

Hyperspectral Image Classification with Squeeze Multi-Bias Network

Leyuan Fang, *Senior Member, IEEE*, Guangyun Liu, Shutao Li, *Senior Member, IEEE*, Pedram Ghamisi, *Senior Member, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

Abstract—Convolutional neural network (CNN) has recently demonstrated its outstanding capability for the classification of hyperspectral images (HSIs). Typical CNN-based methods usually adopt image patches as inputs to the network. But, a fixed-size image patch in HSI with complex spatial contexts may contain multiple ground objects of different classes, which will deteriorate the classification performance of CNN. In addition, traditional convolutional layers adopted in the CNN have a huge amount of parameters needed to be tuned, which will cause high computational cost. To address the above issues, a novel squeeze multi-bias network (SMBN) is proposed for HSI classification. Specifically, the proposed SMBN first introduces the multi-bias module (MBM), which incorporates multi-bias into the rectified linear unit (ReLU) layers. The MBM can decouple the feature maps of input patches into multiple response maps (corresponding to different ground objects), and adaptively select the meaningful maps for classification. Furthermore, the proposed SMBN replaces the traditional convolutional layer with a squeeze convolution module (SCM), which can greatly reduce the number of parameters in the network, thus saving the running time, while still maintaining high classification accuracy. Experimental results on three real HSIs demonstrate the superiority of the proposed SMBN method over several state-of-the-art classification approaches.

Index Terms—Hyperspectral image classification, convolutional neural network, squeeze multi-bias network, multi-bias module, squeeze convolution module.

I. INTRODUCTION

REMOTE sensing hyperspectral images (HSIs) are usually composed of hundreds of spectral channels, which can provide abundant spectral information of land covers. Recently, HSI classification has become a very active research topic in remote sensing [1]–[3], which has been widely applied in many fields, such as precision agriculture [4], environment monitoring [5], and other fields [6]–[8].

In general, there are two major characteristics of HSIs which can be utilized to obtain discriminative features for HSI

This paper was supported by the National Natural Science Fund of China for International Cooperation and Exchanges under Grant 61520106001, the National Natural Science Foundation under Grant No. 61771192, and the Fund of Hunan Province for Science and Technology Plan Project under Grant 2017RS3024.

L. Fang, G. Liu and S. Li are with the College of Electrical and Information Engineering, Hunan University, Changsha, 410082, China, and also with the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Changsha, 410082, China (e-mail: fangleyuan@gmail.com; liuguangyun@hnu.edu.cn; shutao_li@hnu.edu.cn).

P. Ghamisi is with the Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, 09599 Freiberg, Germany (e-mail: p.ghamisi@gmail.com).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland (e-mail: benedikt@hi.is).

classification. One is the spectral information which makes such data a valuable source for the accurate identification of a particular ground material [9], [10]. The other is the spatial information which can be used in addition to spectral features to further improve the classification performance [11].

Recently, researchers on HSI classification communities try to incorporate spatial with spectral information to obtain good classification performance. For instance, in [12], extended morphological profiles (EMPs) were introduced to extract spectral-spatial features, which can be used to obtain a detailed multilevel characterization of HSIs, leading to higher classification accuracies. In [13], [14], some effective multiple kernel learning (MKL) algorithms have been proposed for HSI classification and the MKL uses subspace method to obtain the weights of base kernels in the linear combination. In addition, the sparse representation [15]–[19] is also a widely-used spectral-spatial method for HSI classification, which first exploits the spatial context of the HSIs by defining a local region for each test pixel, and then jointly represents pixels with each region by a set of training samples.

The aforementioned methods depend on human-engineered features [20]–[23]. Nevertheless, it is difficult to find the most optimal parameters to generate features for the classification. Very recently, the deep learning-based methods, which can automatically extract robust and discriminative features in a hierarchical learning way [24], have made a great success in several applications (e.g., object detection [25], natural language processing [26], and image classification [27]). In the remote sensing field, Lu *et al.* have proposed a highly effective and promising feature representation learning method [28], which provides a good idea for remote sensing image processing. A stacked auto-encoder (SAE) was first introduced for HSI classification in [29]. Then, some improved SAE-based methods [30]–[32] have been proposed to obtain better classification performance. In [33], a deep belief network (DBN) model was introduced for HSI classification, which uses the restricted Boltzmann machine and the multiple-layer network to learn features. In [34], a deep convolutional neural network (CNN) was employed to directly classify HSIs in the spectral domain. After that, a series of improved CNN-based methods [35]–[40] were utilized to extract spectral-spatial features for HSI classification. For examples, work in [35] used CNN to extract spatial features, and combined them with spectral features that learned by balanced local discriminant embedding for HSI classification. In [36], a semi-supervised CNN is proposed, which can effectively solve the problem of limited labeled samples via semi-supervised learning. In [37],

Zhang *et al.* proposed a dual stream CNN, where one CNN stream extracts the spectral feature, and the other extracts the spatial-spectral feature. In [38], a CNN-based feature extractor was proposed, which can learn discriminative features from pixel pairs and apply voting strategy to create classification result maps. In [39], Song *et al.* proposed a deep feature fusion network (DFFN) for HSI classification, which can consider correlated information among different layers via a fusion strategy, and use residual learning to alleviate overfitting and gradient vanishing in the deep CNN. The CNN-based methods generally use image patches as inputs of the network, and adopt several convolutional layers and rectified linear unit (ReLU) layers for the feature extraction. The convolutional layer can produce various discriminative feature maps by applying convolution operation on input image patches or feature maps, and the ReLU was used to separate noise from informative signals in the feature maps [27], [41]. But, since the HSIs usually have complex spatial contexts, the fixed-size image patches may contain multiple ground objects of different classes, which will deteriorate the classification performance of CNN. In addition, traditional convolutional layers adopted in the CNN have a huge amount of parameters needed to be tuned, which will result in a high computational cost.

In this paper, to address the aforementioned two issues, a novel squeeze multi-bias network (SMBN) is proposed for HSI classification. Specifically, we first introduce a multi-bias module (MBM), which incorporates multi-bias into the ReLU layers. The MBM can decompose the feature maps of input patches into multiple response maps (corresponding to different ground objects) by adopting multi-bias magnitudes. Then, the MBM can adaptively select meaningful maps for accurate classification. Moreover, instead of adopting the traditional convolutional layer, an efficient squeeze convolution module (SCM) is introduced. The SCM can efficiently improve the running speed of the network by reducing the number of parameters, while still maintaining high classification accuracy.

The remainder of this paper is organized as follows. In Section II, we briefly review the CNN model and the CNN-based HSI classification method. The proposed SMBN method is introduced in Section III. The experimental results are given in Section IV. Finally, we summarize this paper and suggest possible future works in Section V.

II. RELATED WORK

In this section, we briefly review the CNN model and traditional CNN-based HSI classification method.

A. CNN

CNN is one of the most popular deep learning models and has been widely used recently in many fields [42], [43]. CNN was inspired by the hierarchical structure of a visual system, and it delivers excellent performance for image classification and object recognition [27], [44]. The architecture of a typical CNN is shown in Fig. 1.

