number of subimages of each layer is M_n , $n \in [1, N_{\mathcal{L}}]$. For the last layer of the spatial pyramid, the number of subimages is $M_{N_{\mathcal{L}}} = N_{\mathcal{L}}^2$. All of the subimages in the subimages set will be segmented, which means that the segmentation method will be performed $M_{N_{\mathcal{L}}}$ times for the last layer of the spatial pyramid. For the n -level pyramid, we define its subimages set as $\{R_n^k\}_{k=1}^{M_n}$. For a subimage R_n^k in the n -level, we first calculate its spatial-spectral image complexity by (1). Then, we take the complexity value corresponding to subimage R_n^k as the scale parameter of segmentation. The segmentation based on image complexity can be formulated as

$$S(R_n^k, C(R_n^k)) = \left\{ \mathcal{S}_{R_n^k}^c \right\}_{c=1}^{N_{R_n^k}} \quad (2)$$

where $n \in [1, N_{\mathcal{L}}]$, $k \in [1, M_n]$, and $\mathcal{S}_{R_n^k}^c$ denotes the c th segmented region of subimage R_n^k , where the value range of c is $[1, N_{R_n^k}]$. $N_{R_n^k}$ represents the number of subregions of the subimage R_n^k . Therefore, for the n th level, the set of segmented subregions is

$$\mathcal{S}(n) = \left[\left\{ \mathcal{S}_{R_n^1}^c \right\}_{c=1}^{N_{R_n^1}}, \left\{ \mathcal{S}_{R_n^2}^c \right\}_{c=1}^{N_{R_n^2}}, \dots, \left\{ \mathcal{S}_{R_n^{M_n}}^c \right\}_{c=1}^{N_{R_n^{M_n}}} \right]. \quad (3)$$

The procedure of the HSI pyramid segmentation method is given in Algorithm 2.

B. Spatial Fusion Module

In this module, we first determine whether there are labeled samples in a subregion. If there are labels, the region constraint

Algorithm 2 HSI Pyramid Segmentation Method**Input:**An HSI cube $\mathcal{H} \in \mathbb{R}^{X \times Y \times B}$.The level of the spatial pyramid $N_{\mathcal{L}}$ **Output:**The segmented sub-region set $\mathcal{S}(N_{\mathcal{L}})$ of level $N_{\mathcal{L}}$.

- 1: Generate the sub-image set $\{R_{N_{\mathcal{L}}}^k\}_{k=1}^{M_{N_{\mathcal{L}}}}$.
- 2: **for** $R \in \{R_{N_{\mathcal{L}}}^k\}_{k=1}^{M_{N_{\mathcal{L}}}}$ **do**
- 3: Generate the sub-region set $\{\mathcal{S}_R^c\}_{c=1}^{N_R}$ by Eq. (2).
- 4: Add $\{\mathcal{S}_R^c\}_{c=1}^{N_R}$ to $\mathcal{S}(N_{\mathcal{L}})$
- 5: **end for**

loss is generated for the labeled subregion. If there is no label in the subregion, region unified loss is generated for the unlabeled subregion.

1) *Region Constraint Loss*: Given an HSI \mathcal{H} , we denote its labeled samples set and unlabeled samples set as \mathcal{L} and \mathcal{U} , respectively. For a labeled sample $p_j \in \mathcal{L}$, if the subregion of the sample is $\mathcal{S}_{R_n^k}^c$, then the corresponding region constraint loss of the subregion is formulated as

$$\mathcal{L}^C(\mathcal{S}_{R_n^k}^c) = \frac{1}{\Psi(\mathcal{S}_{R_n^k}^c)} \sum_{j=1}^{|\mathcal{S}_{R_n^k}^c|} \sum_{i=1}^{|\mathcal{S}_{R_n^k}^c|} \Delta(f(r_i), \Phi(p_j)) \quad (4)$$

where $\Phi(p_j)$ denotes the label of p_j , Δ is a *cross-entropy* function, $\mathcal{L}^C(\mathcal{S}_{R_n^k}^c)$ is the region constraint loss of subregion $\mathcal{S}_{R_n^k}^c$. r_i is the i th sample in subregion $\mathcal{S}_{R_n^k}^c$ and $f(r_i)$ represents the label generated by the forward propagation of sample r_i . $|\mathcal{S}_{R_n^k}^c|$ and $\Psi(\mathcal{S}_{R_n^k}^c)$ are the total number of samples and the total number of labeled samples in subregion $\mathcal{S}_{R_n^k}^c$, respectively. For the n -level pyramid, the region constraint loss can be formulated as

$$\mathcal{L}^C(n) = \sum_{k=1}^{M_n} \sum_{c=1}^{|\mathcal{R}_n^k|} \mathcal{L}^C(\mathcal{S}_{R_n^k}^c) \quad (5)$$

where M_n is the total number of subimages in n level pyramid, $|\mathcal{R}_n^k|$ is the total number of subregions with labeled samples in subimage R_n^k . Finally, the region constraint losses of each pyramid level are summed and averaged to obtain the final region constraint loss

$$\mathcal{L}^C = \frac{1}{N_{\mathcal{L}}} \sum_{n=1}^{N_{\mathcal{L}}} \mathcal{L}^C(n). \quad (6)$$

2) *Region Unified Loss*: For a subregion without any labeled samples $\mathcal{S}_{R_n^k}^u$, we put the pixels of the subregions into the model and get the set of forward propagation results for all pixels: $\{f(r_i)\}, r_i \in \mathcal{S}_{R_n^k}^u, i \in [1, |\mathcal{S}_{R_n^k}^u|]$. We count the results and find the most classified class, which is denoted as c_{\max} . For subregion $\mathcal{S}_{R_n^k}^u$, the corresponding region unified loss can be formulated as

$$\begin{cases} c_{\max} = \text{argmax}(g(c)), & c \in \mathcal{C} \\ \mathcal{L}^U(\mathcal{S}_{R_n^k}^u) = \sum_{i=1}^{|\mathcal{S}_{R_n^k}^u|} \Delta(f(r_i), c_{\max}), & r_i \in \mathcal{S}_{R_n^k}^u \end{cases} \quad (7)$$

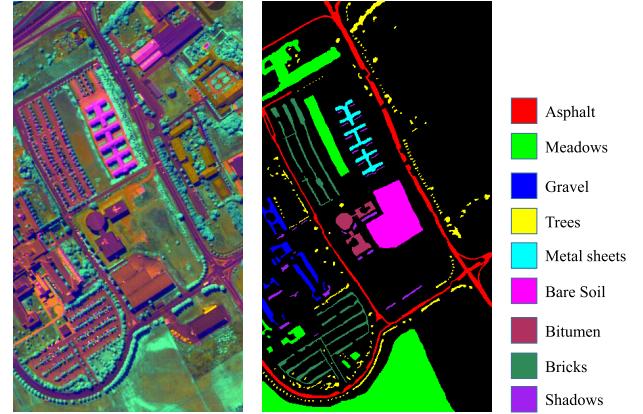
Algorithm 3 Spatial Fusion Module**Input:**An HSI cube $\mathcal{H} \in \mathbb{R}^{X \times Y \times B}$ The level of the spatial pyramid $N_{\mathcal{L}}$ The labeled samples set \mathcal{L} **Output:**Region constraint loss $\mathcal{L}^{\mathcal{C}}$ Region unified loss $\mathcal{L}^{\mathcal{U}}$ 1: $\mathcal{L}^{\mathcal{C}}, \mathcal{L}^{\mathcal{U}} \leftarrow 0$ 2: **for** $n \in [1, N_{\mathcal{L}}]$ **do**3: Generate the sub-regions set for the n -th level $\mathcal{S}(n)$ by Algorithm 2.4: $\alpha, \beta \leftarrow 0$ 5: $\mathcal{L}^{\mathcal{C}}(n) \leftarrow 0$ 6: **for** $p_j \in \mathcal{L}$ **do**7: **for** $\mathcal{S}^c \in \mathcal{S}(n)$ **do**8: **if** $p_j \in \mathcal{S}^c$ **then**9: Compute region constraint loss of \mathcal{S}^c : $\mathcal{L}^{\mathcal{C}}(\mathcal{S}^c)$ by Eq. (4)10: $\alpha \leftarrow \alpha + 1$ 11: $\mathcal{L}^{\mathcal{C}}(n) \leftarrow \mathcal{L}^{\mathcal{C}}(n) + \mathcal{L}^{\mathcal{C}}(\mathcal{S}^c)$ 12: **end if**13: **end for**14: **end for**15: $\mathcal{L}^{\mathcal{C}} = \mathcal{L}^{\mathcal{C}} + \mathcal{L}^{\mathcal{C}}(n)/\alpha$ 16: $\mathcal{L}^{\mathcal{U}}(n) \leftarrow 0$ 17: **for** $\mathcal{S}^u \in \mathcal{S}(N_{\mathcal{L}})$ **do**18: **if** $\mathcal{S}^u \cap \mathcal{L} = \emptyset$ **then**19: Generate results set $\{f(r_i)\}, r_i \in \mathcal{S}^u, i \in [1, |\mathcal{S}^u|]$ by forward propagation20: Compute the number of each class $g(c), c \in \mathcal{C}$ in results set $\{f(r_i)\}$ 21: Compute region unified loss of \mathcal{S}^u : $\mathcal{L}^{\mathcal{U}}(\mathcal{S}^u)$ by Eq. (7)22: $\beta \leftarrow \beta + 1$ 23: $\mathcal{L}^{\mathcal{U}}(n) \leftarrow \mathcal{L}^{\mathcal{U}}(n) + \mathcal{L}^{\mathcal{U}}(\mathcal{S}^u)$ 24: **end if**25: **end for**26: $\mathcal{L}^{\mathcal{U}} = \mathcal{L}^{\mathcal{U}}(n) + \mathcal{L}^{\mathcal{U}}(n)/\beta$ 27: **end for**

Fig. 2. False-color composite image, the corresponding ground-truth map, and the legend of the PU dataset.

