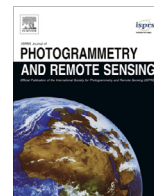




Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: [www.elsevier.com/locate/isprsjprs](http://www.elsevier.com/locate/isprsjprs)

# MugNet: Deep learning for hyperspectral image classification using limited samples

Bin Pan, Zhenwei Shi\*, Xia Xu

Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China  
 State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China  
 Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China

## ARTICLE INFO

### Article history:

Received 9 May 2017

Received in revised form 26 October 2017

Accepted 3 November 2017

Available online xxxx

### Keywords:

Hyperspectral image classification

Deep learning

Multi-grained scanning

MugNet

## ABSTRACT

In recent years, deep learning based methods have attracted broad attention in the field of hyperspectral image classification. However, due to the massive parameters and the complex network structure, deep learning methods may not perform well when only few training samples are available. In this paper, we propose a small-scale data based method, multi-grained network (MugNet), to explore the application of deep learning approaches in hyperspectral image classification. MugNet could be considered as a simplified deep learning model which mainly targets at limited samples based hyperspectral image classification. Three novel strategies are proposed to construct MugNet. First, the spectral relationship among different bands, as well as the spatial correlation within neighboring pixels, are both utilized via a multi-grained scanning approach. The proposed multi-grained scanning strategy could not only extract the joint spectral-spatial information, but also combine different grains' spectral and spatial relationship. Second, because there are abundant unlabeled pixels available in hyperspectral images, we take full advantage of these samples, and adopt a semi-supervised manner in the process of generating convolution kernels. At last, the MugNet is built upon the basis of a very simple network which does not include many hyperparameters for tuning. The performance of MugNet is evaluated on a popular and two challenging data sets, and comparison experiments with several state-of-the-art hyperspectral image classification methods are revealed.

© 2017 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

## 1. Introduction

Hyperspectral images (HSIs) usually have very high spectral resolution, where hundreds of continuous channels can be observed. Based on such abundant spectral information, hyperspectral techniques have been widely used in many fields such as environmental monitoring (Pan et al., 2017a) and materials analysis (Lu et al., 2013; Xu and Shi, 2017). Land cover classification in HSI data is a hot topic, whereas it is still challenging (Khodadadzadeh et al., 2014; Xu et al., 2016; He et al., 2017; Wang et al., 2017). One of the major reasons is that labeled samples which could be used for training are not enough. How to achieve satisfying performance with limited training samples has attracted many researches.

A feasible idea to handle this problem is combining various features or classifiers. Gu et al. proposed a series of multiple kernel learning based HSI classification methods, which have presented the capability of representing different data characteristics (Gu et al., 2012, 2015, 2016). In Li et al. (2013), Li et al. constructed a family of generalized composite kernels by utilizing the spectral and spatial information from HSI data. Multiply feature fusion is also a popular approach (Li et al., 2015). In Benediktsson et al. (2005), extended morphological profiles were proposed to fuse the spectral and spatial features. Huang and Zhang combined spectral, structural and semantic features to improve the classification accuracy (Huang and Zhang, 2013). Zhong et al. compared the performance of hashing based multiple features fusion methods for HSI classification (Zhong et al., 2016). Ensemble learning methods have presented better generalization capability via integrating many individual learners (Zhou, 2012). In Ham et al. (2005) a typical ensemble method, random forest, was investigated for HSI classification. Then improved works based on the random forest were reported (Chan and Paelinckx, 2008; Rodriguez-Galiano

\* Corresponding author at: Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China.

E-mail addresses: [panbin@buaa.edu.cn](mailto:panbin@buaa.edu.cn) (B. Pan), [shizhenwei@buaa.edu.cn](mailto:shizhenwei@buaa.edu.cn) (Z. Shi), [xuxia@buaa.edu.cn](mailto:xuxia@buaa.edu.cn) (X. Xu).

et al., 2012). In Mountrakis et al. (2011), Mountrakis et al. reported the application of ensemble support vector machine (SVM) for HSI classification. Based on SVM several ensemble methods were developed (Pal and Foody, 2010; Santos et al., 2013; Xia et al., 2016a,b). Another idea for small-scale data based classification is semisupervised learning. Although labeled samples in HSIs are usually not rich, there are many unlabeled pixels which could provide extra data distribution information. Using very limited labeled samples and large number of unlabeled pixels, semisupervised methods have presented good performance (Tuia and Camps-Valls, 2009; Yang et al., 2014; Wang et al., 2015; Huo et al., 2015; Wan et al., 2015; Romaszewski et al., 2016; Ma et al., 2016).

Recently, deep learning based methods have shown promising performance in HSI classification. In Chen et al. (2014) Chen et al. explored the application of deep learning methods in HSI classification for the first time. He proposed a hybrid framework of principle component analysis, stacked autoencoders and logistic regression. Inspired by this work some variations were developed, such as convolutional autoencoder (Zhao et al., 2015) and stacked denoising autoencoders (Liu et al., 2015). Another typical network, deep belief network (DBN), were also used for HSI classification. Chen et al. added spatial correlation to the DBN in Chen et al. (2015). Zhong et al. improved the performance of DBN by a diversity promoting prior over latent factors (Zhong et al., 2017). Convolutional neural network (CNN) is one of the most popular deep learning methods, and many studies based on CNN are reported during recent years (Hu et al., 2015; Chen et al., 2016; Li et al., 2017). In order to reduce the computational complexity, Pan et al. used joint spectral-spatial features to simplify the deep learning models (Pan et al., 2016; Pan et al., 2017b). Some applications of deep learning methods on remote sensing can be also observed in Han et al. (2015), Yao et al. (2016), Zhang et al. (2016), and Cheng et al. (2016, 2017).

Generally, the feature representation ability of deep learning models is strongly dependent on large number of training samples. For example, in Chen et al. (2014, 2015) and Pan et al. (2016), about half of all the labeled pixels are used for training. However, manual annotation for hyperspectral data is difficult, which results in the lack of enough labeled pixels. Although deep learning methods sound promising for HSI classification, more studies are necessary in small-scale data condition.

In this paper, we propose a new deep learning based HSI classification method, multi-grained network (MugNet), where only limited training samples are required. The motivations of MugNet derive from three aspects:

First and foremost, since hyperspectral data are composed of continuous spectral vectors, exploring the relationship among different bands may improve the performance of classification methods. Moreover, the spatial relationship among neighboring pixels is proven useful for HSI classification. Multi-grained scanning methods are effective in representing sequential or spatial relationships (Zhou and Feng). Inspired by this recognition, we integrate a new multi-grained scanning strategy with the procedure of feature extraction, so as to generate a more powerful network.

Second, although labeled samples in HSI data are limited, there are plenty of unlabeled pixels. These unlabeled pixels could provide extra data distribution information which may contribute to the final classification. By constructing the classification network in a semi-supervised approach, its representative ability would be improved.

At last, a more immediate idea of reducing the number of required training samples is adopting a simpler network structure. Traditional deep learning models such as DBN and CNN usually have a mass of parameters for optimization, which means sufficient training samples are needed. In Chan et al. (2015) a simplified deep learning method, principal component analysis network

(PCANet), was proposed. PCANet can serve as an effective baseline where more advanced processing components or more sophisticated architectures could be justified.

Based on the above motivations, the solution comes as no surprise. We use PCANet as the foundation, where multi-grain and semi-supervised information are integrated. In addition, research has shown that image filtering is a powerful preprocessing method for HSI classification (Li and Du, 2014; Kang et al., 2014a; Xia et al., 2016b; Pan et al., 2017b,c). Therefore, we use rolling guidance filtering (Zhang et al., 2014) for preprocessing. Similar strategy is also conducted in Pan et al. (2017b,c).

The major contributions of MugNet can be summarized as follows:

- A deep learning based HSI classification method is developed, aiming at the small-scale data classification. Comparison experiments with some state-of-the-art methods have verified the effectiveness of MugNet.
- To enhance the classification performance, three novel strategies are proposed in MugNet, i.e., multi-grain scanning, semi-supervised learning and simple basic framework.

In the next section we give detailed description of the MugNet based HSI classification method. Experimental results and discussion are displayed in Section 3. We conclude this paper in Section 4.

## 2. Methodology

Inspired by the good performance of edge-preserving filtering for preprocessing (Xia et al., 2016b; Pan et al., 2017b,c), we first conduct rolling guidance filtering on the original HSI data, and the filtering results are used as the inputs of the spectral and spatial networks. Rolling guidance filtering will not change the dimensionality of spectral vectors, and it mainly aims at removing small details and noise from the HSI data (Xia et al., 2016b; Pan et al., 2017b). Note that rolling guidance filtering is only a small trick, and we do not pay special attention to it here.

MugNet includes two parallel branches: spectral MugNet and spatial MugNet. Both of the two networks are constructed based on a semi-supervised PCANet,  $S^2$ PCANet. In this section, we give detailed presentations about the three components of the proposed method:  $S^2$ PCANet, spectral MugNet and spatial MugNet. Finally, an overall description about the algorithm flow is provided.

### 2.1. $S^2$ PCANet

PCANet is a cascade linear network which could be regarded as a simple deep learning method. Here, we first give a brief introduction to the network structure of PCANet, and then extend it to a semi-supervised version,  $S^2$ PCANet. PCANet contains four layers: input layer, two convolution layers and output layer.

#### 2.1.1. Input layer

Suppose  $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\}$  denote all the  $N$  training samples.  $\mathbf{X}_i \in \mathbb{R}^{m \times n}$  is the  $i$ th sample in  $\mathbf{X}$  with size  $m \times n$ . For example, if we use spectral vectors directly as samples, then  $m = 1$  and  $n$  is equal to the number of bands. In PCANet,  $\mathbf{X}$  are labeled.