Convolutional layer is an important part of CNN, which convolves the input images or feature maps with multiple

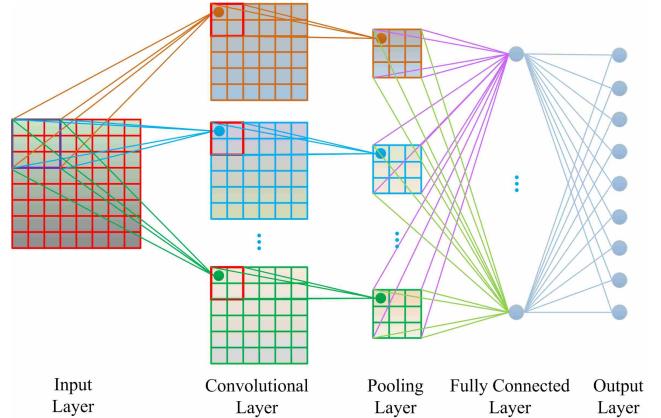


Fig. 1. Architecture of a typical CNN. The purple box in the input layer is a convolution kernel, and the red box in convolutional layer calculates the maximum or average value of this region. For convenience, some layers (e.g., batch-normalization layers and activation layers) are not given.

convolution kernels to produce various feature maps. Mathematically, the convolution operation can be defined as:

$$\mathbf{X}_l = g(\mathbf{X}_{l-1} * \mathbf{W}_l + \mathbf{B}_l), \quad (1)$$

where the \mathbf{X}_{l-1} and \mathbf{X}_l are the input and output of the l -th convolutional layer, and \mathbf{W}_l and \mathbf{B}_l are the convolution kernels and biases of the l -th convolutional layer, respectively. The operator $*$ represents the convolution operation, and $g(\cdot)$ is an activation function.

The activation function has also demonstrated to be very important for deep models, and is usually placed after convolution operation to improve the nonlinearity of the network [45]. The ReLU is found to be particularly effective and widely used in deep neural networks, which can separate noise from informative signals in a feature map obtained from convolutional layer by learning a bias [27], [41]. Mathematically, the ReLU can model a neuron's output g as a function of its input x , $g(x) = \max(0, x)$, which is faster than some saturating nonlinearities (e.g., $g(x) = \tanh(x)$ or $g(x) = (1 + e^{-x})^{-1}$) in terms of training time and gradient descent.

In general, a pooling layer is placed behind the convolutional layer to reduce the size of feature maps from convolutional layer. Then, the size-reduced feature maps are fed into a fully connected layer, where all features are reshaped into one feature vector. In fact, the fully connected layer can be considered as a special case of a convolutional layer, where the size of convolution kernels is the same as the feature maps, and so its mathematical expression is the same as formula (1).

B. CNN-based HSI Classification

In recent years, CNN has also been extended to HSI classification and achieved remarkable performance [46], [47]. In general, typical CNN-based methods extract patch for each pixel of the HSIs and utilize the patches to train the CNN. Specifically, principal component analysis (PCA) [48] is first adopted to reduce the spectral dimension and extract the most informative P components of HSIs [49]. Then, to utilize both spectral and spatial information, for each training pixel, an image patch (of size $S \times S \times P$) is extracted from

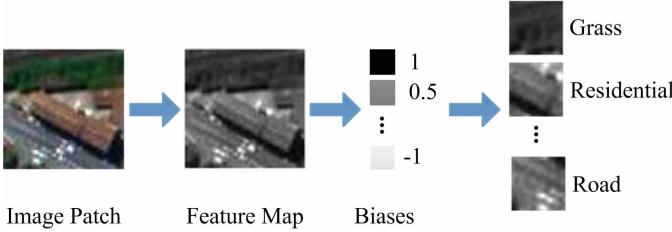


Fig. 2. Example describing how biases can decouple feature maps into different magnitudes of response maps. The grass, residential, and road have different magnitudes in their responses to the brightness and edge differences.

dimension-reduced image as the input of CNN. Assuming the input patches and the corresponding truth label vectors are represented by \mathbf{h}^n ($n = 1, 2, \dots, N$ and N is the number of training pixels) and \mathbf{y}^n , respectively. As described in Section II. A, each input patch \mathbf{h}^n can be transformed into a feature vector \mathbf{v}^n through a series of convolution, pooling, and fully connected operations. Finally, the vector is input to a soft-max layer for classification. The output of the soft-max layer is the probability distribution of each pixel, which is denoted as:

$$\mathbf{p}_c^n = \frac{e^{\mathbf{v}_c^n}}{\sum_{c=1}^C e^{\mathbf{v}_c^n}}, c = 1, 2, \dots, C, \quad (2)$$

where \mathbf{v}_c^n is the c -th value of the input vector \mathbf{v}^n , C is the number of classes, and \mathbf{p}_c^n is the probability of the n -th training pixel belonging to the c -th class. Finally, the predicted label of each pixel can be determined by the maximal probability. Based on the predicted probability, the loss function L_f can be written as:

$$\min L_f = -\sum_{n=1}^N \sum_{c=1}^C \mathbf{y}_c^n \log \mathbf{p}_c^n, \quad (3)$$

where \mathbf{y}_c^n is the c -th value of the truth label vector \mathbf{y}^n .

To minimize the loss function, stochastic gradient descent (SGD) algorithm [50] is widely used to optimize the aforementioned parameters \mathbf{W}_l and \mathbf{B}_l , which propagates the error L_f from the last layer to the first layer, and modifies the parameters at each layer. An optimal CNN model will be obtained, until the loss function convergence is achieved. Then, for an unlabeled test pixel, its possible label can be predicted by the obtained CNN model.

III. PROPOSED SMBN METHOD

For HSI classification, most existing CNN-based methods adopted image patches as inputs to the network. Since HSIs have very complex spatial structures, fixed-size image patches often contain several ground objects, which may mislead the CNN for the classification. The ReLU can be used to separate noise from true signals in the feature maps, but still cannot distinguish different objects. As described in [51], magnitudes of responses can be used to separate different ground objects. Therefore, we incorporate the multiple biases strategy into the ReLU (called as MBM), which can split feature maps of input patches into multiple response maps. An example (patch extracted from the Houston HSI) about how multiple biases can decouple different magnitudes of response feature maps is shown in Fig. 2. We can observe

that there are different brightness and edge contrast on grass, residential, and road. Specifically, the residential class should have higher magnitudes compared to grass and road, since brightness and edge contrast on residential are usually higher than those on grass and road. Thus, we can separate the response maps according to their magnitudes. In addition, to efficiently reduce the parameters and save the running time of network, we incorporate a squeeze strategy into the traditional convolutional layers (called as SCM). The architecture of the proposed SMBN is shown in Fig. 3. As can be seen, MBM and SCM are the two main parts of the proposed SMBN method, which will be detailed as follows.

A. MBM

Fig. 4 illustrates the architecture of MBM. After the convolution operation, we can obtain a set of feature maps $\mathbf{x}_u \in \mathbb{R}^{L_1 L_2}$ ($u = 1, 2, \dots, U$, U is the number of feature maps, L_1 and L_2 are the spatial width and height of the feature map, respectively). Then, by using a multi-bias layer with M biases, each feature map \mathbf{x}_u is separated into M biased maps $\mathbf{x}'_{u,m}$. The element-wise output of the multi-bias layer is defined as:

$$x'_{u,m,i} = g(x_{u,i} + b_{u,m}), \quad (4)$$

where $x_{u,i}$ and $x'_{u,m,i}$ ($i = 1, 2, \dots, L_1 L_2$) are the i -th element in the map \mathbf{x}_u and $\mathbf{x}'_{u,m}$, respectively. $b_{u,m}$ ($m = 1, 2, \dots, M$) is the bias (a scalar) which can be automatically learned through back propagation, and $g(\cdot)$ is a nonlinear function. In this paper, we use ReLU as the nonlinear function, which is described as:

$$x'_{u,m,i} = \begin{cases} x_{u,i} + b_{u,m} & x_{u,i} + b_{u,m} > 0 \\ 0 & x_{u,i} + b_{u,m} \leq 0. \end{cases} \quad (5)$$

Then, these biased maps are concatenated by a concat layer, and the output of the concat layer contains UM concatenated maps $\{\mathbf{x}'_{u,m} | u = 1, 2, \dots, U, m = 1, 2, \dots, M\}$. These maps are then combined into β_t by a convolutional layer, as follows:

$$\beta_t = g(\sum_{u=1}^U \sum_{m=1}^M \mathbf{W}_{u,m,t} * \mathbf{x}'_{u,m} + \mathbf{B}_t), \quad (6)$$

where $t = 1, 2, \dots, T$ and T is the number of convolution kernels in this convolutional layer. $\mathbf{W}_{u,m,t}$ and \mathbf{B}_t are the convolution kernels and biases, respectively.