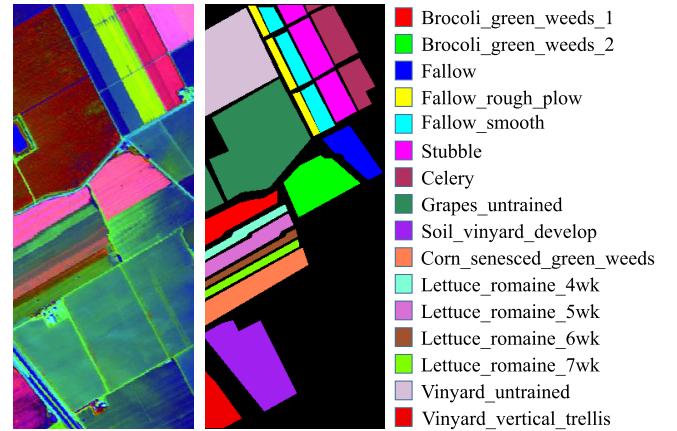


Fig. 3. False-color composite image, the corresponding ground-truth map, and the legend of the SA dataset.

where $g(c)$ represents the number of each object class, $\text{argmax}(g(c))$ returns the object class with the max value of $g(c)$, and $\mathcal{L}^{\mathcal{U}}(\mathcal{S}_{R_n^k})$ is the region unified loss of subregion $\mathcal{S}_{R_n^k}$. \mathcal{C} is the labels set of HSI \mathcal{H} . For the n th level of the spatial pyramid, we define $\lfloor R_n^k \rfloor$ as the total number of unlabeled subregions in subimage R_n^k , then the corresponding region unified loss can be calculated as

$$\lfloor R_n^k \rfloor = N_{R_n^k} - \lceil R_n^k \rceil$$

$$\mathcal{L}^{\mathcal{U}} = \frac{1}{N_{\mathcal{L}}} \sum_{n=1}^{N_{\mathcal{L}}} \sum_{k=1}^{M_n} \sum_{u=1}^{\lfloor R_n^k \rfloor} \mathcal{L}^{\mathcal{U}}(\mathcal{S}_{R_n^k}) \quad (8)$$

where $N_{R_n^k}$ is the total number of subregions in subimage R_n^k , $N_{\mathcal{L}}$ is the total number of pyramid levels.

The procedure of calculating the region constraint loss and region unified loss is given in Algorithm 3. To calculate the final loss, we first calculate the loss of hard labels, which can be formulated as

$$\mathcal{L}^{\mathcal{H}} = \sum_{i=1}^{|\mathcal{L}|} \Delta(f(r_i), \Phi(r_i)), \quad r_i \in \mathcal{L} \quad (9)$$

where $|\mathcal{L}|$ denotes the total number of samples in the labeled samples set \mathcal{L} and $\Phi(r_i)$ denotes the label of sample r_i . Finally, we combine supervised loss with hard label, region constraint loss and region unified loss, and the final loss \mathcal{L}_{all} can be calculated as

$$\mathcal{L}_{\text{all}} = \mathcal{L}^{\mathcal{H}} + \alpha \mathcal{L}^{\mathcal{C}} + \beta \mathcal{L}^{\mathcal{U}} \quad (10)$$

where α and β are the weighting factors for the region constraint loss and the region unified loss, respectively. The proposed method can be combined with mainstream deep learning backbones. In this article, we use 64-layer high-resolution net (HRnet) [74] as the backbone. Due to the characteristics of the end-to-end network model, we directly feed the whole HSI cube into the neural network during training. Finally, we train the HRnet model by optimizing the

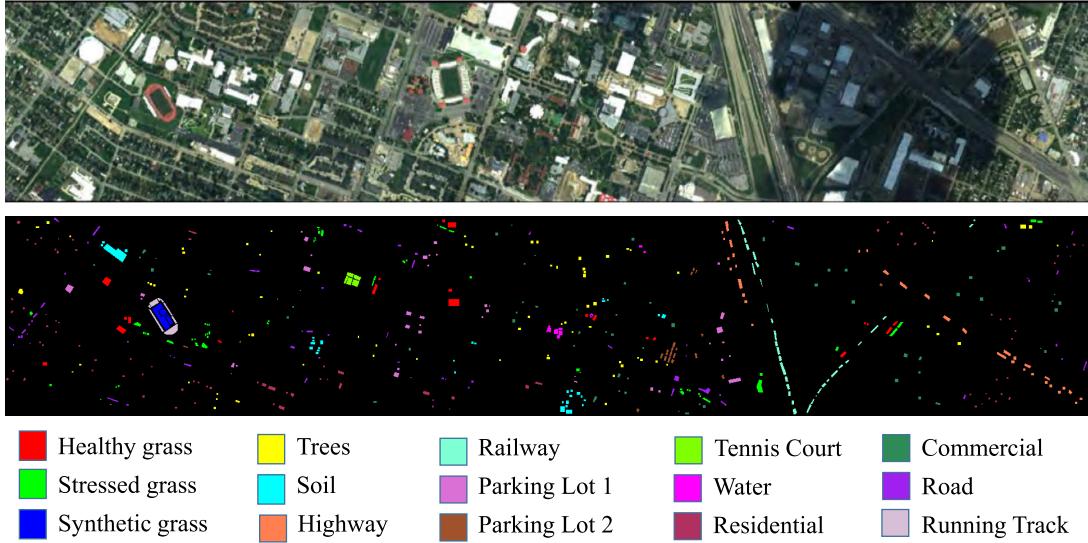


Fig. 4. False-color composite image, the corresponding ground-truth map, and the legend of the HU dataset.

TABLE I

HSI DATASETS, WHERE X AND Y ARE THE WIDTH AND THE HEIGHT OF THE HSI, RESPECTIVELY. B IS THE NUMBER OF CHANNELS AND C IS THE NUMBER OF OBJECT CLASSES OF THE HSI. A IS THE NUMBER OF ACTIVATING PIXELS

Datasets	X	Y	B	C	A
PU	610	340	103	9	42776
SA	512	217	224	16	54129
HU	1905	349	144	15	15029

loss \mathcal{L}_{all} , and generate the HSI classification results through the forward propagation of the trained model.

III. EXPERIMENTS

A. Datasets and Evaluation

To verify and demonstrate the effectiveness of the proposed method, experiments are conducted on three popular HSI datasets including the University of Pavia (PU), Salinas (SA), and Houston (HU). The detailed descriptions of the three datasets are listed in Table I.

1) *PU Dataset*: This dataset was collected by the Reflective Optics Systems Imaging Spectrometer (ROSIS) sensor at the University of Pavia in northern Italy. The spatial resolution of the image is 1.3 m, and the coverage of the image on the ground is 0.8 km \times 0.4 km. The corresponding wavelength range of this image is 0.43–0.86 μm . During data preprocessing, 12 bands that are seriously affected by noise and water absorption were removed, and the remaining 103 spectral channels were utilized in the experiment. Fig. 2 shows the false-color composite image, ground truth map and legend of the image.

2) *SA Dataset*: This dataset was collected by Airborne Visible-Infrared Imaging Spectrometer (AVIRIS) in Salinas

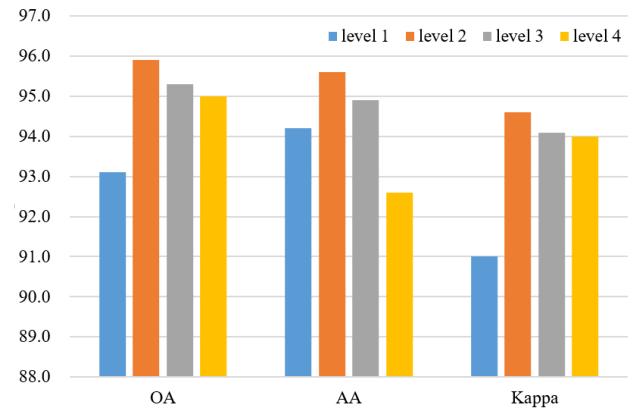


Fig. 5. Accuracies of ASPC with different levels of the spatial pyramid on the PU dataset.

Valley, CA, USA. The spatial resolution of Salinas image is 3.7 m, and the coverage is 1.9 km \times 0.8 km. The corresponding wavelength range of Salinas images is 0.4–2.5 μm . During the data preprocessing, 20 bands (108–112, 154–167 and 224) which are seriously affected by water absorption were removed, and the remaining 204 spectral channels were involved in the experiment. Fig. 3 shows the false color composite image, ground-truth map, and the legend of this image.