#### 2.1.2. Convolution layer

For a given  $\mathbf{X}_i$ , collect all the patches (assuming size  $k_1 \times k_2$ ) around each pixel, then vectorize them and remove patch mean. The patch size  $k_1 \times k_2$  is a user-defined parameter. Still take the spectral vectors for example,  $k_1 \times k_2$  could be set as  $3 \times 1, 5 \times 1$  or others. In this way  $\mathbf{X}_i$  can be expressed by

$$\hat{\mathbf{X}}_i = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t, \dots, \mathbf{x}_{mn}] \in \mathbb{R}^{k_1 k_2 \times mn}. \quad (1)$$

Since there are  $N$  training samples available, totally  $N \times mn$  vectorized patches are available, i.e.,

$$\hat{\mathbf{X}} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t, \dots, \mathbf{x}_{N \times mn}] \in \mathbb{R}^{k_1 k_2 \times Nmn}. \quad (2)$$

Assuming that the number of filters in this layer is  $L$ , the principal components of  $\hat{\mathbf{X}}$  can be obtained by

$$\min_{\mathbf{V} \in \mathbb{R}^{k_1 k_2 \times L}} \|\hat{\mathbf{X}} - \mathbf{V}\mathbf{V}^T \hat{\mathbf{X}}\|_F^2, \quad \text{s.t. } \mathbf{V}^T \mathbf{V} = \mathbf{I}_L, \quad (3)$$

where  $\mathbf{I}_L$  is an identity matrix with size  $L \times L$ , and  $\mathbf{V}$  is the principal eigenvectors of  $\hat{\mathbf{X}}\hat{\mathbf{X}}^T$ . Reshape each column of  $\mathbf{V}$  to a  $k_1 \times k_2$  matrix, i.e.,

$$\hat{\mathbf{V}} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_t, \dots, \mathbf{v}_L\}, \quad \mathbf{v}_t \in \mathbb{R}^{k_1 \times k_2}, \quad (4)$$

and  $\hat{\mathbf{V}}$  are considered as the convolution kernels in this layer. Let  $L_1$  and  $L_2$  denote the number of filters in the first and second convolution layers respectively. Then we can find  $L_1 \times N$  outputs in the first layer. These outputs are used as the inputs of the second layer, and the same process as Eqs. (1)–(4) is conducted in the second layer. Totally  $L_1 L_2 \times N$  outputs are observed in this layer. Theoretically, more layers are also acceptable in PCANet. However, as suggested in Chan et al. (2015), two convolution layers are enough in most cases. Note that there is a patch-mean removal process between the two convolution layers so as to avoid linear cascade.

### 2.1.3. Output layer

In this layer, binary hashing and histogram feature are used for the final feature representation. According to Eqs. (1)–(4), there are  $L_1$  convolved images for a sample  $\mathbf{X}_i$  in the first convolution layer, each of which will generate  $L_2$  outputs in the second convolution layer. For a sample  $\mathbf{X}_i$ , let  $\mathbf{O}_i^j (j = 1, 2, \dots, L_1)$  denote a group of the outputs in the first convolution layer, then we convert all the  $L_2$  outputs in  $\mathbf{O}_i^j$  to a single image:

$$\mathbf{T}_i^j = \sum_{t=1}^{L_2} 2^{t-1} H(\mathbf{O}_i^j), \quad (5)$$

where  $H(\cdot)$  is a Heaviside step function with 1 for positive inputs and 0 otherwise, and  $\mathbf{T}_i^j$  ranges in  $[0, 2^{L_2} - 1]$ . Then  $\mathbf{T}_i^j$  is divided into  $B$  blocks with size  $B_s$ . The histogram features are extracted from these blocks. The features of all the blocks in  $\mathbf{T}_i^j$  are combined and denoted by  $\text{Bhist}(\mathbf{T}_i^j)$ . The final feature expression for  $\mathbf{X}_i$  is calculated by

$$\mathbf{f}_i = [\text{Bhist}(\mathbf{T}_i^1), \text{Bhist}(\mathbf{T}_i^2), \dots, \text{Bhist}(\mathbf{T}_i^{L_1})]. \quad (6)$$

### 2.1.4. Semi-supervised

In  $S^2$ PCANet, we use both labeled and unlabeled samples to generate the convolution kernels. Besides labeled pixels, most HSI data sets also include many unlabeled ones. For example, in Indian Pines data set, 10,249 pixels are labeled, while there are 10,776 unlabeled samples. These unlabeled samples can also provide data distribution information. Let  $\mathbf{Y} = [\mathbf{Y}^l, \mathbf{Y}^u]$  denote all labeled ( $\mathbf{Y}^l$ ) and unlabeled ( $\mathbf{Y}^u$ ) pixels. In  $S^2$ PCANet, we define

$$\hat{\mathbf{Y}} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t, \dots, \mathbf{y}_{(N_l+N_u) \times mn}] \in \mathbb{R}^{k_1 k_2 \times (N_l+N_u)mn}, \quad (7)$$

where  $N_l$  and  $N_u$  are the number of labeled and unlabeled pixels, respectively. Note that although the labels of test samples are not available in the training process, they cannot be considered as “unlabeled pixels”. Test sample do not overlap unlabeled pixels. Then the convolution kernels in the first layer could be obtained by

$$\min_{\mathbf{V} \in \mathbb{R}^{k_1 k_2 \times L_1}} \|\hat{\mathbf{Y}} - \mathbf{V}\mathbf{V}^T \hat{\mathbf{Y}}\|_F^2, \quad \text{s.t. } \mathbf{V}^T \mathbf{V} = \mathbf{I}_{L_1}. \quad (8)$$

Kernels determined by Eq. (8) are used in  $S^2$ PCANet. Similar process is also conducted in the second convolution layer.

Instead of giving pseudo-labels to the unlabeled pixels, in  $S^2$ PCANet we use all the unlabeled pixels to learn more representative convolution kernels, and only use labeled pixels for SVM training. Similar ideas could also be observed in Tuia and Camps-Valls (2009) and Huo et al. (2015).

### 2.2. Spectral MugNet

For a given spectral vector, the values presented in different channels have some correlation. In other words, we can regard hyperspectral vectors as special sequence data. In spectral MugNet, we investigate the data correlation by adopting 1-Dimensional (1-D) convolution for single spectral vectors based on  $S^2$ PCANet. However, it is almost impossible to determine which grains of convolution is the most appropriate. Therefore, a multi-grained scanning based method is proposed for spectral vectors.

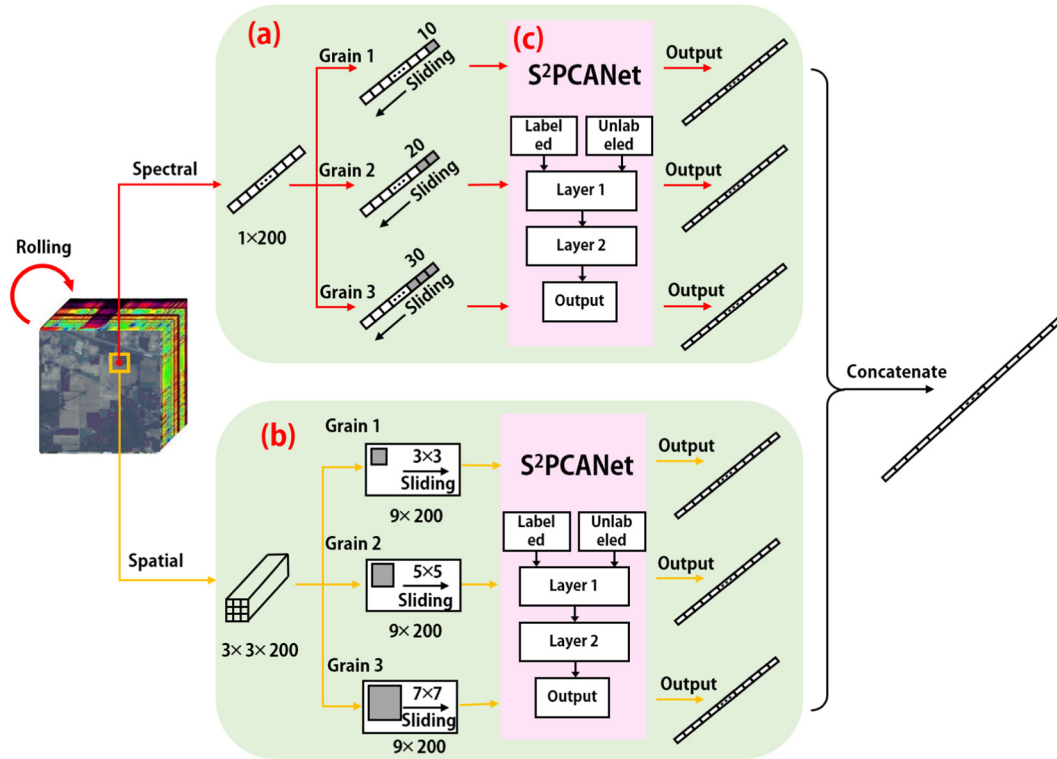
Figs. 1(a) and 2 provide the illustrations for spectral MugNet. Suppose there are 100 training and 1000 testing pixels with 200 bands, the size of sliding window is  $10 \times 1$ , and the step length is 1. Note that training samples could be labeled or unlabeled. For each training sample, we can get 191 patches. After scanning all the training samples, totally 19,100 patches are obtained. These patches are considered as the  $\hat{\mathbf{Y}}$  in Eq. (7). Because spectral MugNet is on the basis of  $S^2$ PCANet, we calculate the convolution kernels by Eqs. (3) and (4). Then these kernels are used in convolution layers, and the network structure in this grain is determined. Finally, the transformed features are obtained by Eqs. (5) and (6), and a linear support vector machine (SVM) (Fan et al., 2008) is employed as the classifier. In the test stage, all the 1000 pixels are directly put into the network with the learned convolution kernels, which leads to 1000 new feature representation for these pixels.