B. SCM

The network computational complexity is generally proportional to the number of parameters in the network [52]. In other words, small networks can be trained faster because they contain fewer parameters. In this paper, following the work in [53], we replace the traditional convolutional layer with the SCM, which can greatly reduce parameters. Fig. 5 illustrates the architecture of SCM, which is composed of a convolutional layer with $k 1 \times 1$ filters, a convolutional layer with $N_2/2 1 \times 1$ filters and a convolutional layer with $N_2/2 3 \times 3$ filters (N_2 is the number of feature maps outputted by SCM). As can be observed, two strategies are adopted in SCM. First, we replace a part of traditional 3×3 filters

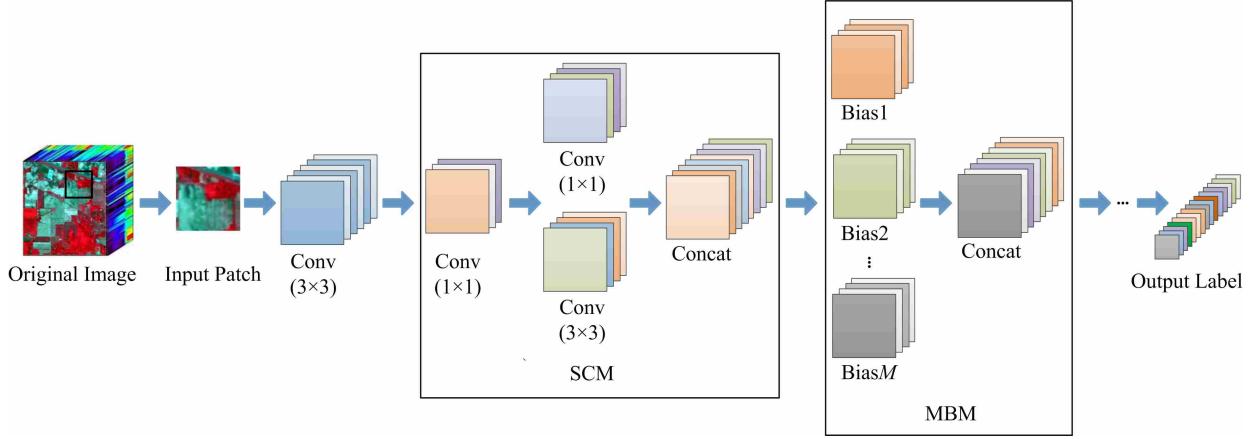


Fig. 3. Brief architecture of the proposed SMBN (some convolutional layers, pooling layers, fully connected layers, and activation layers have been omitted for the sake of simplicity).

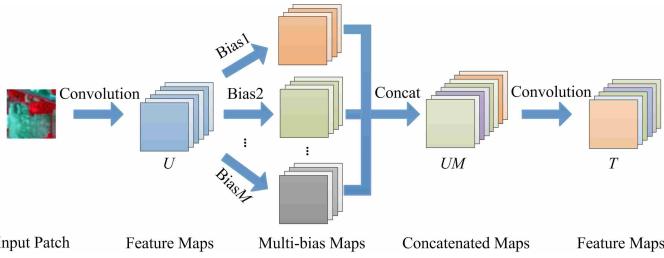


Fig. 4. Architecture of MBM. Adding M biases on U feature maps from convolutional layer to create UM biased maps, and then, these biased maps are concatenated by a concat layer and fed into a convolutional layer.

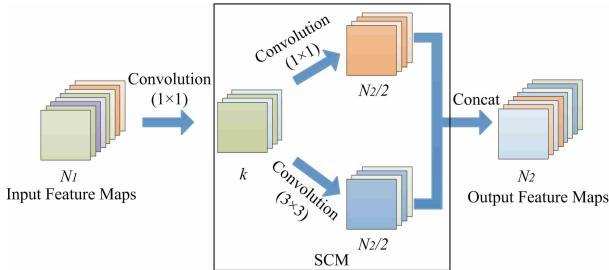


Fig. 5. Organization of convolutional layers in the SCM.

with 1×1 filters, since a 1×1 filter has nine times fewer parameters than a 3×3 filter. Second, we reduce the number of input channels of 3×3 convolution filters and the number of 3×3 convolution filters. Typically, k is set to be much fewer than $N_2/2$, which can reduce the number of input channels for the 3×3 filters. Fig. 6 shows an example which compares the number of parameters between the traditional 3×3 convolutional layer and SCM. We assume the number of input and output feature maps of a 3×3 convolutional layer are 32 and 64, respectively. After calculating the number of parameters, there are 18432 parameters in the traditional 3×3 convolutional layer and 5632 parameters in SCM. Therefore, the SCM has three times fewer parameters than the traditional 3×3 convolutional layer, and thus it can speed up the training and testing process. In addition, the SCM still can maintain high classification accuracy by using 1×1 and 3×3 filters together, which can increase the diversity of feature maps and

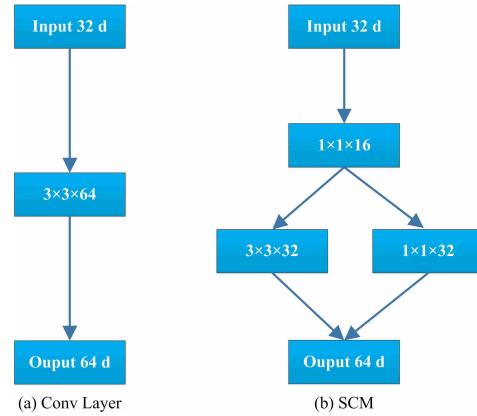


Fig. 6. Example for comparing the number of parameters between traditional 3×3 convolutional layer and SCM. The number of parameters of traditional 3×3 convolutional layer and SCM are 18432 ($32 \times 3 \times 3 \times 64$) and 5632 ($32 \times 16 + 16 \times 3 \times 3 \times 32 + 16 \times 32$), respectively.

enable the network to adaptively select the meaningful maps for accurate classification.

IV. EXPERIMENTAL RESULTS

In this section, in order to verify the effectiveness of the proposed SMBN method, the proposed approach was tested on three widely used HSI images, i.e., the Indian Pines, the Houston University, and the Washington DC. In addition, three quantitative metrics of overall accuracy (OA), average accuracy (AA), and Kappa coefficient are adapted to quantitatively evaluate the classification performance. Specifically, the OA is the total number of the correctly classified samples divided by the total number of test samples. The AA is the mean of the percentage of correctly classified samples for each class. The Kappa is a measure of the degree of agreement. The average value over ten running experiments with randomly selected training samples of each test image is reported for all the indexes. Specifically, 10% samples of all pixels are used as training samples for the Indian Pines image, 50 samples are used as training samples for the Houston University image and Washington DC image. All the experiments were implemented by a PC equipped with a single 1080ti GPU and i7 7700k CPU.

TABLE I
NUMBERS OF SAMPLES IN EACH CLASS ON THE THREE TEST IMAGES.