3) *HU Dataset*: This dataset was collected by ITRES-CASI 1500 sensor on the campus of the University of Houston and its adjacent urban area [47]. The spatial resolution of this image is 2.5 m, and the image is composed of 349 \times 1905 pixels. The corresponding wavelength range of this image is 0.38–1.05 μm . After data preprocessing (attitude processing, radiometric correction, GPS processing, and geological correction), 144 spectral channels participated in the experiment. Fig. 4 shows the false color composite image, ground-truth map, and the legend of Houston image.

TABLE II
CLASSIFICATION RESULTS (%) USING 10, 15, AND 20 LABELED SAMPLES FOR EACH CLASS ON THE PU DATASET

Class	$L = 10$				$L = 15$				$L = 20$			
	CDCNN	SSRN	FDSSC	Ours	CDCNN	SSRN	FDSSC	Ours	CDCNN	SSRN	FDSSC	Ours
1	81.6 \pm 3.2	89.8 \pm 5.0	83.8 \pm 3.2	92.6 \pm 4.1	82.4 \pm 3.0	96.9 \pm 1.8	88.2 \pm 1.8	90.2 \pm 4.6	85.5 \pm 2.7	96.9 \pm 3.5	96.5 \pm 4.6	94.7 \pm 3.9
2	80.5 \pm 4.4	84.7 \pm 4.1	85.3 \pm 3.7	84.4 \pm 6.8	83.7 \pm 4.9	92.6 \pm 2.5	95.5 \pm 2.3	95.0 \pm 2.8	89.9 \pm 3.8	95.3 \pm 2.8	95.4 \pm 2.4	97.0 \pm 1.3
3	54.9 \pm 3.8	72.7 \pm 11.2	80.8 \pm 13.2	99.0 \pm 2.0	61.3 \pm 5.0	74.5 \pm 11.5	75.3 \pm 19.1	98.9 \pm 0.9	78.2 \pm 7.5	72.9 \pm 12.4	79.7 \pm 14.3	99.8 \pm 0.4
4	89.1 \pm 6.4	74.9 \pm 9.7	61.5 \pm 7.1	77.5 \pm 4.5	94.8 \pm 4.1	93.5 \pm 7.8	87.9 \pm 6.7	81.2 \pm 6.4	95.0 \pm 3.8	96.5 \pm 4.7	85.6 \pm 7.6	77.7 \pm 4.7
5	96.4 \pm 2.3	95.6 \pm 4.3	98.3 \pm 1.7	99.9 \pm 0.3	97.4 \pm 2.3	99.9 \pm 0.3	99.7 \pm 0.2	99.9 \pm 0.1	99.3 \pm 0.5	99.9 \pm 0.2	99.7 \pm 0.3	100.0 \pm 0.0
6	48.0 \pm 3.0	61.8 \pm 12.4	95.3 \pm 4.9	98.0 \pm 2.2	53.1 \pm 5.3	82.3 \pm 12.2	87.6 \pm 11.9	98.4 \pm 1.3	56.2 \pm 5.6	88.4 \pm 5.0	91.0 \pm 4.9	99.2 \pm 1.1
7	53.9 \pm 7.9	69.9 \pm 20.3	78.6 \pm 12.6	99.9 \pm 0.2	60.5 \pm 11.1	74.7 \pm 15.3	80.5 \pm 19.6	100.0 \pm 0.0	69.1 \pm 4.0	88.8 \pm 9.8	92.3 \pm 8.6	99.8 \pm 0.4
8	64.1 \pm 5.3	68.9 \pm 5.8	73.6 \pm 5.6	89.6 \pm 5.2	73.7 \pm 3.2	70.8 \pm 6.6	75.1 \pm 13.9	97.3 \pm 1.9	81.5 \pm 2.8	78.8 \pm 8.5	79.1 \pm 11.2	97.7 \pm 1.3
9	61.7 \pm 16.5	98.1 \pm 2.5	98.2 \pm 2.2	86.9 \pm 3.1	82.8 \pm 12.2	97.8 \pm 2.6	93.7 \pm 5.2	83.2 \pm 3.0	82.3 \pm 10.5	99.1 \pm 1.1	96.2 \pm 3.0	84.9 \pm 4.3
OA	74.1 \pm 1.0	80.3 \pm 3.2	83.8 \pm 2.0	89.0\pm3.0	78.4 \pm 2.4	89.1 \pm 2.3	89.7 \pm 2.5	94.1\pm0.6	83.8 \pm 2.0	92.3 \pm 0.9	92.2 \pm 1.8	95.6\pm1.1
AA	70.0 \pm 2.4	79.6 \pm 2.4	83.9 \pm 2.8	92.0\pm1.2	76.6 \pm 2.1	87.0 \pm 2.5	87.1 \pm 3.6	93.8\pm1.0	81.9 \pm 1.1	90.7 \pm 1.7	90.6 \pm 2.9	94.5\pm1.3
κ	67.4 \pm 1.2	75.2 \pm 3.7	79.5 \pm 2.3	85.8\pm3.7	72.6 \pm 2.8	85.9 \pm 3.0	86.7 \pm 3.2	92.2\pm0.8	79.2 \pm 2.4	89.9 \pm 1.1	89.8 \pm 2.3	94.2\pm1.5

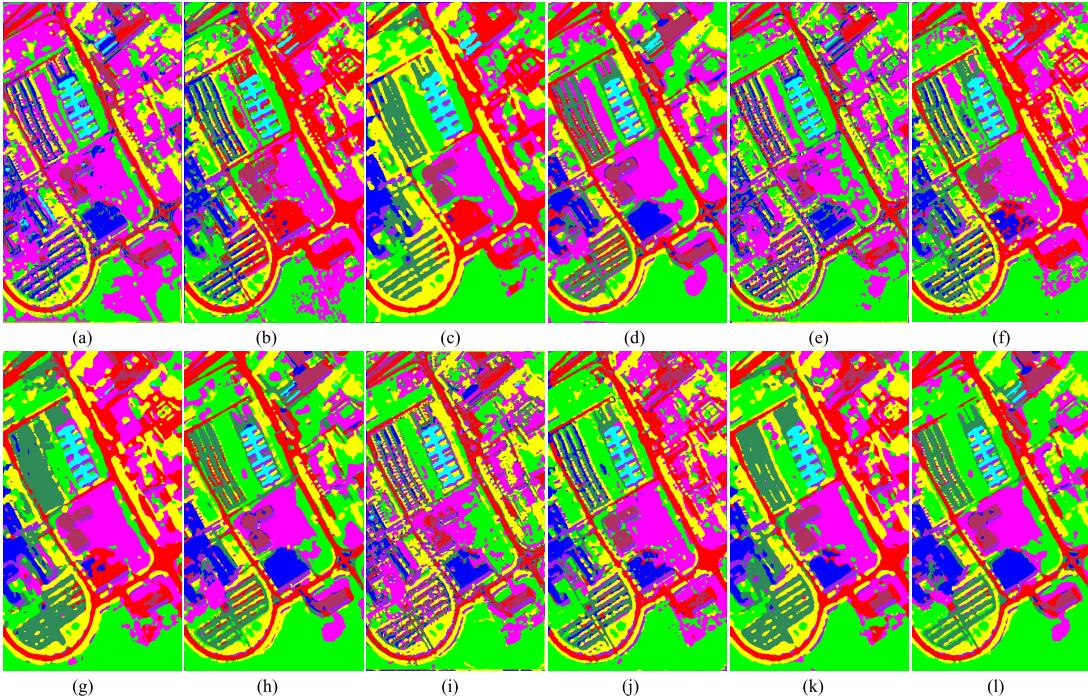


Fig. 6. Classification map of the PU dataset. (a) CDCNN ($L = 10$). (b) SSRN ($L = 10$). (c) FDSSC ($L = 10$). (d) Proposed ASPC ($L = 10$). (e) CDCNN ($L = 15$). (f) SSRN ($L = 15$). (g) FDSSC ($L = 15$). (h) Proposed ASPC ($L = 15$). (i) CDCNN ($L = 20$). (j) SSRN ($L = 20$). (k) FDSSC ($L = 20$). (l) Proposed ASPC ($L = 20$).

On each HSI dataset, we take L samples from each class for training, and the remaining active pixels A are used for testing. To avoid accidental errors, we repeated each experiment ten times and reported the average results and standard deviation of the 10 experiments. The accuracy indicators reported in this article include accuracy of each class, overall accuracy, average accuracy, and kappa coefficient.

B. Parameter Settings

To verify the influence of the pyramid level on the final accuracy and determine the best pyramid level, we set different pyramid levels and test their accuracy on the PU dataset. From the test results shown in Fig. 5, we can easily see that the

accuracy is the highest when the pyramid level is 2, so we set the pyramid level to 2 in the following experiment.