A natural doubt is that it is almost impossible to judge which size of sliding window can produce the best performance. In spectral MugNet, we use multi-grained scanning (Zhou and Feng) to combine several different grains. Then, we fuse all the extracted features in different grains using a concatenating strategy. By selecting multiple sizes of sliding windows, the final features will include more spectral relationship information.

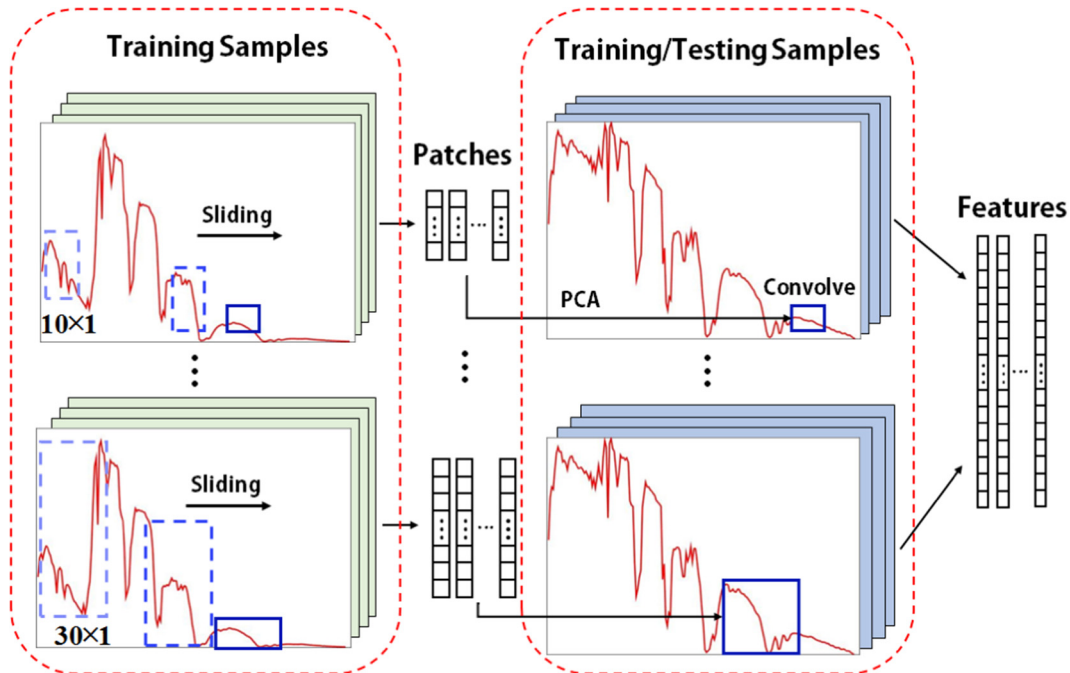
### 2.3. Spatial MugNet

As proved in many works (Kang et al., 2014a; Ghamisi et al., 2015; Kang et al., 2015; Xia et al., 2016b), properly exploiting the spatial correlation among neighboring pixels will lead to more accurate classification maps. Based on this recognition, besides spectral MugNet, we design the spatial MugNet to further utilize the spatial relationships in hyperspectral images.

Spatial MugNet is also on the basis of  $S^2$ PCANet. An illustration for spatial MugNet is shown in Figs. 1(b) and 3. Follow the same assumption as spectral MugNet (100/1000 samples for training/testing, 200 bands, 1 step length). For each training sample, we collect its  $3 \times 3$  neighboring pixel vectors, and flatten them into a matrix with size  $9 \times 200$ . This matrix is considered as an input samples in spatial MugNet. Subsequently, we segment it to many patches with a certain size, e.g.,  $3 \times 3$ . Totally  $1386 \times 100$  patches can be obtained for all the training samples. According to Eq. (7), we can get the  $\hat{\mathbf{Y}}$  by transforming these patches to a  $9 \times 138,600$  matrix. Finally,  $\hat{\mathbf{Y}}$  is used to learn the filters by Eq. (8) and extract the high level features by Eqs. (5) and (6). Similar process is also



**Fig. 1.** The flowchart of MugNet. We take Indian Pines data set for example. The number of channels in this data set is 200, and we set the neighboring size as  $3 \times 3$ . (a) and (b) are spectral and spatial MugNets, respectively. (c) is the semi-supervised PCANet.



**Fig. 2.** An illustration for spectral MugNet.

conducted in the test stage, and all the 1000 test samples are expressed by more representative features.

Similar to spectral MugNet, we extend the spatial network to a multi-grained version. Several sizes of sliding windows are selected to construct different networks which could extract repre-

sentative features from raw HSI data. At last, vector stacking strategy is used to generate the final feature expression for all the samples. According to the above discussion, we can conclude that spatial MugNet mainly targets at exploring the relationship between neighboring pixels in nearby bands.



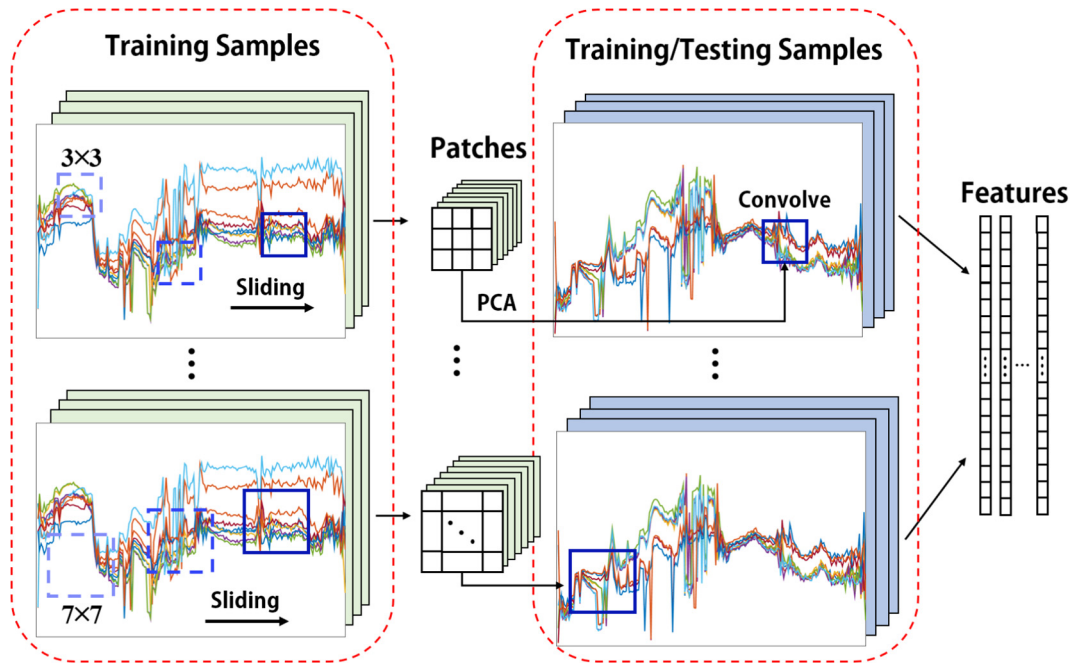


Fig. 3. An illustration for spatial MugNet.

#### 2.4. Classification with MugNet

The integrated MugNet based HSI classification can be obtained by combining the spectral and spatial parts. A flowchart of MugNet is displayed in Fig. 1. First, the same rolling guidance filtering as (Pan et al., 2017b) is used on the original HSI data. Then, two parallel branches are observed, i.e., spectral and spatial MugNets. The inputs of spectral MugNet are raw spectral vectors, while those of spatial MugNet are data cubes surrounding training/testing samples. After the process by the networks, features extracted in different grains are combined via a vector stacking strategy. Finally, a simple linear SVM is learned and utilized for classification. The regularization parameter in SVM is determined by cross validation, which is set as 10.

### 3. Experiments and discussion

In this section, we discuss the performance of MugNet in three public data sets. Spectral and spatial MugNet are evaluated in Sections 3.2 and 3.3. Three popular metrics, overall accuracy (OA), average accuracy (AA) and kappa coefficient ( $\kappa$ ) are used for evaluation. The results of MugNet are compared with those of 5 recently proposed methods (Li and Du, 2014; Kang et al., 2014a, b; Pan et al., 2017b,c). We have run all the methods for 50 times and the mean values as well as the standard deviations are reported. Finally, the statistical evaluation is implemented to further verify the effectiveness of MugNet.

Since MugNet is developed under the basis of PCANet (Chan et al., 2015), we use the default values to define the shared parameters between MugNet and PCANet, i.e.,  $L_1 = L_2 = 8$ ,  $B_s = 7 \times 7$ . The unique parameters in MugNet mainly include spectral grains, spatial grains and the number of combined models. However, because MugNet is an ensemble based method, setting different parameters can just generate various individual learners. Therefore, analysis about these parameters is embedded in Sections 3.2 and 3.3.