Indian Pines			Houston University Image			Washington DC Image		
Class	Name	Number	Class	Name	Number	Class	Name	Number
1	Alfalfa	46	1	Healthy grass	1251	1	Roof	3129
2	Corn-no till	1428	2	Stressed grass	1254	2	Road	1790
3	Corn-min till	830	3	Synthetic grass	697	3	Grass	1402
4	Corn	237	4	Trees	1244	4	Trail	1264
5	Grass/Pasture	483	5	Soil	1242	5	Trees	1194
6	Grass/Trees	730	6	Water	325	6	Shadow	1120
7	G/Pasture-m	28	7	Residential	1268			
8	Hay-win	478	8	Commercial	1244			
9	Oats	20	9	Road	1252			
10	Soy-no till	972	10	Highway	1227			
11	Soy-min till	2455	11	Railway	1235			
12	Soy-clean	593	12	Parking Lot1	1233			
13	Wheat	205	13	Parking Lot2	469			
14	Woods	1265	14	Tennis court	428			
15	BGT-Drives	386	15	Running track	660			
16	S-s Towers	93						
	Total	10249		Total	15029		Total	9899

The code of the proposed SMBN method will be released on the website¹.

A. Description of Data Sets

1) *Indian Pines Image*: The Indian Pines image was acquired by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor in northwestern Indiana. The spatial size and spatial resolution of the Indian Pines image are 145×145 and 20m per pixel, respectively. The image consists of 220 spectral channels ranging from 200nm to 240nm. In general, 20 spectral channels have been discarded since they are corrupted by water absorption effects. The original ground truth of this image contains sixteen classes, and the number of samples for each class is reported in Table I. Fig. 10 (a) and (b) show the false-color composite of the Indian Pines image and the corresponding ground truth.

2) *Houston University Image*: The Houston University image was acquired over the Houston University campus and its neighboring region, and was used in the 2013 GRSS Data Fusion Contest. This image contains 144 spectral channels ranging from 380nm to 1050nm, and its spatial size is 349×1905 with the spatial resolution of 2.5m. The ground truth of this image includes fifteen classes, and the number of samples for each class is reported in Table I. The false-color composite of the Houston University image and the corresponding ground truth are shown in Fig. 11 (a) and (b).

3) *Washington DC Image*: The Washington DC image was acquired by the Hyperspectral Digital Image Collection Experiment (HYDICE) sensor over the Washington DC Mall, which contains 210 spectral channels. In this experiment, we only use 191 spectral channels ranging from 400nm to 2400nm and 19 channels ranging from 900nm to 14000nm were abandoned. The spatial size of this image is 280×307 . Its ground truth has six classes, and the number of samples for each class is reported in Table I. Fig. 12 (a) and (b) show the false-color composite of Washington DC image and the corresponding ground truth.

Input	Size: $P \times P$, Channel #: B
Conv Block-1	Kernel Size: 3×3 , Kernel #: 32, Padding: 1, Stride: 1
Max Pool-1	Kernel Size: 2×2 , Stride: 2
Conv Block-2	Kernel Size: 3×3 , Kernel #: 64, Padding: 1, Stride: 1
Max Pool-2	Kernel Size: 2×2 , Stride: 2
Conv Block-3	Kernel Size: 3×3 , Kernel #: 128, Padding: 1, Stride: 1
Max Pool-3	Kernel Size: 2×2 , Stride: 2
Full Connection-1	Kernel Size: 1×1 , Kernel #: 300
Full Connection-2	Kernel Size: 1×1 , Kernel #: C
Softmaxloss	Probability Vector: $[1 \times C]$

Fig. 7. Detailed information of the baseline network. Each conv block contains three convolutional layers, and kernel # represents the number of kernels of three convolutional layers in each conv block. C represents the number of classes.

Input	Size: $P \times P$, Channel #: B
Conv Block	Kernel Size: 3×3 , Kernel #: 32, Padding: 1, Stride: 1
Max Pool-1	Kernel Size: 2×2 , Stride: 2
SM Block-1	Kernel #: (32, 64, 64), Bias #: 4
Max Pool-2	Kernel Size: 2×2 , Stride: 2
SM Block-2	Kernel #: (64, 128, 128), Bias #: 4
Max pool-3	Kernel Size: 2×2 , Stride: 2
Full Connection-1	Kernel Size: 1×1 , Kernel #: 300
Full Connection-2	Kernel Size: 1×1 , Kernel #: C
Softmaxloss	Probability Vector: $[1 \times C]$

Fig. 8. Detailed information of SMBN. SM is the combination of SCM and MBM, and each SM block contains three SMs. bias # represents the number of biases in the MBM.

B. Network Design and Parameters Discussion

Fig. 7 describes the detailed information of our baseline network (it contains traditional 3×3 convolutional layers, and does not adopt MBM and SCM), which includes three convolution blocks, where each block contains three convolutional layers. Each block has [32-32-32], [64-64-64], and [128-128-128] filters, respectively. For all convolutional layers, the kernel size is 3×3 and padded by 1 pixel with stride 1. After each convolution block, it is a max pooling layer with the

¹<https://sites.google.com/site/leyuanfang/home>.

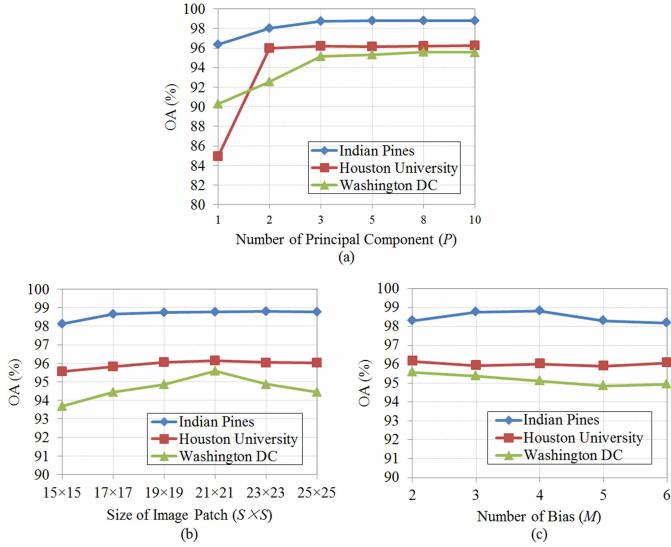


Fig. 9. Classification accuracies (OA) of the proposed SMBN method on the three test images with different parameters: (a) Number of principal component, (b) Size of image patch, and (c) Number of bias.

kernel size of 2×2 and stride 2. The kernel numbers of two fully connected layers are 300 and C (the number of classes), respectively. The final layer is a soft-max layer. Based on this baseline network, the proposed SMBN method introduces MBMs and SCMs to improve the classification performance and the operation speed of the network. The detailed network information of SMBN is described in Fig. 8. All convolutional layers are replaced by SCMs except for the convolution block-1. Since the convolution features are more and more complex and abstract with the increasing of convolutional layers [35], the MBMs are placed behind the convolutional layers of the convolution block-2 and convolution block-3.

Except that weights and biases can be automatically learned by the network during training, there are still other important parameters which need to be determined by our prior knowledge and empirical study, such as the number of principal components (P), the size of image patches ($S \times S$), and the number of biases in the MBM (M).

Fig. 9 (a) shows the effects of principal component number P to the performance (on OA) of the proposed SMBN method on the three test images. It can be seen that as the P increases, the OA first increases, and then, becomes stable. Considering that if P is too small, spectral information is not fully utilized, which will influence on the classification accuracy, and if P is too large, the computing time will increase. Therefore, the P is set to be 5, 5, and 8 for Indian Pines, Houston University, and Washington DC images, respectively.