We train the HRnet model by optimizing (10), where α and β are set to 0.8 and 0.05, respectively. When training the model, the initial learning rate is set to 0.08. For every 100 epochs, the learning rate is multiplied by 0.9. The computer environment for the experiments is as follows: the processor is “Intel(R) Core i7-8700K CPU”; the graphics card is “NVIDIA GeForce GTX 1080 Ti” with “CUDA version 9.2.148”; the programming language and the deep learning platform are Python (version 3.6.10) and PyTorch (version 1.2.0), respectively.¹

¹The source code is available at <https://github.com/vision-zhu>

TABLE III
CLASSIFICATION RESULTS (%) USING 10, 15, AND 20 LABELED SAMPLES FOR EACH CLASS ON THE SA DATASET

Class	$L = 10$				$L = 15$				$L = 20$			
	CDCNN	SSRN	FDSSC	Ours	CDCNN	SSRN	FDSSC	Ours	CDCNN	SSRN	FDSSC	Ours
1	90.2 \pm 17.5	99.9 \pm 0.2	100.0 \pm 0.0	99.0 \pm 1.3	84.4 \pm 20.8	99.8 \pm 0.4	100.0 \pm 0.0	100.0 \pm 0.0	97.2 \pm 5.5	100.0 \pm 0.0	100.0 \pm 0.0	100.0 \pm 0.0
2	91.2 \pm 5.7	99.5 \pm 1.1	99.7 \pm 0.6	99.9 \pm 0.1	97.5 \pm 5.8	98.4 \pm 1.9	99.6 \pm 0.9	99.9 \pm 0.1	99.7 \pm 0.4	99.9 \pm 0.1	100.0 \pm 0.0	100.0 \pm 0.0
3	78.0 \pm 10.0	94.5 \pm 5.5	97.9 \pm 2.4	100.0 \pm 0.0	89.7 \pm 5.1	92.7 \pm 7.5	99.3 \pm 0.9	100.0 \pm 0.0	93.1 \pm 2.8	97.7 \pm 2.6	99.9 \pm 0.2	100.0 \pm 0.0
4	97.9 \pm 0.8	98.3 \pm 0.8	98.2 \pm 1.1	100.0 \pm 0.0	98.4 \pm 1.0	97.0 \pm 2.4	98.1 \pm 1.3	100.0 \pm 0.0	96.8 \pm 1.9	97.7 \pm 2.1	98.7 \pm 0.9	100.0 \pm 0.0
5	98.0 \pm 2.4	97.7 \pm 4.8	99.9 \pm 0.1	89.1 \pm 29.7	98.7 \pm 1.3	98.7 \pm 1.9	99.9 \pm 0.1	99.2 \pm 0.2	97.7 \pm 3.2	99.4 \pm 0.6	99.5 \pm 1.1	99.2 \pm 0.1
6	98.7 \pm 1.5	99.9 \pm 0.1	99.9 \pm 0.2	100.0 \pm 0.0	98.1 \pm 4.0	100.0 \pm 0.0	99.9 \pm 0.1	100.0 \pm 0.0	99.2 \pm 0.7	99.9 \pm 0.2	100.0 \pm 0.0	99.9 \pm 0.1
7	98.5 \pm 1.4	99.8 \pm 0.5	100.0 \pm 0.0	99.8 \pm 0.3	96.2 \pm 9.4	99.6 \pm 1.0	99.7 \pm 0.6	99.9 \pm 0.1	99.8 \pm 0.5	100.0 \pm 0.0	100.0 \pm 0.0	100.0 \pm 0.0
8	68.2 \pm 2.8	83.6 \pm 4.3	70.9 \pm 12.4	95.7 \pm 1.6	73.0 \pm 5.3	82.1 \pm 2.3	76.5 \pm 10.7	79.8 \pm 0.8	77.9 \pm 6.1	89.6 \pm 3.2	88.7 \pm 4.3	99.2 \pm 0.6
9	82.3 \pm 4.7	95.3 \pm 1.4	98.6 \pm 0.2	100.0 \pm 0.0	88.9 \pm 3.2	98.5 \pm 0.8	99.8 \pm 0.3	100.0 \pm 0.0	92.7 \pm 2.3	96.5 \pm 1.4	99.3 \pm 0.4	100.0 \pm 0.0
10	78.6 \pm 6.3	93.2 \pm 2.5	98.4 \pm 2.1	99.7 \pm 0.3	78.3 \pm 3.1	92.3 \pm 2.0	94.2 \pm 2.8	99.9 \pm 0.1	81.3 \pm 6.3	95.9 \pm 1.7	96.1 \pm 2.8	99.9 \pm 0.1
11	86.1 \pm 6.1	97.5 \pm 2.5	98.1 \pm 1.3	89.3 \pm 29.8	93.9 \pm 4.4	94.6 \pm 2.7	97.3 \pm 1.6	99.0 \pm 3.2	94.5 \pm 3.7	96.0 \pm 2.0	97.8 \pm 0.9	100.0 \pm 0.0
12	93.7 \pm 8.5	96.8 \pm 2.0	98.1 \pm 1.3	99.3 \pm 0.6	97.5 \pm 2.2	99.4 \pm 0.8	99.7 \pm 0.8	99.8 \pm 0.6	98.8 \pm 0.5	99.5 \pm 0.5	100.0 \pm 0.0	100.0 \pm 0.0
13	90.3 \pm 7.2	98.3 \pm 2.0	99.7 \pm 0.4	99.7 \pm 0.6	94.3 \pm 3.5	96.4 \pm 2.5	99.9 \pm 0.1	99.8 \pm 0.1	96.1 \pm 1.8	98.8 \pm 2.7	99.5 \pm 1.0	99.4 \pm 0.3
14	95.1 \pm 5.0	98.0 \pm 1.1	97.8 \pm 1.1	97.1 \pm 1.0	94.1 \pm 3.4	99.7 \pm 0.7	99.2 \pm 0.5	97.8 \pm 0.5	93.1 \pm 5.8	99.2 \pm 1.0	99.7 \pm 0.3	97.1 \pm 0.7
15	50.9 \pm 6.6	74.0 \pm 9.1	88.0 \pm 6.8	99.5 \pm 0.3	63.2 \pm 6.4	81.6 \pm 3.3	92.2 \pm 3.9	99.6 \pm 0.3	64.8 \pm 7.8	76.0 \pm 4.6	86.5 \pm 3.4	99.9 \pm 0.2
16	92.3 \pm 8.7	99.5 \pm 0.7	100.0 \pm 0.0	99.9 \pm 0.2	91.2 \pm 7.0	99.5 \pm 0.5	99.6 \pm 0.5	100.0 \pm 0.0	89.0 \pm 12.9	99.4 \pm 0.7	99.9 \pm 0.2	100.0 \pm 0.0
OA	80.3 \pm 2.4	91.4 \pm 1.4	91.6 \pm 2.4	98.1\pm2.4	84.5 \pm 2.1	92.4 \pm 0.5	93.4 \pm 1.8	99.4\pm0.2	87.3 \pm 1.9	93.5 \pm 0.5	95.4 \pm 0.5	99.7\pm0.1
AA	86.9 \pm 1.8	95.4 \pm 0.7	96.6 \pm 0.6	98.0\pm3.8	89.8 \pm 1.7	95.6 \pm 0.6	97.2 \pm 0.5	99.6\pm0.2	92.0 \pm 1.5	96.6 \pm 0.4	97.8 \pm 0.2	99.7\pm0.0
κ	78.1 \pm 2.6	90.5 \pm 1.6	90.7 \pm 2.6	97.9\pm2.6	82.8 \pm 2.3	91.5 \pm 0.6	92.7 \pm 1.9	99.3\pm0.2	85.9 \pm 2.1	92.8 \pm 0.5	94.8 \pm 0.5	99.7\pm0.1

IV. RESULTS AND DISCUSSION

A. Accuracies

To verify the effectiveness of the proposed method with limited training samples, L training samples are randomly selected from each object class. In this experiment, L was set to 10, 15, and 20, respectively. From the labeled samples set other than the training set, three labeled samples of each object class are randomly selected for validation, and the rest are used for testing. In this study, we compared the proposed method with several recently proposed state-of-the-art HSI classification methods. The comparison methods include contextual deep convolutional neural network (CDCNN) [75], spectral-spatial residual network (SSRN) [45], and fast dense spectral-spatial convolution (FDSSC) [76]. These three models are end-to-end models, and achieved state-of-the-art HSI classification performance. As a result, we choose these SOTA methods to verify the effectiveness of the proposed method based on ASPC.

For the PU dataset, we report the results of the proposed method and the comparison methods in Table II. From this table, we can see that compared with the CDCNN method, the OA of ASPC is increased by 14.9% when $L = 10$. Compared with the FDSSC method, the AA and κ of ASPC are increased by 8.1% and 6.3% when $L = 10$, respectively. Compared with the SSRN method, the OA of ASPC is increased by 5% when $L = 15$. Compared with the FDSSC method, the OA of ASPC is increased by 3.4% when $L = 20$. The classification maps of the proposed method and the comparison methods are shown in Fig. 6.