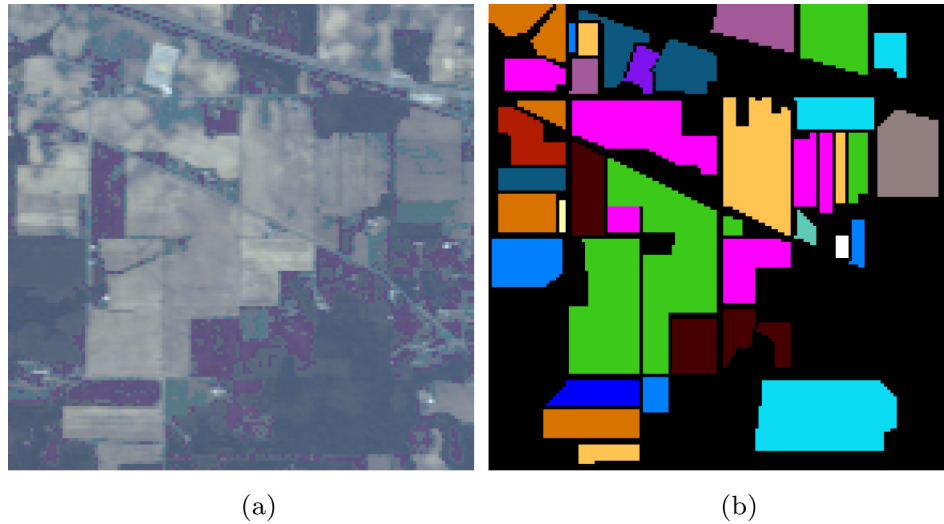
#### 3.1. Data sets

Three hyperspectral data sets are used to evaluate the performance of all the methods: Indian Pines,<sup>1</sup> (IEEE GRSS Data Fusion Contest, 2013; Liao et al., 2015; Debes et al., 2014) and grss\_dfc\_2014 (IEEE GRSS Data Fusion Contest, 2014; Liao et al., 2015)<sup>2</sup>:

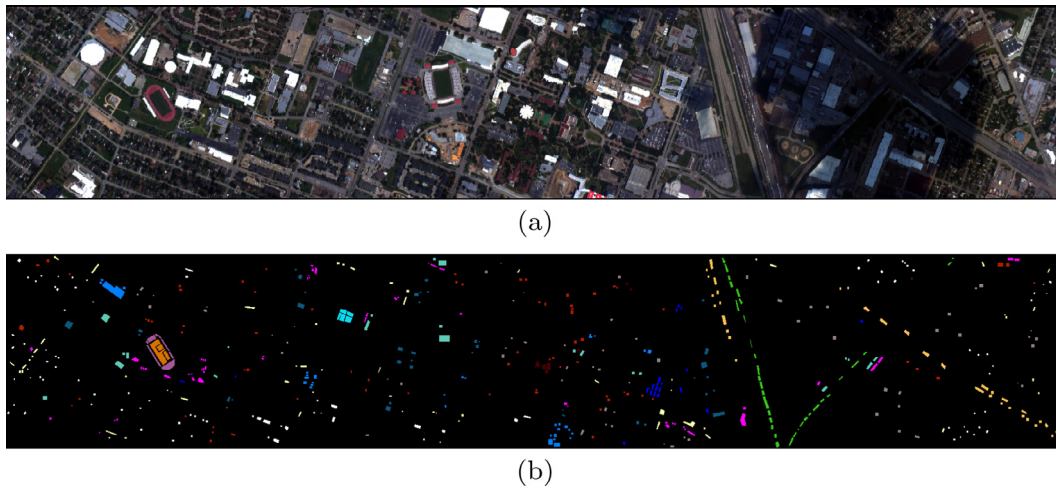
- Indian Pines is a very popular hyperspectral data set which has been widely used in many classification works. This data set was acquired by airborne visible/infrared imaging spectrometer (AVIRIS) in Northwestern Indiana, with 20 m spatial resolution and 0.4–2.5  $\mu\text{m}$  wavelength range. It covers  $145 \times 145$  pixels, where 10,249 pixels are labeled, and the rests are not. The labeled pixels are classified into 16 different classes. 200 spectral bands are observed after removing the water absorption bands. Fig. 4(a) and (b) are false color composite image and the corresponding ground truth for this data set.
- Grss\_dfc\_2013 is a public hyperspectral data set used in 2013 IEEE GRSS Data Fusion Contest. It is acquired over the University of Houston campus and the neighboring urban area in 2012, with  $349 \times 1905$  pixels size and 2.5 m spatial resolution. This hyperspectral imagery has 144 spectral bands between 0.4 and 1.0  $\mu\text{m}$  region. All the 15,029 labeled pixels are classified into 15 classes. In Fig. 5 a false color composite image and the ground truth map are displayed.
- Grss\_dfc\_2014 is a more challenging hyperspectral data set which is used in the 2014 IEEE GRSS Data Fusion Contest. It is acquired from long-wave infrared (LWIR, thermal infrared) bands between 7.8 and 11.5  $\mu\text{m}$  wavelengths. It covers an urban area near Thetford Mines in Québec, Canada, with 84 channels and 1 m spatial resolution. The size of this data set is

<sup>1</sup> Available online: [http://www.ehu.es/ccwintco/index.php?title=Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes), grss\_dfc\_2013.

<sup>2</sup> Available online: <http://www.grss-ieee.org/community/technical-committees/data-fusion/>.



**Fig. 4.** Indian Pines data set. (a) A false composite image with R-G-B = bands 36-17-11. (b) The ground truth map. Each color corresponds to a certain class. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 5.** Grss\_dfc\_2013 data set. (a) A false composite image with R-G-B = bands 70-30-12. (b) The ground truth map. Each color corresponds to a certain class. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

795 × 564 pixels with 22,532 pixels labeled. A ground truth with 7 land cover classes is provided. Since this data set is collected from LWIR bands, its quality is relatively lower. Therefore, this data set is more challenging for classification. A false color composite image and the corresponding ground truth are shown in Fig. 6(a) and (b).

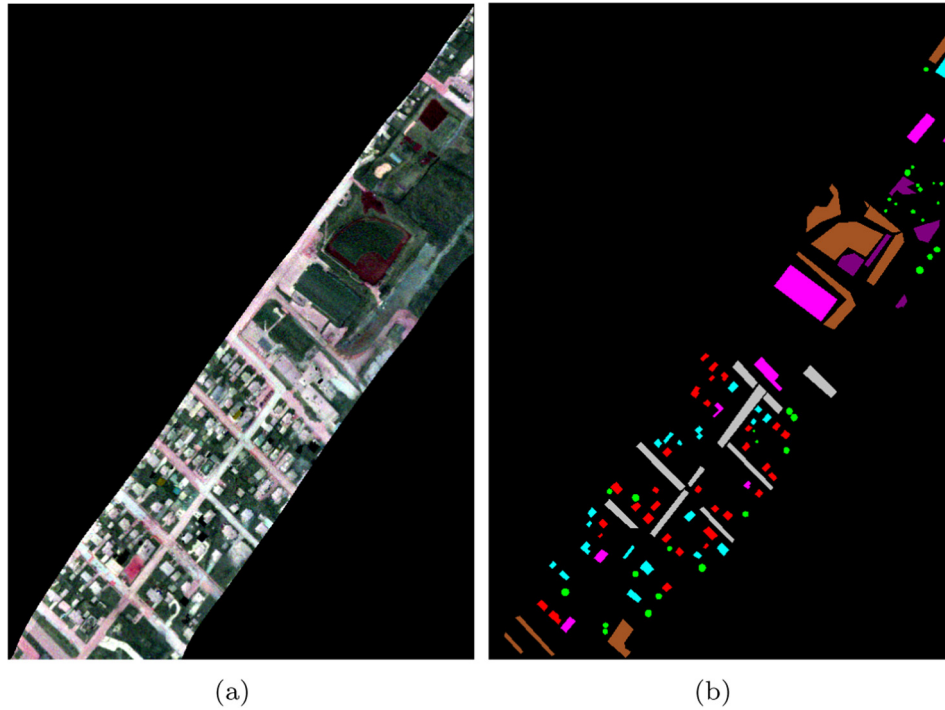
For Indian Pines/grss\_dfc\_2013/grss\_dfc\_2014 data sets we randomly select 20/20/50 samples per class for training, and the rests for testing. Unlabeled pixels used in MugNet refer to the black regions in Figs. 4(b), 5(b) and 6(b), and they do not overlap the test samples. Some classes such as “Grass-pasture-mowed” and “Oats” only have nearby 20 labeled pixels. For these classes we use half for training and the rest for testing. All the methods are conducted 50 times with different training samples, and the average accuracies are reported. Specially, in spatial MugNet, we utilize a patch around each training sample for feature extraction. This operation may lead to overlap between training and testing sets. To avoid this problem, we remove the pixels in patches from the testing set. Note that using neighborhood information is not equal to increasing training samples number, because the neighboring sam-

ples are not labeled. This is also a widely used strategy in HSI classification. For spectral MugNet, this phenomenon does not exist.

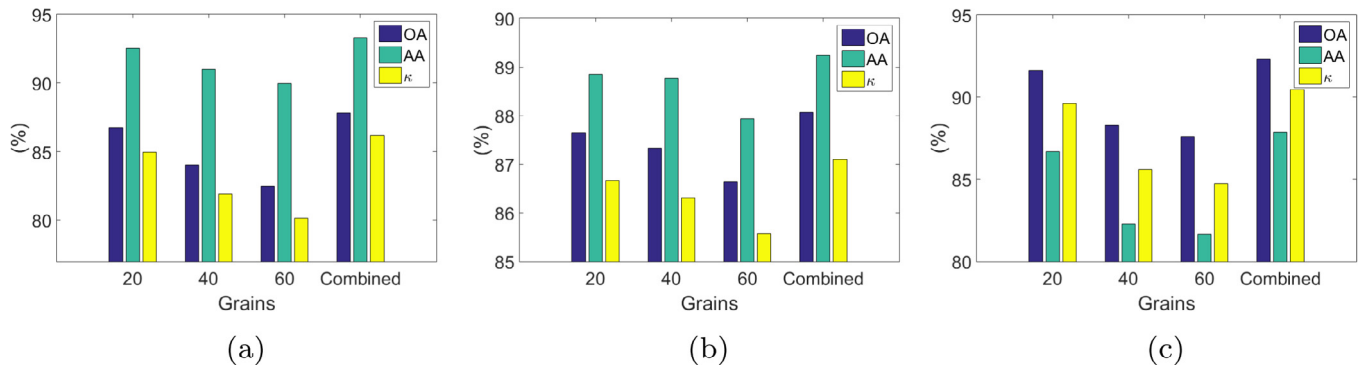
### 3.2. Spectral classification

Spectral MugNet applies only spectral information for classification. Here, we mainly focus on effect of single grain in order to compare with the combined grains. We set the grains as  $20 \times 1$ ,  $40 \times 1$  and  $60 \times 1$ , respectively. Then all the features extracted in different grains are combined via a vector stacking strategy.