Fig. 9 (b) shows the effects of image patches size $S \times S$ on the performance (on OA) of the proposed SMBN method over the three test images. As can be seen, with the increasing size of the image patches, the OA first improves, and then, degrades for the three test images. In our experiments, we set $S \times S$ to be 23×23 , 21×21 and 21×21 for the Indian Pines, Houston University, and Washington DC images, respectively. The main reason is that the Indian Pines image

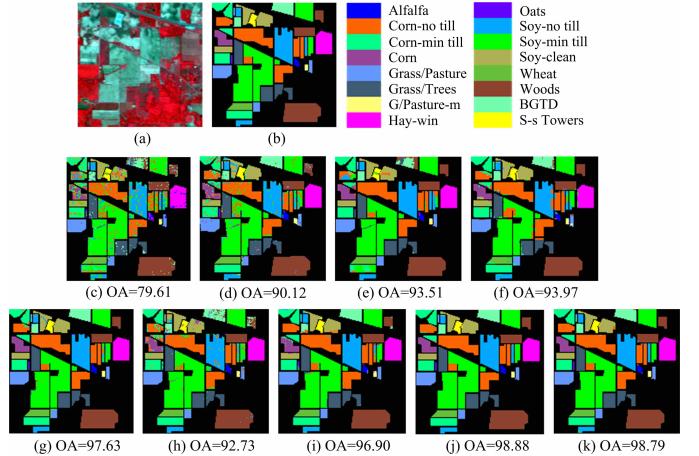


Fig. 10. Reference map and classification results on Indian Pines image. (a) False-color composite image, (b) Reference map, and the classification results (OA in %) obtained by the (c) SVM [54], (d) EMP [12], (e) EPF [55], (f) SRC [18], (g) SC-MK [56], (h) MNFL [57], (i) CNN [46], (j) MBN, and (k) SMBN methods.

has larger smooth regions, whereas the Houston University and Washington DC images have more detailed regions.

Fig. 9 (c) shows the effects of different number of bias M on the performance of the proposed method. As can be observed, when M is set to 4, 2, and 2 for the Indian Pines, Houston University, and Washington DC images, respectively, the OA reaches the maximum. Setting a larger M does not improve the classification performance. Therefore, in our experiments, we set M to 4, 2, and 2 for the Indian Pines, Houston University, and Washington DC images, respectively. The main reason that M in the Indian Pines image is larger than that in the Houston University and Washington DC images is that the objects distribution in Indian Pines is more intensive, and one image patch may include more ground objects.

C. Comparison Results of Different Methods

The classification performance of the proposed SMBN method on the three test images was compared with several well-known classification methods, such as SVM [54], EMP [12], EPF [55], pixel-wise sparse representation classification (SRC) [18], superpixel-based classification via multiple kernels (SC-MK) [56], multiple nonlinear feature learning with multivariate logistic regression (MNFL) [57], the CNN [46], and multi-bias network (MBN). SVM is a spectral-based classifier, which does not consider spatial information. EMP and EPF are spatial-spectral-based classification methods, which exploit the spatial information of HSI by the morphological method and edge-filtering, respectively. SRC is a sparse representation-based method, which adopts the spatial context within fixed-size patches. SC-MK method adopts a superpixel strategy to extract spatial and spectral information. For the MNFL method, multiple features learned from HSI were flexibly combined together for classification. The CNN is a baseline network, which does not contain MBMs and SCMs. The MBN is an improved network, which has introduced the MBMs to optimize CNN. In addition, the parameters of these

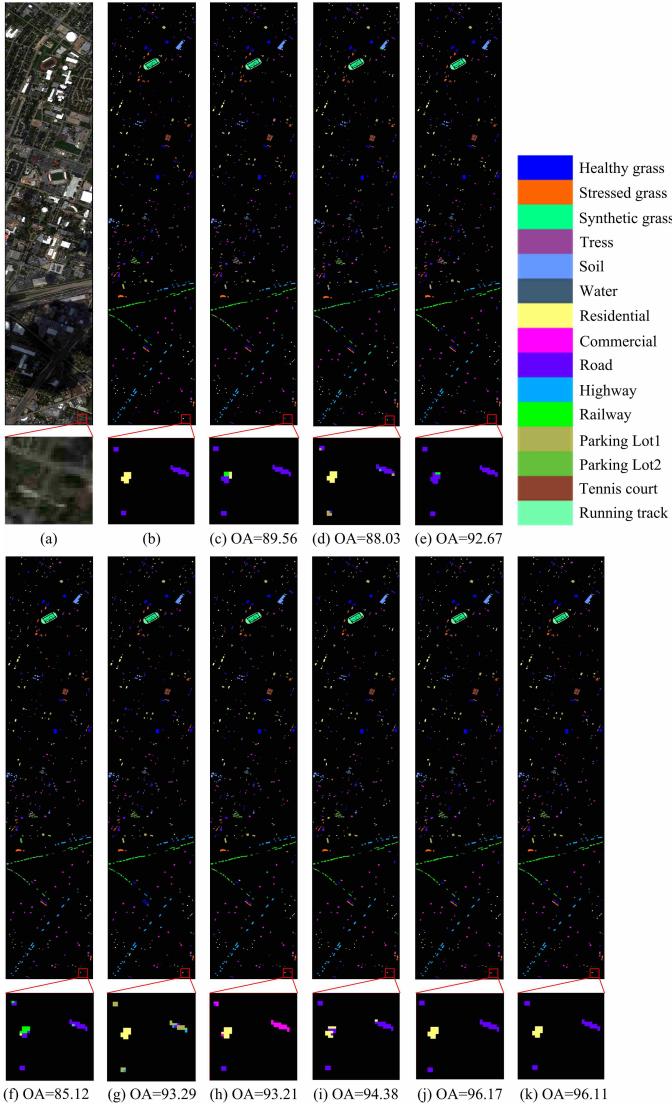


Fig. 11. Reference map and classification results on Houston University image. (a) False-color composite image, (b) Reference map, and the classification results (OA in %) obtained by the (c) SVM [54], (d) EMP [12], (e) EPF [55], (f) SRC [18], (g) SC-MK [56], (h) MNFL [57], (i) CNN [46], (j) MBN, and (k) SMBN methods.

TABLE II

EXECUTION TIME (IN SECOND) OF TRAINING AND TEST PROCESSES ON THE THREE TEST IMAGES USING DIFFERENT METHODS.

Method		Indian Pines	Houston University	Washington DC
CNN	Training	879	772	764
	Test	0.9	1.1	0.9
MBN	Training	1774	1403	1304
	Test	1.5	1.7	1.6
SMBN	Training	931	901	893
	Test	1.0	1.2	1.1

methods are set according to the default values reported in [12], [18], [46], [54]–[57]. Fig. 10 (c)-(k) show classification maps obtained by different methods on the Indian Pines image. It can be observed that the SVM method creates a very noisy estimation in the corresponding classification map, since it only considers spectral information. Some other methods (e.g.,

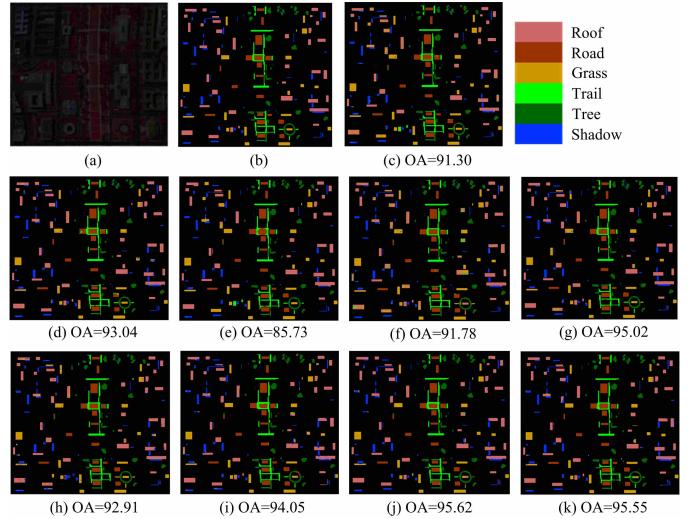


Fig. 12. Reference map and classification results on Washington DC image. (a) False-color composite image, (b) Reference map, and the classification results (OA in %) obtained by the (c) SVM [54], (d) EMP [12], (e) EPF [55], (f) SRC [18], (g) SC-MK [56], (h) MNFL [57], (i) CNN [46], (j) MBN, and (k) SMBN methods.