To intuitively show the separability of the features generated by the proposed method in the feature space, we use t-distributed stochastic neighbor embedding (t-SNE) [77] to visualize the features (500 samples are randomly selected from the testing set for each class, $L = 20$), and the results

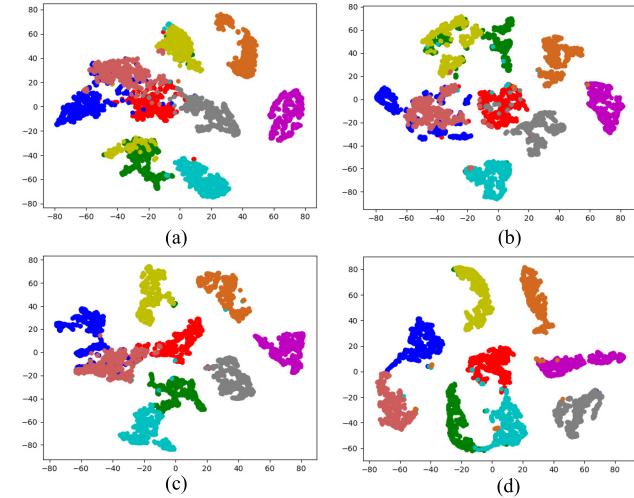


Fig. 7. t-SNE visualization results on the PU dataset. 500 samples are randomly selected from the testing set for each object class. (a) CDCNN ($L = 20$). (b) SSRN ($L = 20$). (c) FDSSC ($L = 20$). (d) Proposed ASPC ($L = 20$).

are shown in Fig. 7. From the results of t-SNE, we can see that the features generated by the proposed method have higher distinguishability than the features generated by the comparison methods.

For the SA dataset, we report the results of the proposed method and the comparison methods in Table III. From this table, we can see that compared with the CDCNN method, the OA of ASPC is increased by 17.8% when $L = 10$. Compared with FDSSC method, the OA of ASPC is increased by 4.3% when $L = 20$. Compared with the SSRN method, the OA of ASPC is increased by 7% when $L = 15$. In the case of $L = 10$, the OA, AA, and κ of ASPC are increased by 6.5%, 1.4%, and 7.2% when compared with FDSSC, respectively.

TABLE IV
CLASSIFICATION RESULTS (%) USING 10, 15, AND 20 LABELED SAMPLES FOR EACH CLASS ON THE HU DATASET

Class	$L = 10$				$L = 15$				$L = 20$			
	CDCNN	SSRN	FDSSC	Ours	CDCNN	SSRN	FDSSC	Ours	CDCNN	SSRN	FDSSC	Ours
1	91.0±6.5	96.0±3.3	95.8±3.8	80.4±5.1	91.2±5.5	88.6±8.9	90.2±1.9	92.2±4.9	88.6±8.0	90.5±3.5	86.8±6.7	98.1±1.4
2	69.6±6.6	74.6±6.4	78.9±4.2	84.8±3.7	84.6±10.4	82.3±8.8	91.2±6.7	85.2±3.5	84.5±8.2	91.9±8.7	91.9±5.5	86.3±4.7
3	59.6±12.5	77.6±5.5	81.4±4.7	99.3±0.9	75.7±10.7	79.6±3.0	83.9±3.3	99.0±0.8	80.6±15.6	80.0±4.1	86.8±3.7	99.4±0.6
4	92.7±5.1	90.8±6.2	81.1±5.1	62.8±6.2	94.9±3.0	93.0±4.3	84.8±4.3	65.3±6.4	95.3±1.9	89.2±4.2	87.5±4.6	71.6±3.4
5	89.0±3.8	91.9±6.6	93.3±5.3	99.5±0.7	89.7±1.8	91.9±4.7	89.5±5.6	99.9±0.2	95.5±2.2	92.2±4.4	97.7±3.0	99.6±0.7
6	87.9±13.1	93.0±9.0	96.1±6.4	71.4±24.3	92.4±5.3	98.4±2.1	94.1±3.6	79.9±5.3	92.9±9.6	91.4±16.2	97.9±2.8	79.1±6.0
7	75.5±2.5	78.2±12.5	78.0±12.7	77.6±6.8	78.4±5.6	77.9±10.8	82.9±6.4	78.7±3.3	84.8±3.7	89.9±3.6	87.5±7.3	80.1±4.4
8	77.4±7.7	92.8±4.4	95.5±6.4	62.3±4.6	85.4±5.9	92.5±11.7	92.6±5.5	77.7±6.2	86.6±10.0	94.1±5.2	94.3±6.0	78.9±6.2
9	73.4±10.7	79.1±8.2	76.8±12.7	68.2±6.7	82.9±5.2	82.2±9.3	77.3±11.8	74.2±4.3	83.5±5.6	86.0±6.3	86.2±12.6	77.9±3.5
10	68.6±6.6	69.5±14.1	69.8±8.6	98.7±1.3	74.3±7.5	73.0±11.8	72.3±7.3	97.2±2.7	74.5±9.2	76.7±8.6	78.9±7.7	99.8±0.5
11	72.4±8.6	67.8±8.8	65.9±8.4	87.6±0.6	76.9±4.9	84.8±7.1	82.3±8.1	93.2±4.1	75.8±6.8	80.2±10.4	85.4±11.0	99.0±0.9
12	69.6±8.1	70.3±8.0	73.2±12.4	85.8±5.9	75.2±5.1	79.0±9.7	74.5±10.9	81.6±3.2	86.5±4.1	82.7±7.8	84.7±8.3	91.7±4.3
13	71.4±7.3	93.3±6.9	80.8±8.6	77.3±6.2	75.3±9.6	89.6±7.3	75.0±11.3	86.8±3.7	79.9±7.1	87.8±8.1	77.4±6.5	83.3±5.1
14	82.0±10.8	96.0±4.7	90.1±3.6	99.6±0.6	73.1±13.2	91.1±14.4	98.0±2.0	98.6±1.5	88.6±6.1	98.5±2.5	97.6±3.5	99.4±0.9
15	91.6±5.5	77.3±4.8	96.6±2.9	99.6±0.9	94.5±4.5	80.5±3.4	98.3±1.0	97.9±1.8	91.1±4.4	97.6±1.6	98.9±1.2	99.6±0.5
OA	77.8±1.6	81.8±1.6	82.2±2.9	82.6±0.7	83.1±0.9	84.7±2.1	84.8±3.4	86.1±0.7	85.6±1.8	87.9±2.7	88.7±1.5	89.2±0.8
AA	78.1±1.7	83.2±1.2	83.6±2.1	83.7±1.8	83.0±1.0	85.6±1.6	85.8±2.5	87.2±0.6	85.9±1.5	88.6±1.9	89.3±0.9	89.6±0.8
κ	74.2±1.8	80.5±1.8	80.8±3.1	81.2±0.8	80.5±1.0	83.8±2.3	84.3±3.7	85.0±0.7	83.1±2.0	87.3±2.9	87.7±1.6	88.3±0.9

TABLE V
RESULTS OF TEN RUNS FOR THE PU, SA, AND HU DATASETS IN THE ABLATION STUDY

Datasets	PU				SA				HU			
	$L = 10$	$L = 15$	$L = 20$	$L = 10$	$L = 15$	$L = 20$	$L = 10$	$L = 15$	$L = 20$	$L = 10$	$L = 15$	$L = 20$
Method	The proposed ASPC											
OA(%)	89.0±3.0	94.1±0.6	95.6±1.1	98.1±2.4	99.4±0.2	99.7±0.1	82.6±0.7	86.1±0.7	89.2±0.8			
AA(%)	92.0±1.2	93.8±1.0	94.5±1.3	98.0±3.8	99.6±0.2	99.7±0.0	83.7±1.8	87.2±0.6	89.6±0.8			
κ (%)	85.8±3.7	92.2±0.8	94.2±1.5	97.9±2.6	99.3±0.2	99.7±0.1	81.2±0.8	85.0±0.7	88.3±0.9			
Method	ASPC without region unified loss											
OA(%)	86.4±1.6	92.4±2.5	95.3±1.6	97.3±2.0	99.0±0.3	99.1±0.2	81.7±0.8	85.6±0.9	89.0±0.7			
AA(%)	91.0±1.4	92.7±3.7	95.3±1.1	96.7±3.2	99.4±0.2	99.5±0.2	82.8±1.0	86.9±0.7	89.0±1.4			
κ (%)	82.7±1.9	90.1±3.2	93.9±2.1	97.0±2.2	98.9±0.4	99.0±0.2	80.2±0.9	84.5±1.0	88.1±0.7			
Method	ASPC without region constraint loss											
OA(%)	83.8±5.5	85.0±3.0	95.2±1.0	94.2±1.2	94.8±1.0	95.6±0.8	78.3±1.7	82.8±2.0	87.3±1.1			
AA(%)	87.3±1.8	91.5±1.9	94.4±1.1	95.8±0.6	97.0±0.6	97.5±0.3	79.7±1.5	84.5±1.8	88.1±1.1			
κ (%)	79.4±6.5	81.0±3.6	93.6±1.2	93.5±1.3	94.3±1.2	95.2±0.9	76.5±1.9	81.4±2.1	86.3±1.2			
Method	ASPC without spatial fusion module											
OA(%)	78.1±2.4	82.4±1.8	90.9±2.0	93.6±0.9	94.6±0.7	95.5±0.5	78.0±1.0	82.3±1.2	86.6±1.6			
AA(%)	84.1±1.9	89.6±1.5	93.9±1.1	95.5±0.6	96.9±0.3	97.5±0.3	80.1±0.9	84.2±1.3	87.8±1.4			
κ (%)	72.5±2.8	77.8±2.1	88.2±2.4	92.8±1.0	94.0±0.8	95.0±0.5	76.2±1.0	80.9±1.3	85.5±1.7			

In the case of $L = 20$, the OA, AA, and κ of ASPC are increased by 6.2%, 3.1%, and 6.9% when compared with SSRN, respectively. The classification maps of the proposed method and the comparison methods are shown in Fig. 8.