Fig. 7 shows the results by spectral MugNet on the three data sets. For Indian Pines data set, it is observed that smaller grains could lead to better performance, but the gaps are not obvious. By combining all the grains, spectral MugNet outperforms single grain based networks, as shown in the 4th column at Fig. 7(a). Similar phenomenon can also be observed from the results on grss\_dfc\_2013 and grss\_dfc\_2014 data sets. The OA/AA/ $\kappa$  of spectral MugNet in Indian Pines, grss\_dfc\_2013 and grss\_dfc\_2014 data sets are 87.83%/93.32%/86.17%, 88.06%/89.25%/87.10% and 92.37%/87.86%/90.47%, respectively. Note that it is not appropriate to combine too many grains. In each grain, we have to train a cer-



**Fig. 6.** Grss\_dfc\_2014 data set. (a) A false composite image with R-G-B = bands 30-45-66. (b) The ground truth map. Each color corresponds to a certain class. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 7.** Classification accuracies by spectral MugNet on (a) Indian Pines, (b) grss\_dfc\_2013 and (c) grss\_dfc\_2014 data sets.

tain network. If too many grains were used in spectral MugNet, the computational cost and memory usage will be unacceptable. To balance this conflict, we only use 3 grains for ensemble.

### 3.3. Spatial classification

The effectiveness of spatial correlation is evaluated in this section. Because spatial MugNet is constructed using training samples and their neighborhoods, to guarantee the objectivity of the results, the neighboring pixels around training samples are not used for testing. Here, we set the grains in spatial MugNet as  $3 \times 3$ ,  $5 \times 5$  and  $7 \times 7$ . Their respective and combined accuracies are all reported.

The classification results by spatial MugNet are revealed in Fig. 8. We can see that the accuracies decrease in larger grains. This is mainly because the input data of spatial MugNet are small images with size  $9 \times 200$ , in this case large grains such as  $7 \times 7$  tend to represent more global information. However, according to this experiments, local information is more beneficial. Com-

pared with single grain network, the ensemble networks present slightly better performance, as shown in the 4th columns at Fig. 8(a)–(c). The OA/AA/ $\kappa$  of spatial MugNet in Indian Pines, grss\_dfc\_2013 and grss\_dfc\_2014 data sets are 89.86%/94.29%/88.78%, 89.49%/90.55%/88.47% and 92.19%/89.29%/90.59%, respectively. Overall, spectral and spatial MugNets have reported similar classification accuracies.

### 3.4. Joint spectral and spatial features for classification

In this section, we combine the spectral and spatial features to construct the final classification model, MugNet. The performance of MugNet is compared with those of some recently proposed methods which are based on nearest regularized subspace (NRS) (Li and Du, 2014), edge-preserving filtering (EPF) (Kang et al., 2014a), image fusion and recursive filtering (IFRF) (Kang et al., 2014b), simplified deep learning (RVCA Net) (Pan et al., 2017b) and ensemble learning (HiFi) (Pan et al., 2017c). To validate the effectiveness of the semi-supervised strategy, supervised MugNet

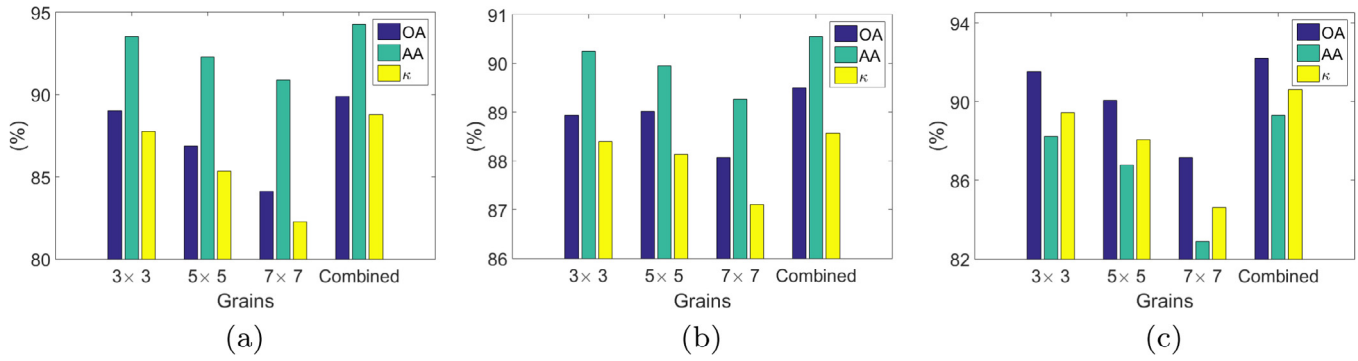


Fig. 8. Classification accuracies by spatial MugNet on (a) Indian Pines, (b) grss\_dfc\_2013 and (c) grss\_dfc\_2014 data sets.

**Table 1**  
Classification accuracies of different methods on indian pines dataset.

Class	Samples	Methods						
		Train/Test	IFRF	EPF	NRS	RVCANet	HiFi	MugNet-S
OA(%)		88.85 ± 2.51	77.54 ± 3.10	79.86 ± 1.82	83.06 ± 2.32	89.06 ± 1.70	88.42 ± 1.87	<b>90.65 ± 1.77</b>
AA(%)		94.03 ± 1.25	87.04 ± 1.88	88.74 ± 1.03	91.56 ± 0.89	94.11 ± 0.77	93.94 ± 0.63	<b>95.48 ± 0.78</b>
$\kappa \times 100$		87.35 ± 2.81	74.66 ± 3.41	77.14 ± 2.02	80.87 ± 2.56	87.51 ± 1.90	86.87 ± 2.08	<b>89.36 ± 1.99</b>
Alfalfa	20/26	99.69 ± 1.05	98.84 ± 1.78	99.92 ± 0.55	99.00 ± 1.70	99.46 ± 1.35	98.46 ± 1.92	100.0 ± 0.00
Corn-notill	20/1408	85.68 ± 5.96	56.53 ± 11.1	70.28 ± 7.42	63.94 ± 6.85	81.91 ± 5.58	78.35 ± 6.06	78.34 ± 5.46
Corn-mintill	20/810	90.10 ± 6.43	67.27 ± 10.5	75.47 ± 7.71	79.91 ± 7.05	91.49 ± 4.52	89.81 ± 2.78	89.88 ± 6.82
Corn	20/217	97.54 ± 4.43	96.56 ± 4.60	94.58 ± 4.59	98.59 ± 2.22	96.78 ± 3.84	97.93 ± 2.53	99.09 ± 1.88
Grass-pasture	20/463	92.59 ± 4.73	91.09 ± 4.56	87.69 ± 5.18	93.60 ± 3.02	90.06 ± 3.88	93.67 ± 3.79	93.38 ± 3.83
Grass-trees	20/710	97.07 ± 4.42	96.97 ± 3.93	92.55 ± 4.19	98.36 ± 1.10	97.92 ± 1.80	98.74 ± 1.54	99.16 ± 1.43
Grass-pasture-mowed	14/14	100.0 ± 0.00	96.85 ± 3.58	100.0 ± 0.00	99.42 ± 1.95	96.75 ± 5.54	95.42 ± 3.49	99.85 ± 1.03
Hay-windrowed	20/458	99.65 ± 0.84	96.65 ± 5.47	98.20 ± 2.30	98.76 ± 0.63	99.39 ± 0.92	99.75 ± 0.45	99.94 ± 0.10
Oats	10/10	100.0 ± 0.00	99.80 ± 1.41	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00
Soybean-notill	20/952	87.98 ± 6.18	83.09 ± 7.85	72.45 ± 8.16	87.43 ± 3.79	88.16 ± 6.63	86.99 ± 4.44	90.20 ± 4.02
Soybean-mintill	20/2435	80.94 ± 7.82	69.55 ± 9.23	71.75 ± 8.38	72.01 ± 6.49	79.82 ± 5.86	79.89 ± 6.38	85.17 ± 5.10
Soybean-clean	20/573	87.75 ± 4.81	73.26 ± 10.1	81.24 ± 7.00	90.49 ± 4.08	93.31 ± 3.18	93.43 ± 3.76	95.01 ± 3.59
Wheat	20/185	99.36 ± 0.41	99.39 ± 0.32	99.18 ± 1.28	99.49 ± 0.31	99.41 ± 0.29	99.43 ± 0.29	99.89 ± 0.21
Woods	20/1245	92.24 ± 6.02	88.51 ± 7.76	87.09 ± 5.11	94.24 ± 3.49	96.96 ± 2.79	95.69 ± 4.99	98.78 ± 2.19
Buildings-Grass-Trees-Drives	20/366	95.37 ± 4.42	81.44 ± 10.6	90.43 ± 6.09	90.65 ± 4.05	95.23 ± 2.72	96.71 ± 2.59	99.47 ± 1.28
Stone-Steel-Towers	20/73	98.46 ± 3.39	96.93 ± 5.68	98.96 ± 2.02	99.06 ± 1.90	99.07 ± 0.65	98.79 ± 1.07	99.57 ± 0.85