EMP, EPF, SRC, SC-MK, and MNFL), which incorporate the spatial context of the HSI, deliver a smoother visual result. However, these methods still fail to correctly classify the pixels in some regions (e.g., the upper left region of the Indian Pines image). In the detailed and edge regions (e.g., Soy-no till and Soy-min till in the ten and eleven classes), whose spatial contexts are complex, the MBN and SMBN show better classification performance compared to CNN. This is because they utilize MBM to decouple the feature maps of input patches into multiple response maps, and then select meaningful maps in a flexible way to classify. In addition, as can be seen from Tables II and III, by introducing the SCM, the SMBN can greatly reduce the computational time of the network compared with MBN, while still maintaining high classification accuracy.

Figs. 11 and 12 show different classification maps obtained by different investigated methods on the Houston University and Washington DC images, respectively, and the quantitative results are reported in Tables IV and V. As can be seen, the MBN and SMBN methods, which adopt MBMs to optimize the network, perform better than all other compared methods in terms of OA, AA, and Kappa. In the zoomed regions of Fig. 11, MBN and SMBN provide the best visual classification results. In addition, in some classes, the accuracies of the proposed method are less than those of the other best ones, and the two main reasons are as follows: (1) our method uses fixed-sized image patches as the input to the network, which might not well represent the detailed structural regions; (2) some regions are too scattered and small, and image patches might contain few target ground object but with lots of ground objects from other classes, thus misleading the classification result. In general, our proposed method still deliver the best performance for most classes, OA, AA, and Kappa. Furthermore, as can be seen from Table II, for the three images, SMBN can also greatly save the computing time, and

TABLE III
CLASSIFICATION ACCURACIES (IN %) OF THE INDIAN PINES IMAGE OBTAINED BY DIFFERENT METHODS WITH 10% TRAINING SAMPLES PER CLASS.

Class	SVM [54]	EMP [12]	EPF [55]	SRC [18]	SC-MK [56]	MNFL [57]	CNN [46]	MBN	SMBN
1	54.35	86.94	66.67	87.50	100.0	80.98	87.80	92.74	92.68
2	74.12	83.40	94.35	94.63	97.37	88.26	98.13	99.10	98.83
3	72.95	89.75	97.54	89.22	96.81	89.34	98.53	99.77	99.87
4	64.48	78.29	96.83	89.71	98.69	69.34	91.55	100.0	100.0
5	86.85	89.13	99.31	88.19	97.05	92.51	97.24	97.87	97.47
6	87.11	92.37	92.80	90.29	100.0	98.55	99.54	100.0	100.0
7	84.67	90.43	100.0	79.17	99.78	87.60	80.00	100.0	100.0
8	91.28	99.12	93.68	97.08	98.99	99.58	100.0	100.0	100.0
9	30.56	47.78	33.33	15.88	99.67	66.67	61.11	100.0	100.0
10	74.24	86.76	88.35	86.00	93.87	88.28	94.28	99.51	99.31
11	77.93	94.06	91.33	90.90	98.75	95.59	97.69	99.35	99.45
12	67.51	82.89	93.08	86.86	96.21	90.39	92.50	95.78	95.50
13	93.11	98.43	100.0	93.18	99.57	97.83	97.83	100.0	100.0
14	92.95	98.05	94.99	98.81	99.98	96.55	99.21	99.55	99.65
15	68.22	96.39	96.70	99.10	97.29	86.74	93.08	97.72	97.12
16	99.45	94.94	91.62	82.50	99.04	87.35	92.77	98.56	98.80
OA	79.42	90.77	93.47	93.77	97.98	92.33	97.08	98.87	98.81
AA	76.24	88.05	89.46	85.34	98.21	88.47	92.58	98.75	98.67
Kappa	76.46	89.68	92.53	90.57	97.70	91.24	96.68	99.09	99.06

TABLE IV
CLASSIFICATION ACCURACIES (IN %) OF THE HOUSTON UNIVERSITY IMAGE OBTAINED BY DIFFERENT METHODS WITH 50 TRAINING SAMPLES PER CLASS.

Class	SVM [54]	EMP [12]	EPF [55]	SRC [18]	SC-MK [56]	MNFL [57]	CNN [46]	MBN	SMBN
1	96.43	98.14	98.41	93.76	98.57	95.50	96.25	98.04	97.92
2	96.12	95.32	98.07	98.67	88.62	92.94	96.59	98.76	98.75
3	99.84	94.65	100.0	99.69	100.0	100.0	99.85	99.87	99.85
4	98.29	94.96	97.99	94.72	96.48	90.20	94.39	96.67	96.57
5	93.37	91.66	99.67	96.90	98.79	98.15	99.16	99.69	99.75
6	100.0	99.02	100.0	99.64	99.27	97.45	91.27	100.0	100.0
7	89.44	92.68	93.10	82.84	92.53	86.37	89.57	92.26	92.04
8	78.76	85.84	86.79	76.38	82.88	78.64	80.74	90.27	90.15
9	84.43	74.05	92.55	76.79	88.92	86.02	87.10	92.65	91.85
10	85.00	81.00	95.95	91.67	97.64	97.45	95.07	98.75	98.90
11	83.44	74.69	93.07	70.38	95.61	95.02	95.61	98.10	98.06
12	77.40	73.58	76.57	67.54	92.00	92.14	94.51	97.41	97.30
13	63.22	86.97	44.23	44.63	84.77	82.10	98.09	99.55	99.52
14	97.69	94.21	98.51	99.21	100.0	97.88	100.0	100.0	100.0
15	99.48	99.21	100.0	98.52	99.87	99.84	99.84	100.0	100.0
OA	89.43	88.15	92.83	85.09	93.79	93.00	94.48	96.20	96.15
AA	89.53	89.07	91.73	86.09	94.36	92.65	94.54	97.40	97.37
Kappa	88.27	86.64	92.93	84.54	93.28	91.36	93.34	96.57	96.54

maintain competitive classification accuracy.

D. Effect of Different Number of Training Samples

In practical situations, limited training samples may be insufficient to accurately classify each class, and so it is necessary to investigate the sensitivity of the number of training samples. Fig. 13 illustrates the classification performance of the three test images with different numbers of training samples. As can be seen, for the Indian Pines image, the number of training samples per class is changed from 6% to 20% of all labeled pixels with an interval of 2%, and the number of training samples per class ranges from 30 to 100 with an interval of 10 for Washington DC and Houston images, respectively. As can be seen, SMBN still performs better than other methods, which verifies that the multi-bias strategy can indeed improve the classification performance. The main reasons are that combining the response maps in different magnitude ranges in a flexible manner can achieve better classification performance and sufficient training samples can guarantee the parameters of the network to be finely tuned.