For the HU dataset, we report the results of the proposed method and the comparison methods in Table IV. From this table, we can see that compared with the CDCNN method, the OA of ASPC is increased by 4.8% when $L = 10$. In the case of $L = 15$, the OA, AA, and κ of ASPC are increased by 3%, 4.2%, and 4.5% when compared with CDCNN, respectively. In the case of $L = 20$, the OA, AA, and κ of ASPC are increased by 3.6%, 3.7%, and 5.2%

when compared with CDCNN, respectively. Compared with the SSRN method, the OA of ASPC is increased by 1.4% when $L = 15$. Compared with the FDSSC method, the OA of ASPC is increased by 0.5% when $L = 20$. The classification maps of the proposed method and the comparison methods are shown in Fig. 9.

B. Ablation Study

To verify the effectiveness of each module of the proposed method, we conducted an ablation study. In our experimental setup, ASPC without the region constraint loss, ASPC without

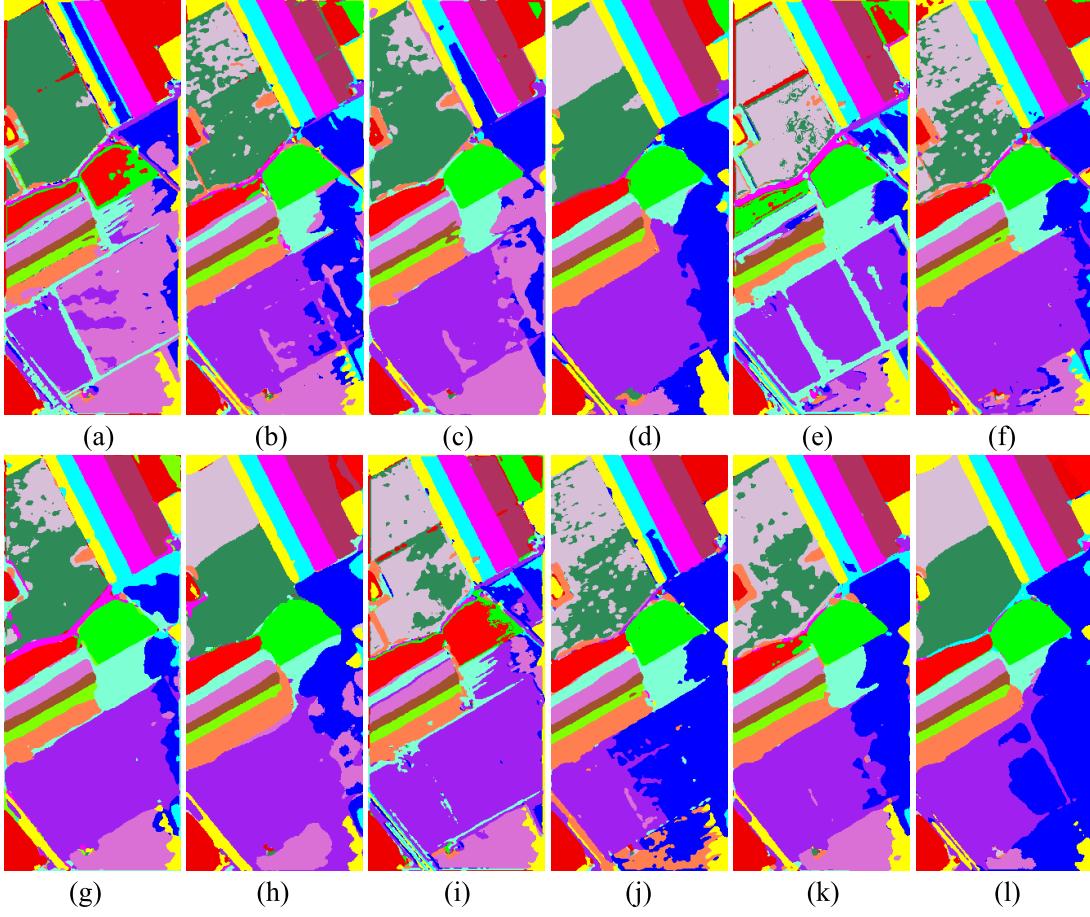


Fig. 8. Classification map of the SA dataset. (a) CDCNN ($L = 10$). (b) SSRN ($L = 10$). (c) FDSSC ($L = 10$). (d) Proposed ASPC ($L = 10$). (e) CDCNN ($L = 15$). (f) SSRN ($L = 15$). (g) FDSSC ($L = 15$). (h) Proposed ASPC ($L = 15$). (i) CDCNN ($L = 20$). (j) SSRN ($L = 20$). (k) FDSSC ($L = 20$). (l) Proposed ASPC ($L = 20$).

the region unified loss and ASPC without the spatial fusion module were compared with our proposed method. The results of the ablation study are shown in Table V.

For the PU dataset, when the region constraint loss is removed, the accuracy (OA) of ASPC decreases by 5.2% in the case of $L = 10$. The accuracy (OA) of ASPC decreases by 10.9% on the PU dataset ($L = 10$) when the spatial fusion module is removed. In the case of $L = 15$, when the region unified loss is removed, the accuracy (OA) of ASPC decreases by 1.7%. When the region constraint loss is removed, the AA and κ of ASPC decreases by 4.7% and 6.4% ($L = 10$). In the case of $L = 20$, when the spatial fusion module is removed, the accuracy (OA) of ASPC decreases by 4.7%.

For the SA dataset, when the region unified loss is removed, the accuracy (OA) of ASPC decreases by 0.8% in the case of $L = 10$. When the spatial fusion module is removed, the accuracy (OA) of ASPC decreases by 4.5% on SA dataset ($L = 10$). In the case of $L = 15$, when the region constraint loss is removed, the accuracy (OA) of ASPC decreases by 4.6%. When the region constraint loss is removed, the AA and κ of ASPC decreases by 2.6% and 5% ($L = 15$). Meanwhile, in the case of $L = 20$, when the spatial fusion module is removed, the accuracy (OA) of ASPC decreases by 4.2%.

For the HU dataset, when the spatial fusion module is removed, the accuracy (OA) of ASPC decreases by 2.6% in the case of $L = 20$. When the region constraint loss is removed, the accuracy (OA) of ASPC decreases by 1.9% on HU dataset ($L = 20$). In the case of $L = 15$, when the spatial fusion module is removed, the accuracy (OA) of ASPC decreases by 3.8%. In the case of $L = 10$, when the region unified loss is removed, the accuracy (OA) of ASPC decreases by 0.9%. When the region constraint loss is removed, the AA and κ of ASPC decreases by 1.5% and 2% ($L = 20$). Therefore, this experiment proves the effectiveness of the region constraint loss, region unified loss and spatial fusion module proposed in this article.

C. Computational Cost

In this article, we compare the training time and test time between the proposed method and the comparison methods. For the PU dataset, the training time and inference time are shown in Table VI. For the SA dataset, the training time and inference time are shown in Table VII. For the HU dataset, the training time and inference time are shown in Table VIII. The experimental results show that the training time and