**Table 2**  
Classification accuracies of different methods on grss\_dfc\_2013 dataset.

Class	Samples	Methods						
	Train/Test	IFRF	EPF	NRS	RVCANet	HiFi	MugNet-S	MugNet
OA(%)		76.86 ± 1.38	87.01 ± 2.10	84.15 ± 1.57	87.51 ± 1.02	86.31 ± 1.29	89.73 ± 0.76	<b>90.82 ± 1.00</b>
AA(%)		78.86 ± 1.26	86.83 ± 2.23	85.88 ± 1.38	89.20 ± 1.01	87.56 ± 1.10	91.12 ± 0.68	<b>91.91 ± 0.93</b>
$\kappa \times 100$		75.00 ± 1.49	85.95 ± 2.27	82.87 ± 1.70	86.50 ± 1.10	85.20 ± 1.39	88.90 ± 0.83	<b>90.08 ± 1.08</b>
Healthy grass	20/1231	72.30 ± 8.27	94.15 ± 4.78	87.96 ± 5.15	91.62 ± 3.80	93.01 ± 4.03	90.39 ± 3.36	91.26 ± 4.36
Stressed grass	20/1234	61.18 ± 5.76	93.80 ± 4.28	86.52 ± 6.73	91.66 ± 3.54	96.11 ± 3.22	94.49 ± 1.94	91.58 ± 3.98
Synthetic grass	20/677	93.31 ± 4.90	98.91 ± 2.10	95.68 ± 3.42	99.99 ± 0.02	99.68 ± 0.12	99.74 ± 0.10	100.0 ± 0.00
Trees	20/1224	58.42 ± 6.38	95.82 ± 5.15	86.48 ± 6.03	91.07 ± 3.19	94.76 ± 3.64	92.24 ± 2.54	93.28 ± 2.93
Soil	20/1222	88.21 ± 4.52	94.19 ± 3.38	94.73 ± 3.69	94.88 ± 3.27	98.83 ± 1.11	99.06 ± 0.61	93.89 ± 3.73
Water	20/305	74.18 ± 9.15	93.96 ± 7.16	91.24 ± 6.09	98.57 ± 2.60	95.87 ± 3.39	97.45 ± 0.83	97.31 ± 2.97
Residential	20/1248	74.19 ± 5.68	84.74 ± 7.04	81.65 ± 5.51	82.22 ± 5.13	89.74 ± 4.58	86.90 ± 3.13	89.15 ± 4.54
Commercial	20/1224	78.55 ± 8.62	84.81 ± 6.08	73.04 ± 7.29	93.19 ± 4.14	58.78 ± 7.12	72.44 ± 7.58	94.80 ± 4.21
Road	20/1232	63.50 ± 7.67	78.93 ± 7.51	92.48 ± 3.30	78.83 ± 4.17	80.33 ± 3.83	83.05 ± 4.40	85.54 ± 4.91
Highway	20/1207	83.79 ± 5.48	87.05 ± 6.13	72.43 ± 9.02	80.99 ± 4.87	82.21 ± 9.61	92.15 ± 6.44	86.54 ± 5.94
Railway	20/1215	87.98 ± 6.50	83.05 ± 5.57	72.98 ± 8.65	80.33 ± 6.13	76.45 ± 7.60	88.63 ± 4.96	91.14 ± 4.44
Parking Lot 1	20/1213	87.35 ± 4.45	72.89 ± 7.75	72.31 ± 8.14	81.76 ± 7.19	76.69 ± 6.57	82.44 ± 5.57	86.51 ± 7.19
Parking Lot 2	20/449	75.86 ± 10.8	49.94 ± 16.2	85.12 ± 5.79	80.32 ± 8.21	73.63 ± 4.71	88.89 ± 2.89	85.98 ± 7.75
Tennis Court	20/408	92.69 ± 6.73	90.29 ± 8.89	95.86 ± 3.34	99.12 ± 1.75	97.81 ± 1.44	98.89 ± 0.60	97.21 ± 3.47
Running Track	20/640	91.33 ± 5.72	99.88 ± 0.37	99.81 ± 0.60	93.43 ± 2.58	99.54 ± 0.53	99.98 ± 0.05	94.50 ± 2.68

(MugNet-S) using labeled samples only is also compared. The 50-times average accuracies and the corresponding standard deviations are reported in Tables 1–3, and the classification maps by different methods are shown in Figs. 9–11.

### 3.4.1. Results on Indian Pines data set

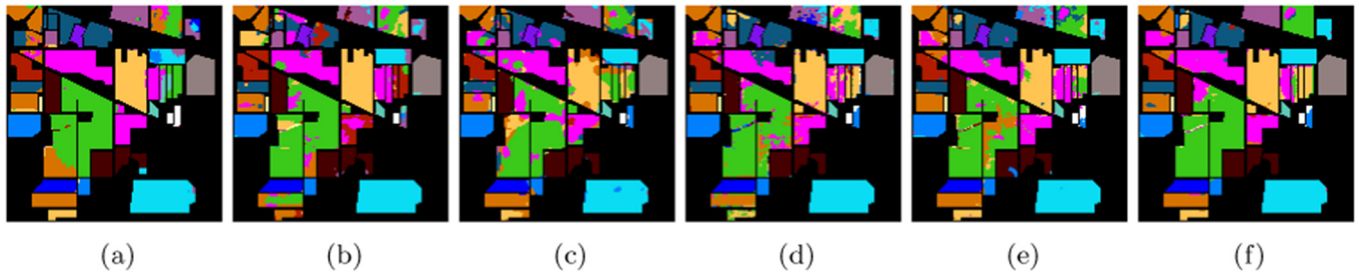
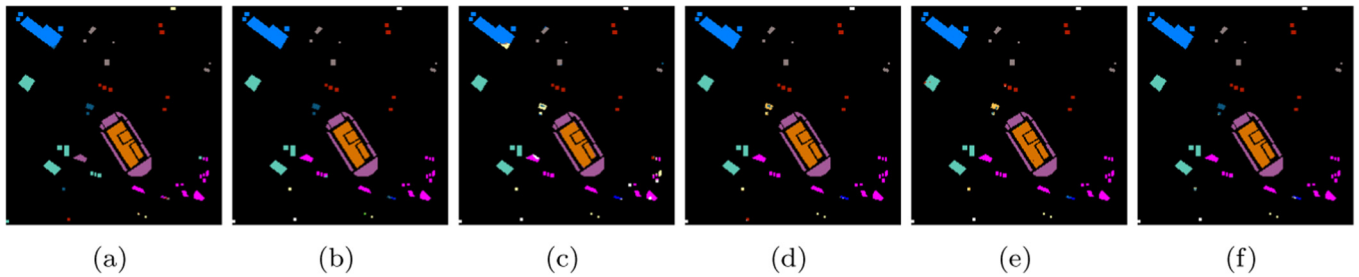
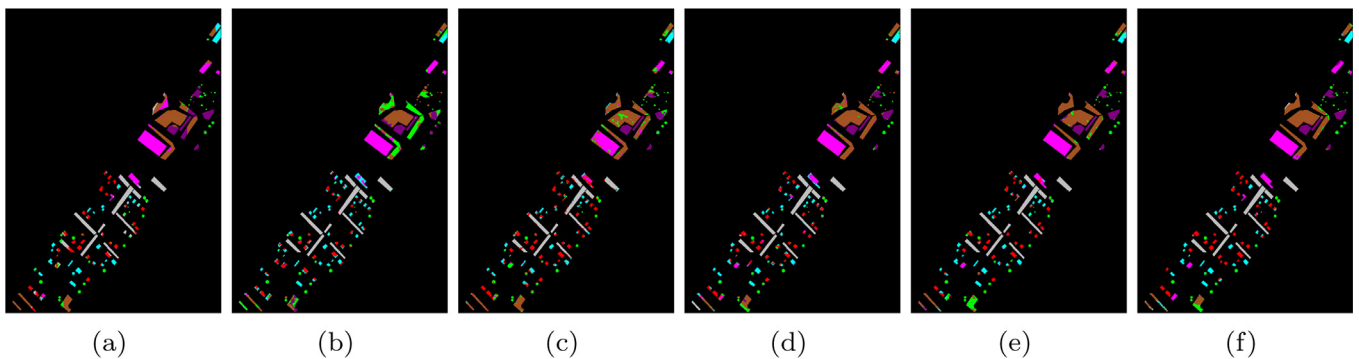
In this data set, IFRF, HiFi and MugNet perform better than the others, among which MugNet achieves about 1–5% advantage. A notable detail is that samples in this data set are quite unbalanced,



**Table 3**

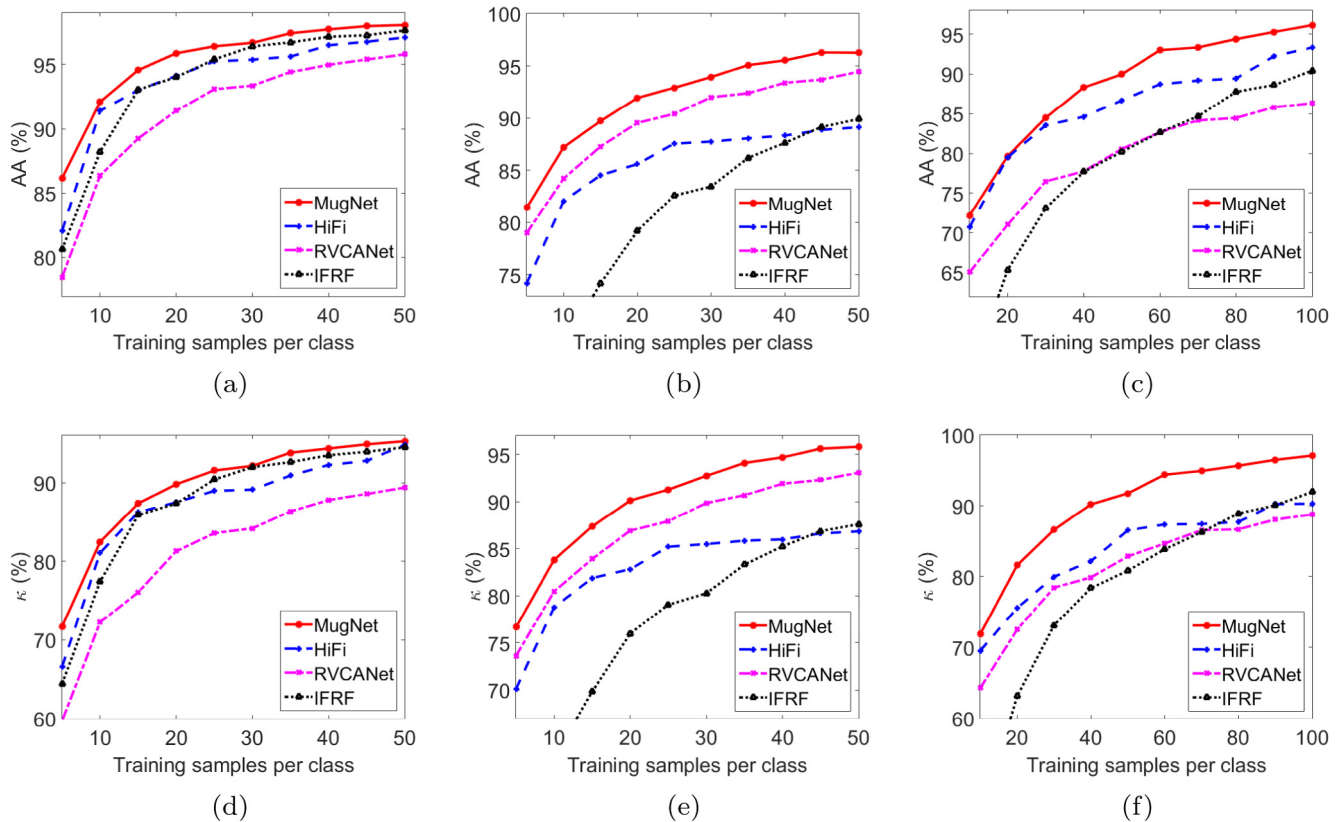
Classification accuracies of different methods on grss\_dfc\_2014 dataset.