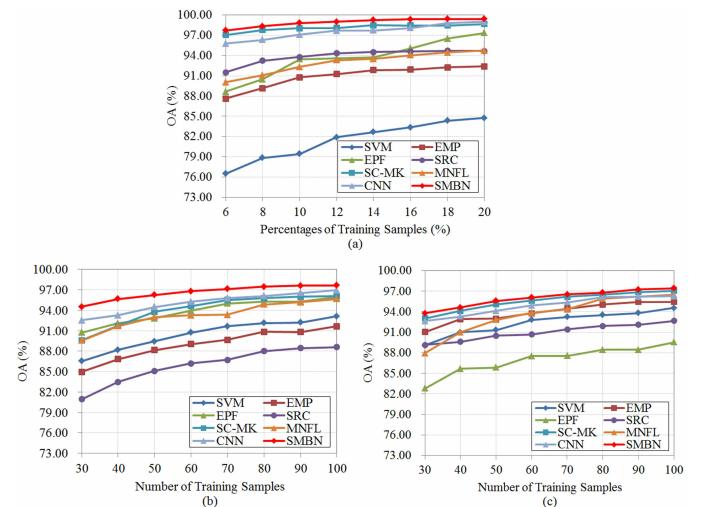


Fig. 13. Effects of different number of training samples on SVM [54], EMP [12], EPF [55], SRC [18], SC-MK [56], MNFL [57], CNN [46] and the proposed SMBN for the (a) Indian Pines, (b) Houston University, and (c) Washington DC.

TABLE V
CLASSIFICATION ACCURACIES (IN %) OF THE WASHINGTON DC IMAGE OBTAINED BY DIFFERENT METHODS WITH 50 TRAINING SAMPLES PER CLASS.

Class	SVM [54]	EMP [12]	EPF [55]	SRC [18]	SC-MK [56]	MNFL [57]	CNN [46]	MBN	SMBN
1	94.04	89.47	91.91	86.52	92.39	87.16	93.18	94.71	94.36
2	97.64	97.30	79.16	98.16	97.87	95.18	96.32	98.21	98.17
3	69.42	83.65	64.54	81.58	95.15	93.19	94.45	93.12	92.97
4	93.69	97.69	92.54	97.52	97.68	97.87	97.60	98.09	98.01
5	97.13	96.65	94.33	95.72	96.85	97.40	95.98	97.63	97.73
6	99.34	96.12	98.46	98.50	96.82	93.94	90.00	92.24	91.84
OA	91.33	92.97	85.85	91.88	95.07	92.80	94.12	95.65	95.58
AA	92.14	93.48	86.82	93.47	95.13	94.13	94.56	95.80	95.75
Kappa	89.73	90.69	81.01	89.92	94.37	91.09	93.35	94.60	94.21

V. CONCLUSIONS

In this paper, we presented a novel SMBN for HSIs classification. Specifically, we place an MBM behind the convolutional layer to decouple feature maps to multiple maps according to the magnitudes of responses, and then, the response maps in different magnitude ranges are combined flexibly by the subsequent layer to achieve better classification performance. In addition, we replace the 3×3 convolutional layer by SCM, which can reduce the executable time of the network and maintain competitive classification accuracy. Experimental results on three real HSIs demonstrate the superiority of the proposed SMBN over traditional CNN-based method and other well-known classifiers.

In the future publications, we will first divide the feature maps from convolutional layers, and feed them into multiple response modules (i.e., multiple nonlinear activation functions). Then, we combine them together as the input of the subsequent convolutional layer, which may further improve the expressive power of the network and obtain more discriminative information hidden in the magnitude of response.

VI. ACKNOWLEDGEMENT

The authors gratefully acknowledge the Associate Editor and the two Anonymous Reviewers for their outstanding comments and suggestions, which greatly helped us to improve the technical quality and presentation of our manuscript.

REFERENCES

- [1] P. Zhong and R. Wang, "Jointly learning the hybrid CRF and MLR model for simultaneous denoising and classification of hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 7, pp. 1319–1334, Jul. 2014.
- [2] L. Fang, C. Wang, S. Li, and J. A. Benediktsson, "Hyperspectral image classification via multiple-feature-based adaptive sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1646–1657, Jul. 2017.
- [3] J. Zhu, L. Fang, and P. Ghamisi, "Deformable convolutional neural networks for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: 10.1109/LGRS.2018.2830403.
- [4] M. Dalponte, H. O. Orka, T. Gobakken, D. Gianelle, and E. Naesset, "Tree species classification in boreal forests with hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2632–2645, May 2013.
- [5] X. Yang and Y. Yu, "Estimating soil salinity under various moisture conditions: An experimental study," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2525–2533, May 2017.
- [6] X. Lu, W. Zhang, and X. Li, "A hybrid sparsity and distance-based discrimination detector for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1704–1717, Dec. 2018.
- [7] D. Chen and C. Hao, "Scale-invariant amplitude spectrum modulation for visual saliency detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 8, pp. 1206–1214, Aug. 2012.
- [8] H. He and Y. Cao, "SSC: a classifier combination method based on signal strength," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1100–1117, Jul. 2012.
- [9] L. Zhang, Y. Zhong, B. Huang, J. Gong, and P. Li, "Dimensionality reduction based on clonal selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4172–4186, Dec. 2007.
- [10] X. Lu, B. Wang, X. Li, and X. Zheng, "Exploring models and data for remote sensing image caption generation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2183–2195, Apr. 2018.
- [11] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.
- [12] J. A. Benediktsson, J. A. Palmason, and J. R. Steinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [13] Y. Gu, C. Wang, D. You, Y. Zhang, S. Wang, and Y. Zhang, "Representative multiple kernel learning for classification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2852–2865, Jul. 2012.
- [14] Y. Gu, J. Chanussot, X. Jia, and J. A. Benediktsson, "Multiple kernel learning for hyperspectral image classification: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6547–6565, Nov. 2017.
- [15] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.
- [16] W. Fu, S. Li, L. Fang, X. Kang, and J. A. Benediktsson, "Hyperspectral image classification via shape-adaptive joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 556–567, Sep. 2016.
- [17] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, Aug. 2015.
- [18] 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.
- [19] L. Fang, H. Zhuo, and S. Li, "Super-resolution of hyperspectral image via superpixel-based sparse representation," *Neuro.*, vol. 273, pp. 171–177, Aug. 2017.
- [20] L. Fang, N. He, S. Li, A. J. Plaza, and J. Plaza, "Extinction profiles fusion for hyperspectral images classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1803–1815, Mar. 2018.
- [21] Y. Wang, L. Lee, B. Xue, L. Wang, M. Song, C. Yu, S. Li, and C. Chang, "A posteriori hyperspectral anomaly detection for unlabeled classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3091–3106, Jun. 2018.
- [22] X. Lu, W. Zhang, and X. Li, "A coarse-to-fine semi-supervised change detection for multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3587–3599, Jun. 2018.
- [23] L. Fang, N. He, S. Li, P. Ghamisi, and J. A. Benediktsson, "A new spatial-spectral feature extraction method for hyperspectral images using local covariance matrix representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3534–3546, Jun. 2018.
- [24] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in