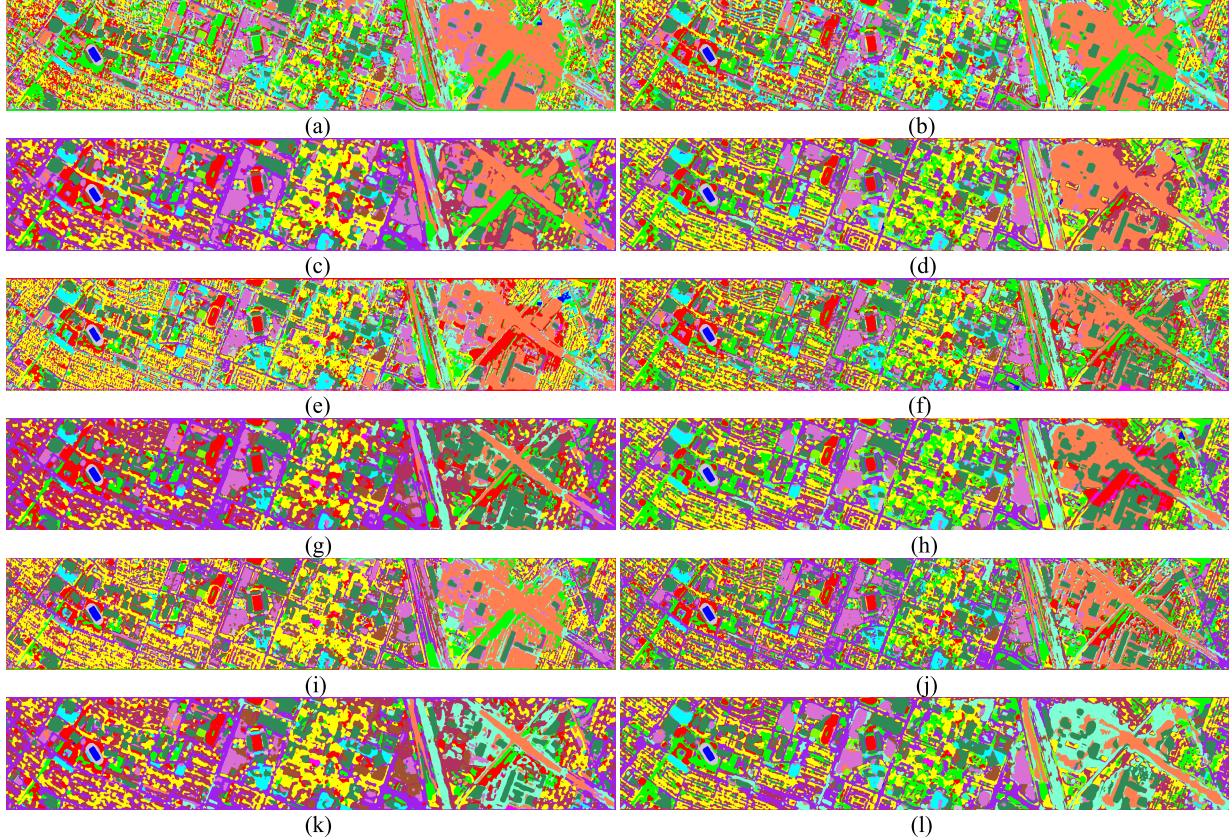


Fig. 9. Classification map of the HU dataset using 10, 15, and 20 labeled samples for each class. (a) CDCNN ($L = 10$). (b) SSRN ($L = 10$). (c) FDSSC ($L = 10$). (d) Proposed ASPC ($L = 10$). (e) CDCNN ($L = 15$). (f) SSRN ($L = 15$). (g) FDSSC ($L = 15$). (h) Proposed ASPC ($L = 15$). (i) CDCNN ($L = 20$). (j) SSRN ($L = 20$). (k) FDSSC ($L = 20$). (l) Proposed ASPC ($L = 20$).

TABLE VI
TRAINING AND INFERENCE TIME (s) ON THE PU DATASET

Methods	Training time			Inference time		
	$L = 10$	$L = 15$	$L = 20$	$L = 10$	$L = 15$	$L = 20$
CDCNN	65.50	92.71	118.49	29.81	29.80	29.40
SSRN	78.07	107.08	129.14	62.94	62.61	62.37
FDSSC	150.10	237.81	281.56	75.67	75.97	75.90
ASPC	49.51	47.35	38.81	5.52	5.52	5.53

TABLE VIII
TRAINING AND INFERENCE TIME (s) ON THE HU DATASET

Methods	Training time			Inference time		
	$L = 10$	$L = 15$	$L = 20$	$L = 10$	$L = 15$	$L = 20$
CDCNN	65.16	168.23	93.32	11.74	11.32	10.84
SSRN	202.74	266.77	444.57	27.09	26.93	26.81
FDSSC	573.38	872.84	1009.94	32.74	32.57	32.23
ASPC	137.79	223.67	244.38	5.65	5.62	5.58

TABLE VII
TRAINING AND INFERENCE TIME (s) ON THE SA DATASET

Methods	Training time			Inference time		
	$L = 10$	$L = 15$	$L = 20$	$L = 10$	$L = 15$	$L = 20$
CDCNN	48.44	78.51	81.66	48.12	48.19	48.45
SSRN	366.26	554.92	856.44	135.87	135.87	135.94
FDSSC	954.16	1676.74	2081.07	162.50	162.63	162.32
ASPC	26.48	35.55	44.38	5.45	5.43	5.42

inference time of the proposed method are less than those of the comparison methods.

V. CONCLUSION

This article proposed an HSI classification method with limited labeled samples based on ASPC, including the spatial

pyramid module and spatial fusion module. The spatial pyramid module includes the spatial-spectral image complexity evaluation method and HSI multiscale complexity evaluation method based on the spatial pyramid. Through this module, we can evaluate the complexity of ground objects in each subimages of an HSI image, and this parameter can be used as the basis for determining the separation scale. At the same time, based on the multiscale complexity evaluation, a set of HSI subregions is generated adaptively. The spatial fusion module determines whether there are labeled samples in each subregion. For the labeled subregion, the region constraint loss is generated. If there are no labeled samples in a subregion, then the region unified loss is generated for this subregion. The main conclusions are as follows.

- 1) Experimental results on three public datasets suggest that the proposed method works more accurately than

- the existing state-of-the-art methods under limited training samples.
- 2) The ablation study and the parameter analysis with the pyramid level demonstrate the effectiveness of both the spatial pyramid module and the spatial fusion module.
 - 3) The time consumption experiments show that the proposed method takes less time than the compared methods in the training phase, and is more efficient in the inference phase.

In the future, we will further explore the object-based HSI complexity evaluation method and spatial constraint method for HSI classification with limited labeled samples.

ACKNOWLEDGMENT

The authors would like to thank the National Center for Airborne Laser Mapping (NCALM), the University of Houston and the IEEE Image Analysis and Data Fusion Technical Committee for providing the Houston dataset, and the University of Pavia and Grupo de Inteligencia Computacional (GIC) for providing the Salinas dataset and the University of Pavia dataset. They would also like to thank the editors and anonymous reviewers for their valuable comments and suggestions.

REFERENCES

- [1] J. A. Benediktsson and P. Ghamisi, *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*. Norwood, MA, USA: Artech House, 2015.
- [2] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, Mar. 2017.
- [3] S. Liu, Q. Shi, and L. Zhang, "Few-shot hyperspectral image classification with unknown classes using multitask deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 1–18, Jun. 2020.
- [4] S. Prasad and L. M. Bruce, "Limitations of principal components analysis for hyperspectral target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 625–629, Oct. 2008.
- [5] Q. Du and H. Ren, "Real-time constrained linear discriminant analysis to target detection and classification in hyperspectral imagery," *Pattern Recognit.*, vol. 36, no. 1, pp. 1–12, Jan. 2003.
- [6] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.
- [7] J. R. Harris, D. Rogge, R. Hitchcock, O. Ijewliw, and D. Wright, "Mapping lithology in Canada's arctic: Application of hyperspectral data using the minimum noise fraction transformation and matched filtering," *Can. J. Earth Sci.*, vol. 42, no. 12, pp. 2173–2193, Dec. 2005.
- [8] B. Rasti *et al.*, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Dec. 2020.
- [9] 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.
- [10] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 894–898, Sep. 2011.
- [11] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognit.*, vol. 43, no. 1, pp. 331–341, Jan. 2010.
- [12] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Improved manifold coordinate representations of large-scale hyperspectral scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2786–2803, Oct. 2006.
- [13] Y. W. Chen and X. H. Han, *Classification of High-Resolution Satellite Images Using Supervised Locality Preserving Projections*. Berlin, Germany: Springer, 2008.
- [14] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, Mar. 2005.
- [15] P. R. Marpu, M. Pedergnana, M. D. Mura, J. A. Benediktsson, and L. Bruzzone, "Automatic generation of standard deviation attribute profiles for spectral-spatial classification of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 293–297, Mar. 2013.
- [16] W. Liao, R. Bellens, A. Pizurica, W. Philips, and Y. Pi, "Classification of hyperspectral data over urban areas using directional morphological profiles and semi-supervised feature extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1177–1190, Aug. 2012.
- [17] P. Ghamisi and J. A. Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 309–313, Feb. 2015.
- [18] P. Ghamisi, J. A. Benediktsson, G. Cavallaro, and A. Plaza, "Automatic framework for spectral-spatial classification based on supervised feature extraction and morphological attribute profiles," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2147–2160, Jun. 2014.
- [19] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Tensor discriminative locality alignment for hyperspectral image spectral-spatial feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 242–256, Jan. 2013.
- [20] H. Pu, Z. Chen, B. Wang, and G. Jiang, "A novel spatial-spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7008–7022, Nov. 2014.
- [21] Y. Zhou, J. Peng, and C. L. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, Feb. 2015.
- [22] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.
- [23] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
- [24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, vol. 1. Red Hook, NY, USA: Curran Associates, 2012, pp. 1097–1105.
- [25] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
- [26] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [27] Z. Tian, H. Zhao, M. Shu, Z. Yang, R. Li, and J. Jia, "Prior guided feature enrichment network for few-shot segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Aug. 3, 2020, doi: [10.1109/TPAMI.2020.3013717](https://doi.org/10.1109/TPAMI.2020.3013717).
- [28] Q. Shi *et al.*, "Domain adaption for fine-grained urban village extraction from satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1430–1434, Aug. 2020.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [30] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [31] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.
- [32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
- [33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [34] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Mar. 23, 2020, doi: [10.1109/TPAMI.2020.2982166](https://doi.org/10.1109/TPAMI.2020.2982166).
- [35] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