Class	Samples		Methods					
	Train/Test	IFRF	EPF	NRS	RVCANet	HiFi	MugNet-S	MugNet
OA(%)		84.26 ± 2.07	68.52 ± 4.76	83.37 ± 1.60	86.40 ± 1.45	88.72 ± 1.51	92.76 ± 1.26	<b>93.15 ± 1.21</b>
AA(%)		79.70 ± 2.40	68.81 ± 2.77	82.30 ± 1.09	80.78 ± 1.77	87.81 ± 1.49	88.76 ± 1.22	<b>89.60 ± 1.82</b>
$\kappa \times 100$		80.65 ± 2.47	62.49 ± 5.09	79.46 ± 1.88	83.21 ± 1.72	86.06 ± 1.82	90.99 ± 1.55	<b>91.49 ± 1.49</b>
Road	50/4393	89.36 ± 2.96	97.21 ± 0.98	94.64 ± 1.94	95.55 ± 1.55	98.95 ± 0.82	95.98 ± 3.13	96.08 ± 1.89
Trees	50/1043	60.08 ± 8.18	17.09 ± 3.69	76.76 ± 4.16	54.47 ± 7.73	81.09 ± 6.25	67.25 ± 9.50	68.25 ± 10.4
Red roof	50/ 1804	74.43 ± 5.99	52.93 ± 8.23	73.90 ± 5.12	70.94 ± 5.12	83.33 ± 3.71	86.89 ± 3.49	87.58 ± 4.63
Grey roof	50/2076	75.89 ± 6.44	56.89 ± 9.11	85.40 ± 4.77	76.57 ± 3.88	83.89 ± 4.30	88.02 ± 2.63	88.72 ± 4.21
Concrete roof	50/3838	83.23 ± 5.56	92.89 ± 6.45	72.19 ± 5.64	91.86 ± 4.14	87.88 ± 3.91	97.68 ± 1.33	97.84 ± 1.51
Vegetation	50/7307	96.96 ± 1.42	87.65 ± 3.70	83.73 ± 5.20	95.35 ± 1.98	85.57 ± 3.75	94.08 ± 3.80	97.78 ± 1.11
Bare soil	50/1721	77.95 ± 7.83	77.03 ± 12.1	89.49 ± 4.12	80.69 ± 7.13	93.98 ± 3.19	91.39 ± 2.14	90.91 ± 3.65

**Fig. 9.** Classification maps for Indian Pines data set by (a) IFRF, (b) EPF, (c) NRS, (d) RVCANet, (e) HiFi and (f) MugNet.**Fig. 10.** Classification maps for grss\_dfc\_2013 data set by (a) IFRF, (b) EPF, (c) NRS, (d) RVCANet, (e) HiFi and (f) MugNet. Because this data set is too large, to provide a better visual effect, we only show part of it.**Fig. 11.** Classification maps for grss\_dfc\_2014 data set by (a) IFRF, (b) EPF, (c) NRS, (d) RVCANet, (e) HiFi and (f) MugNet.

e.g., classes alfalfa, grass-pasture-mowed and oats. All the 6 methods perform well in these classes although training samples are limited. However, the reason may be that data distribution in such small-scale sample sets are quite fixed. In other words, the results by all the methods are actually overfitting. Therefore, it is more meaningful to focus on the classes with many testing samples. In

this case, OA is a more objective metric than AA for Indian Pines data set. We note that MugNet outperforms IFRF and HiFi in OA by about 1.5%, and semi-supervised manner presents slightly better results. Overall, due to the problem of sample unbalance, this data set alone cannot verify the effectiveness of MugNet. More data sets are necessary.



**Fig. 12.** Influence of training samples number. AA on (a) Indian Pines, (b) grss\_dfc\_2013, (c) grss\_dfc\_2014.  $\kappa$  on (d) Indian Pines, (e) grss\_dfc\_2013, (f) grss\_dfc\_2014.

### 3.4.2. Results on grss\_dfc\_2013 data set

This data set is originally designed for classification contest (IEEE GRSS Data Fusion Contest, 2013), thus the results on it are more convincing. Actually this data set is more challenging than some popular ones such as Pavia University, according to Pan et al. (2017c). Although there are only 20 pixels per class for training, MugNet can still work well. Among all the six methods only MugNet achieves more than 90% OA, AA and  $\kappa$ . When evaluating the performance of small-scale samples classifiers, the accuracy in each class is also important. We note that MugNet reports at least 85% accuracies for all the classes, while all of the others have several classes below 80%.

### 3.4.3. Results on grss\_dfc\_2014 data set

Compared with the other two data sets, grss\_dfc\_2014 is much more challenging. The results reported in Table 3 are obtained with 50 training samples per class, rather than 20. However, because there are only 7 classes in this data set, totally 350 pixels are used for training which is very similar to the other two (320 in Indian Pines and 300 in grss\_dfc\_2013). Since there are more than 22,000 pixels are labeled, our setting could be still considered as “limited samples”. The advantage of MugNet in this data set is more significant, especially in OA and  $\kappa$ . The accuracy of MugNet on trees is not satisfying (below 70%), however, MugNet performs well in the other 6 classes. Among all the 7 classes, MugNet outperforms others in 4 ones. The effectiveness of semi-supervised strategy is also verified, according to Table 3.

Overall, MugNet has achieved better performance than the compared methods in the selected data sets. Another point we have to emphasize is that as an ensemble method, the individual learners in MugNet are the feature extraction processes in single grains. Although the improvements from spectral/spatial MuNet

to MugNet are not very apparent, we just hope that they are significant from individual learners to the final ensemble model.

## 3.5. Analysis and discussion

### 3.5.1. Influence of training samples

Since we mainly target at small-scale classification problem, in this section we discuss the influence of training samples number. Fig. 12 displays the curves of AA and  $\kappa$  under different number of training samples. We only show the curves of IFRF, HiFi, RVCA-Net and MugNet because they have presented the closest performance in Tables 1–3.

We can see from Fig. 12 that MugNet outperforms the compared methods in most cases. In Indian Pines and grss\_dfc\_2013 data sets, MugNet has achieved more than 85% AA and 70%  $\kappa$  with only 5 training samples per class. Classification on grss\_dfc\_2014 data set is more challenging, but the AA and  $\kappa$  by MugNet are still higher than other methods. As is expected, the accuracies are promoted with the increase of training samples' number. However, we note that the gaps among different methods shrink with sufficient training samples. For example, in Fig. 12(a)–(d), IFRF, HiFi and MugNet have reported very close results with 50 training samples per class. In this case, it is not safe to conclude which method is better. Therefore, the results obtained by limited samples may be more important for evaluation. Results in Fig. 12 indicates that MugNet has about 2–5% advantage over the compared methods with 10–20 training samples per class.

### 3.5.2. Statistical evaluation

To further validate that the observed advantages acquired by MugNet is statistically significant, we conduct paired t-test on OA, AA and  $\kappa$ . In paired t-test, we accept the hypothesis that  $\bar{m}_1$  is larger than  $\bar{m}_2$  only if