- deep convolutional networks for visual recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, May 2015.
- [26] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 3104–3112.
- [27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1106–1114.
- [28] X. Lu, X. Zheng, and Y. Yuan, “Remote sensing scene classification by unsupervised representation learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5148–5157, Sep. 2017.
- [29] 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. 2017.
- [30] X. Ma, H. Wang, and J. Geng, “Spectral-spatial classification of hyperspectral image based on deep auto-encoder,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4073–4085, Sep. 2016.
- [31] Y. Liu, G. Cao, Q. Sun, and M. Siegel, “Hyperspectral classification via deep networks and superpixel segmentation,” *Int. J. Remote Sens.*, vol. 36, no. 13, pp. 3459–3482, Jul. 2015.
- [32] C. Tao, H. Pan, Y. Li, and Z. Zou, “Unsupervised spectral-spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification,” *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2438–2442, Dec. 2015.
- [33] Y. Chen, X. Zhao, and X. 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, Jul. 2015.
- [34] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, “Deep convolutional neural networks for hyperspectral image classification,” *J. Sensors*, vol. 2015, no. 2, pp. 1–12, Jul. 2015.
- [35] 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.
- [36] B. Liu, X. Yu, P. Zhang, X. Tan, A. Yu, and Z. Xue, “A semi-supervised convolutional neural network for hyperspectral image classification,” *Remote Sens. Lett.*, vol. 8, no. 9, pp. 839–848, Sep. 2017.
- [37] H. Zhang, Y. Li, Y. Zhang, and Q. Shen, “Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network,” *Remote Sens. Lett.*, vol. 8, no. 5, pp. 438–447, May 2017.
- [38] 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.
- [39] W. Song, S. Li, L. Fang, and T. Lu, “Hyperspectral image classification with deep feature fusion network,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3173–3184, Jun. 2018.
- [40] X. Lu, Y. Chen, and X. Li, “Hierarchical recurrent neural hashing for image retrieval with hierarchical convolutional features,” *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 106–120, Jan. 2018.
- [41] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.
- [42] A. Stuhlsatz, J. Lippel, and T. Zielke, “Feature extraction with deep neural networks by a generalized discriminant analysis,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 4, pp. 596–608, Apr. 2012.
- [43] L. Shao, D. Wu, and X. Li, “Learning deep and wide: a spectral method for learning deep networks,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2303–2308, Dec. 2014.
- [44] M. Schwarz, H. Schulz, and S. Behnke, “RGB-D object recognition and pose estimation based on pre-trained convolutional neural network features,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2015, pp. 1329–1335.
- [45] J. Bruna and S. Mallat, “Invariant scattering convolution networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013.
- [46] 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, Jul. 2016.
- [47] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, “Multisource remote sensing data classification based on convolutional neural network,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 937–949, Feb. 2018.
- [48] K. W. Jorgensen and L. K. Hansen, “Model selection for gaussian kernel PCA denoising,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 1, pp. 163–168, Jan. 2012.
- [49] G. Licciardi, P. R. Marpu, J. Chanussot, and J. A. Benediktsson, “Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles,” *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 447–451, May 2012.
- [50] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Dec. 1998.
- [51] H. Li, W. Ouyang, and X. Wang, “Multi-bias non-linear activation in deep neural networks,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2016.
- [52] F. N. Iandola, M. W. Moskewicz, K. Ashraf, and K. Keutzer, “FireCaffe: Near-linear acceleration of deep neural network training on compute clusters,” in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 2592–2600.
- [53] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, E. Dumitru, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 1–9.
- [54] 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.
- [55] X. Kang, S. Li, and J. A. Benediktsson, “Spectral-spatial hyperspectral image classification with edge-preserving filtering,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.
- [56] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, “Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.
- [57] J. Li, X. Huang, P. Gamba, J. M. Bioucas-Dias, L. Zhang, J. A. Benediktsson, and A. Plaza, “Multiple feature learning for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015.

Leyuan Fang (S’10-M’14-SM’17) 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 Postdoctoral Research Fellow with the Department of Biomedical Engineering, Duke University, Durham, NC, USA. Since Jan. 2017, he has been an associate professor with the College of Electrical and Information Engineering, Hunan University. His research interests include sparse representation and multi-resolution analysis in remote sensing and medical image processing. He has won the Scholarship Award for Excellent Doctoral Student granted by Chinese Ministry of Education in 2011.

Guangyun Liu received the B.S. degree from Jiangxi University of Science and Technology, ganzhou, China, in 2015. He is currently working toward the M.S. degree in the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Hunan University, Changsha, China.



His research interest is remote sensing image classification.



Shutao Li (M'07-SM'15) received the B.S., M.S., and Ph.D. degrees from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively. He was a Research Associate with the Department of Computer Science, the Hong Kong University of Science and Technology, Hong Kong, in 2011. From 2002 to 2003, he was a Post-Doctoral Fellow with the Royal Holloway College, University of London, London, U.K., with Prof. John Shawe-Taylor. In 2005, he visited the Department of Computer Science, Hong Kong University of Science and

Technology as a Visiting Professor. He joined the College of Electrical and Information Engineering, Hunan University, in 2001. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He has authored or co-authored over 160 refereed papers. His current research interests include compressive sensing, sparse representation, image processing, and pattern recognition.

He is now an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, and a member of the Editorial Board of the Information Fusion and the Sensing and Imaging. He was a recipient of two 2nd-Grade National Awards at the Science and Technology Progress of China in 2004 and 2006.



Jón Atli Benediktsson (S'84-M'90-SM'99-F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984 and the M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

From 2009 to 2015, he was the Pro Rector of Science and Academic Affairs and a Professor of electrical and computer engineering with the University of Iceland. On July 1, 2015, he became the Rector of the University of Iceland. He has published

extensively in his fields of interest. His research interests include remote sensing, image analysis, pattern recognition, biomedical analysis of signals, and signal processing.

Dr. Benediktsson is a member of the Association of Chartered Engineers at Iceland (VFI), the Societas Scinetiarum Islandica, and the Tau Beta Pi. He is a Fellow of the International Society for Optics and Photonics (SPIE). He was a recipient of the Stevan J. Kristof Award from Purdue University in 1991 as an Outstanding Graduate Student in Remote Sensing. He was also a recipient of the Icelandic Research Council's Outstanding Young Researcher Award in 1997, the IEEE Third Millennium Medal in 2000, the Yearly Research Award from the Engineering Research Institute at the University of Iceland in 2006, the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society in 2007, and the IEEE/VFI Electrical Engineer of the Year Award in 2013. He was a co-recipient of the University of Iceland's Technology Innovation Award in 2004, the 2012 IEEE Transactions on Geoscience and Remote Sensing Paper Award, the IEEE GRSS Highest Impact Paper Award in 2013, and the International Journal of Image and Data Fusion Best Paper Award in 2014. He was the 2011-2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS) and has been on the GRSS Administrative Committee since 2000. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008 and has been an Associate Editor of TGRS since 1999, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003, and the IEEE ACCESS since 2013. He was the Chairman of the Steering Committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2007 to 2010. He serves on the Editorial Board of the IEEE PROCEEDINGS and the International Editorial Board of the International Journal of Image and Data Fusion. He is the Co-Founder of a biomedical startup company Oxymap.



Pedram Ghamsi (S'12-M'15-SM'18) received the B.Sc. degree in civil (survey) engineering from the Tehran South Campus of Azad University, Tehran, Iran, the M.Sc. degree (Hons.) in remote sensing from the K. N. Toosi University of Technology, Tehran, in 2012, and the Ph.D. degree in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2015.

From 2013 to 2014, he spent seven months with the School of Geography, Planning and Environmental Management, The University of Queensland, Brisbane, QLD, Australia. He was a Post-Doctoral Research Fellow with the University of Iceland. He has been a Post-Doctoral Research Fellow with the Technical University of Munich, Munich, Germany, and Heidelberg University, Heidelberg, Germany, since 2015. He has also been a Researcher with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany, since 2015. His research interests include remote sensing and image analysis, with a special focus on spectral and spatial techniques for hyperspectral image classification, multisensor data fusion, machine learning, and deep learning. Dr. Ghamsi received the prestigious Alexander von Humboldt Fellowship in 2015. In the 2010C2011 academic year, he received the Best Researcher Award for M.Sc. students in the K. N. Toosi University of Technology. In 2013, he presented at the IEEE International Geoscience and Remote Sensing Symposium, Melbourne, VIC, Australia, and received the IEEE Mikio Takagi Prize for winning the conference Student Paper Competition against almost 70 people. In 2016, he was selected as a talented international researcher by the Irans National Elites Foundation. In 2017, he won the Data Fusion Contest 2017 organized by the Image Analysis and Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society. His model was the most accurate among over 800 submissions.