- [36] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2480–2495, Jul. 2021.
- [37] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, "Denoising prior driven deep neural network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2305–2318, Oct. 2019.
- [38] Q. Shi, X. Tang, T. Yang, R. Liu, and L. Zhang, "Hyperspectral image denoising using a 3-D attention denoising network," *IEEE Trans. Geosci. Remote Sens.*, early access, Jan. 8, 2021, doi: 10.1109/TGRS.2020.3045273.
- [39] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.
- [40] L. Jiao, M. Liang, H. Chen, S. Yang, H. Liu, and X. Cao, "Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5585–5599, Oct. 2017.
- [41] X. Ma, A. Fu, J. Wang, H. Wang, and B. Yin, "Hyperspectral image classification based on deep deconvolution network with skip architecture," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4781–4791, Aug. 2018.
- [42] J. Yue, S. Mao, and M. Li, "A deep learning framework for hyperspectral image classification using spatial pyramid pooling," *Remote Sens. Lett.*, vol. 7, no. 9, pp. 875–884, Sep. 2016.
- [43] Y. S. Chen, X. Zhao, and X. P. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.
- [44] P. Zhong, Z. Gong, S. Li, and C.-B. Schonlieb, "Learning to diversify deep belief networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3516–3530, Jun. 2017.
- [45] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.
- [46] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral-spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 740–754, Feb. 2019.
- [47] 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.
- [48] C. Zhang, J. Yue, and Q. Qin, "Deep quadruplet network for hyperspectral image classification with a small number of samples," *Remote Sens.*, vol. 12, no. 4, p. 647, Feb. 2020.
- [49] K. Gao, W. Guo, X. Yu, B. Liu, A. Yu, and X. Wei, "Deep induction network for small samples classification of hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3462–3477, Jun. 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9119099>
- [50] C. Zhang, J. Yue, and Q. Qin, "Global prototypical network for few-shot hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4748–4759, Aug. 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9170766>
- [51] M. S. Aydemir and G. Bilgin, "Semisupervised hyperspectral image classification using small sample sizes," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 621–625, May 2017.
- [52] S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," in *Proc. ICLR*, 2016, pp. 1–11.
- [53] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4077–4087.
- [54] B. Hariharan and R. Girshick, "Low-shot visual recognition by shrinking and hallucinating features," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3018–3027.
- [55] S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taixé, D. Cremers, and L. Van Gool, "One-shot video object segmentation," in *Proc. CVPR*, Jul. 2017, pp. 221–230.
- [56] E. Schwartz *et al.*, "Delta-encoder: An effective sample synthesis method for few-shot object recognition," in *Proc. NIPS*, 2018, pp. 1–11.
- [57] W. Li, L. Wang, J. Xu, J. Huo, Y. Gao, and J. Luo, "Revisiting local descriptor based image-to-class measure for few-shot learning," in *Proc. CVPR*, Jun. 2019, pp. 7260–7268.
- [58] X. Dong, L. Zheng, F. Ma, Y. Yang, and D. Meng, "Few-example object detection with model communication," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1641–1654, Jul. 2019.
- [59] B. Liu, X. Yu, A. Yu, P. Zhang, G. Wan, and R. Wang, "Deep few-shot learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2290–2304, Apr. 2019.
- [60] J. Kang, R. Fernandez-Beltran, Z. Ye, X. Tong, P. Ghamisi, and A. Plaza, "Deep metric learning based on scalable neighborhood components for remote sensing scene characterization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8905–8918, Dec. 2020.
- [61] L. Dinh, D. Krueger, and Y. Bengio, "NICE: Non-linear independent components estimation," 2014, *arXiv:1410.8516*. [Online]. Available: <https://arxiv.org/abs/1410.8516>
- [62] I. Misra and L. van der Maaten, "Self-supervised learning of pretext-invariant representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6707–6717.
- [63] C. Doersch, A. Gupta, and A. A. Efros, "Unsupervised visual representation learning by context prediction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1422–1430.
- [64] G. Larsson, M. Maire, and G. Shakhnarovich, "Learning representations for automatic colorization," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2016, pp. 577–593.
- [65] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4681–4690.
- [66] Y. Wang *et al.*, "Self-supervised low-rank representation (SSLRR) for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5658–5672, Oct. 2018.
- [67] Y. Wang *et al.*, "Self-supervised feature learning with CRF embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2628–2642, May 2019.
- [68] S. Jia, B. Deng, J. Zhu, X. Jia, and Q. Li, "Local binary pattern-based hyperspectral image classification with superpixel guidance," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 749–759, Feb. 2018.
- [69] Z. Xue, S. Zhou, and P. Zhao, "Active learning improved by neighborhoods and superpixels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 469–473, Mar. 2018.
- [70] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," in *Proc. ICLR*, 2018, pp. 1–16.
- [71] L. Jing, X. Yang, J. Liu, and Y. Tian, "Self-supervised spatiotemporal feature learning via video rotation prediction," 2018, *arXiv:1811.11387*. [Online]. Available: <https://arxiv.org/abs/1811.11387>
- [72] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, Jul. 2010. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S003132031000049X>
- [73] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.
- [74] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5686–5696.
- [75] 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.
- [76] W. Wang, S. Dou, Z. Jiang, and L. Sun, "A fast dense spectral-spatial convolution network framework for hyperspectral images classification," *Remote Sens.*, vol. 10, no. 7, p. 1068, Jul. 2018, doi: 10.3390/rs10071068.
- [77] L. van der Maaten, "Accelerating t-SNE using tree-based algorithms," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, Oct. 2014.



Jun Yue received the B.Eng. degree in geodesy from Wuhan University, Wuhan, China, in 2013, and the Ph.D. degree in GIS from Peking University, Beijing, China, in 2018.

He is an Assistant Professor with the Department of Geomatics Engineering, Changsha University of Science and Technology, Changsha, China. His research interests include satellite image understanding, pattern recognition, and few-shot learning.

Dr. Yue serves as a Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the ISPRS Journal of Photogrammetry and Remote Sensing, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, the IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, Information Fusion, Information Sciences, the International Journal of Remote Sensing, and Remote Sensing Letters.



Dingshun Zhu received the B.S. degree from South-Central University for Nationalities, Wuhan, China, in 2020. He is pursuing the M.S. degree with the College of Electrical and Information Engineering, Hunan University, Changsha, China.

His research interests include remote sensing image processing, object detection, and lightweight models.



Pedram Ghamisi (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2015.

He works as the Head of the Machine Learning Group, Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Dresden, Germany, and as the Visiting Professor with the Institute of Advanced Research in Artificial Intelligence (IARAI), Vienna, Austria. He is also the CTO and a Co-Founder of VasoGnosis Inc., Milwaukee, WI, USA. His research interests include interdisciplinary research on machine (deep) learning, image and signal processing, and multisensor data fusion.

Dr. Ghamisi was a recipient of the Best Researcher Award for M.Sc. students at the K. N. Toosi University of Technology in the academic year of 2010–2011, the IEEE Mikio Takagi Prize for winning the Student Paper Competition at the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) in 2013, the Talented International Researcher by Iran's National Elites Foundation in 2016, the First Prize of the Data Fusion Contest organized by the Image Analysis and Data Fusion Technical Committee (IADF) of the IEEE-GRSS in 2017, the Best Reviewer Prize of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL) in 2017, the IEEE Geoscience and Remote Sensing Society 2020 Highest Impact Paper Award, the Alexander von Humboldt Fellowship from the Technical University of Munich, and the High Potential Program Award from HZDR. He is also the Co-Chair of the IEEE Image Analysis and Data Fusion Committee. For detailed info, please see <http://pedram-ghamisi.com/>.



Leyuan Fang (Senior Member, IEEE) received the Ph.D. degree from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2015.

From September 2011 to September 2012, he was a Visiting Ph.D. Student with the Department of Ophthalmology, Duke University, Durham, NC, USA, supported by the China Scholarship Council. From August 2016 to September 2017, he was a Post-Doctoral Researcher with the Department of Biomedical Engineering, Duke University, Durham, NC, USA. He is a Professor with the College of Electrical and Information Engineering, Hunan University, and an Adjunct Researcher with the Peng Cheng Laboratory, Shenzhen, China. His research interests include sparse representation and multiresolution analysis in remote sensing and medical image processing.

Dr. Fang was a recipient of one 2nd-Grade National Award at the Nature and Science Progress of China in 2019. He is an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and *Neurocomputing*.



Yaowei Wang (Member, IEEE) received the Ph.D. degree in computer science from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2005.

He is a tenured Associate Professor with the Peng Cheng Laboratory, Shenzhen, China. He was with the Department of Electronics Engineering, Beijing Institute of Technology, Beijing, from 2005 to 2019. He was a Professor with the National Engineering Laboratory for Video Technology Shenzhen (NELVT), Peking University Shenzhen Graduate School, Shenzhen, in 2019. From 2014 to 2015, he was an Academic Visitor with the Vision Laboratory, Queen Mary University of London, London, U.K. He has authored or coauthored more than 60 refereed journals and conference papers. His research interests include machine learning, multimedia content analysis, and understanding.

Dr. Wang is a member of the China Computer Federation (CCF), the Chinese Institute of Electronics (CIE), and the China Society of Image and Graphics (CSIG). He was a recipient of the Second Prize of the National Technology Invention in 2017 and the First Prize of the CIE Technology Invention in 2015. His team was ranked as one of the best performers in the TRECVID CCD/SED tasks from 2009 to 2012 and in PETS 2012.