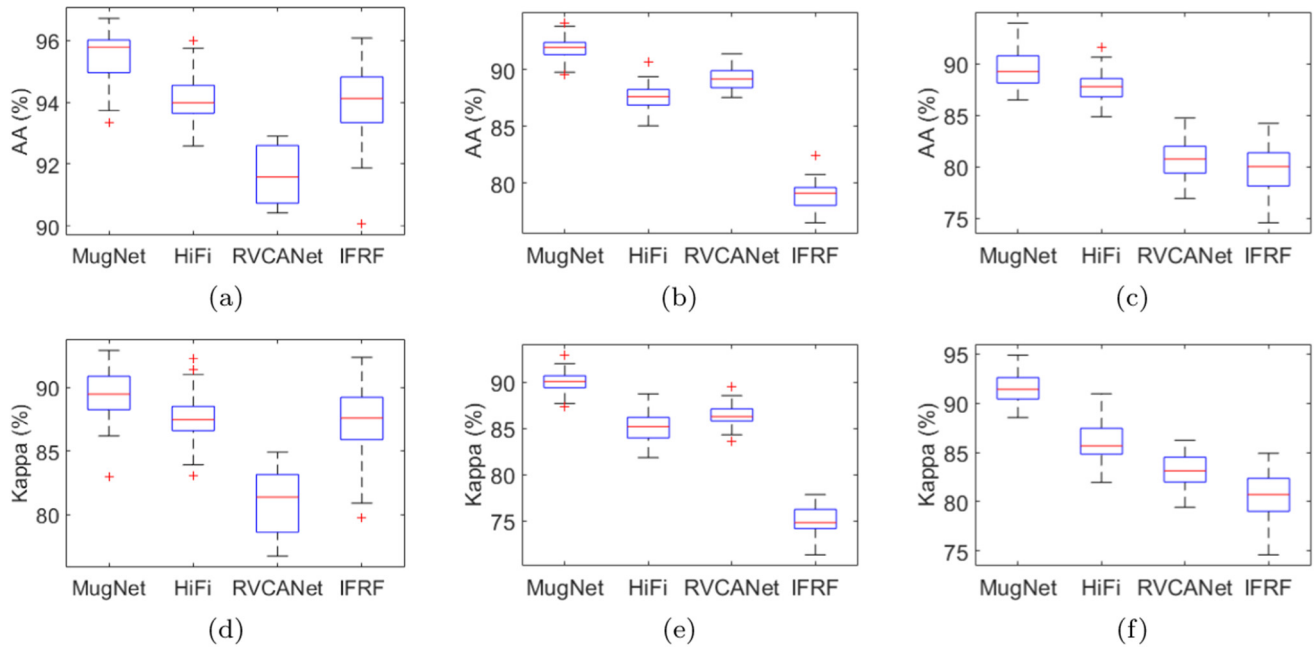


Fig. 13. Statistical evaluation. AA on (a) Indian Pines, (b) grss\_dfc\_2013, (c) grss\_dfc\_2014.  $\kappa$  on (d) Indian Pines, (e) grss\_dfc\_2013, (f) grss\_dfc\_2014.

$$\frac{(\bar{m}_1 - \bar{m}_2)\sqrt{n_1 + n_2 - 2}}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)(n_1 s_1^2 + n_2 s_2^2)}} > t_{1-\alpha}[n_1 + n_2 - 2], \quad (9)$$

where  $\bar{m}_1$  and  $\bar{m}_2$  are the mean values of OA, AA or  $\kappa$ ,  $n_1$  and  $n_2$  are running times which is set as 50 in our experiments,  $s_1$  and  $s_2$  are standard deviations, and  $t_{1-\alpha}$  is the  $\alpha$ th best quantile of the Student's law. According to the results in Tables 1–3, MugNet outperforms all the compared methods at confidence level 95%. The detailed statistics of AA and  $\kappa$  is shown in Fig. 13. Fig. 13 contains several box plots where the distribution of results are illustrated. Similar to the above discussion, only results by MugNet, HiFi, RVCANet and IFRF are presented in Fig. 13 because they achieve the closest accuracies. In a box plot, the red<sup>3</sup> line denotes the median of multiple running result, the box height denotes interquartile range, and the red “+” are outliers. Fig. 13 also demonstrates the effectiveness of MugNet.

#### 4. Conclusion

Although deep learning has presented excellent performance in many research fields, more studies are still required to further explore its application in HSI classification. One of the major challenge is that labeled pixels in HSI is not sufficient, which may lead to overfitting or underfitting. Fortunately, compared with visual images, HSIs have provided abundant spectral signatures which could produce extra discriminant information. Moreover, the spatial relationship between neighboring pixels can also contribute for classification.

Inspired by this recognition, we develop a new deep learning method, MugNet, which aims at making full use of the spectral and spatial correlations. MugNet is constructed on the basis of a simplified network, PCANet, where required hyper-parameters have been significantly reduced. In MugNet, multi-grained scanning strategy is utilized to represent the spectral and spatial relationships in different grains. In each grain there is an individual

deep learning model, and the final classification results are obtained by an ensemble approach. Therefore, MugNet could be regarded as an ensemble deep learning based method. Furthermore, to better deal with the problem of limited samples, we introduce semi-supervised method to MugNet. The experiments have shown that MugNet outperforms some state-of-the-art HSI classification methods even in the case of small-scale training set.

In our future work, we will try to transform MugNet to a completely end-to-end manner, and further improve its computational efficiency.

#### Acknowledgement

The work was supported by the National Natural Science Foundation of China under the Grants 61671037, the Beijing Natural Science Foundation under the Grant 4152031, and the Excellence Foundation of BUAA for PhD Students under the Grants 2017057.

The authors would like to express their great appreciation to NCALM for providing the grss\_dfc\_2013 data set, to Xiong Zhou, Minshan Cui, Abhinav Singhania, Dr. Juan Carlos Fernández Díaz and the data fusion technical committee for their preparation and pre-process for this data set.

The authors would also like to thank Telops Inc., Québec City, QC, Canada, for acquiring and providing the grss\_dfc\_2014 data set used in this paper, the IEEE GRSS Image Analysis and Data Fusion Technical Committee and Dr. M. Shimoni at the Signal and Image Center, Royal Military Academy, Brussels, Belgium, for organizing the 2014 Data Fusion Contest, the Center de Recherche Public Gabriel Lippmann (CRPGL), Luxembourg, and Dr. M. Schlerf at CRPGL, for their contribution of the Hyper-Cam LWIR sensor, and Dr. M. De Martino at the University of Genoa, Genoa, Italy, for her contribution to data preparation.

#### References

- Benediktsson, J.A., Palmason, J.A., Sveinsson, J.R., 2005. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. Geosci. Rem. Sens.* 43 (3), 480–491.

<sup>3</sup> For interpretation of color in Fig. 13, the reader is referred to the web version of this article.

- Chan, J.C.-W., Paelinckx, D., 2008. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Rem. Sens. Environ.* 112 (6), 2999–3011.
- Chan, T.H., Jia, K., Gao, S., Lu, J., Zeng, Z., Ma, Y., 2015. PCANet: a simple deep learning baseline for image classification? *IEEE Trans. Image Process.* 24 (12), 5017–5032.
- Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y., 2014. Deep learning-based classification of hyperspectral data. *IEEE J. Select. Top. Appl. Earth Observ. Rem. Sens.* 7 (6), 2094–2107.
- Chen, Y., Zhao, X., Jia, X., 2015. Spectral–spatial classification of hyperspectral data based on deep belief network. *IEEE J. Select. Top. Appl. Earth Observ. Rem. Sens.* 8 (6), 2381–2392.
- Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P., 2016. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans. Geosci. Rem. Sens.* 54 (10), 6232–6251.
- Cheng, G., Zhou, P., Han, J., 2016. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Trans. Geosci. Rem. Sens.* 54 (12), 7405–7415.
- Cheng, G., Han, J., Lu, X., 2017. Remote sensing image scene classification: benchmark and state of the art. *Proc. IEEE* (99), 1–19.
- Debes, C., Merentitis, A., Heremans, R., Hahn, J., Frangiadakis, N., van Kasteren, T., Liao, W., Bellens, R., Pizurica, A., Gautama, S., Philips, W., Prasad, S., Du, Q., Pacifici, F., 2014. Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest. *IEEE J. Select. Top. Appl. Earth Observ. Rem. Sens.* 7 (6), 2405–2418.
- Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., Lin, C.-J., 2008. LIBLINEAR: a library for large linear classification. *J. Mach. Learn. Res.* 9 (August), 1871–1874.
- Ghamisi, P., Dalla Mura, M., Benediktsson, J.A., 2015. A survey on spectral–spatial classification techniques based on attribute profiles. *IEEE Trans. Geosci. Rem. Sens.* 53 (5), 2335–2353.
- Gu, Y., Wang, C., You, D., Zhang, Y., Wang, S., Zhang, Y., 2012. Representative multiple kernel learning for classification in hyperspectral imagery. *IEEE Trans. Geosci. Rem. Sens.* 50 (7), 2852–2865.
- Gu, Y., Wang, Q., Wang, H., You, D., 2015. Multiple kernel learning via low-rank nonnegative matrix factorization for classification of hyperspectral imagery. *IEEE J. Select. Top. Appl. Earth Observ. Rem. Sens.* 8 (6), 2739–2751.
- Gu, Y., Liu, T., Jia, X., Benediktsson, J.A., 2016. Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification. *IEEE Trans. Geosci. Rem. Sens.* 54 (6), 3235–3247.
- Ham, J., Chen, Y., Crawford, M.M., Ghosh, J., 2005. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Trans. Geosci. Rem. Sens.* 43 (3), 492–501.
- Han, J., Zhang, D., Cheng, G., Guo, L., Ren, J., 2015. Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. *IEEE Trans. Geosci. Rem. Sens.* 53 (6), 3325–3337.
- He, L., Li, J., Plaza, A., Li, Y., 2017. Discriminative low-rank Gabor filtering for spectral–spatial hyperspectral image classification. *IEEE Trans. Geosci. Rem. Sens.* 55 (3), 1381–1395.
- Huang, X., Zhang, L., 2013. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE Trans. Geosci. Rem. Sens.* 51 (1), 257–272.
- Hu, W., Huang, Y., Wei, L., Zhang, F., Li, H., 2015. Deep convolutional neural networks for hyperspectral image classification. *J. Sens.* 2015 (2), 1–12.
- Huo, L.Z., Tang, P., Zhang, Z., Tuia, D., 2015. Semisupervised classification of remote sensing images with hierarchical spatial similarity. *IEEE Geosci. Rem. Sens. Lett.* 12 (1), 150–154.
- 2013 IEEE GRSS Data Fusion Contest <<http://www.grss-ieee.org/community/technical-committees/data-fusion/>>.
- 2014 IEEE GRSS Data Fusion Contest <<http://www.grss-ieee.org/community/technical-committees/data-fusion/>>.
- Kang, X., Li, S., Benediktsson, J.A., 2014a. Spectral–spatial hyperspectral image classification with edge-preserving filtering. *IEEE Trans. Geosci. Rem. Sens.* 52 (5), 2666–2677.
- Kang, X., Li, S., Benediktsson, J.A., 2014b. Feature extraction of hyperspectral images with image fusion and recursive filtering. *IEEE Trans. Geosci. Rem. Sens.* 52 (6), 3742–3752.
- Kang, X., Li, S., Fang, L., Benediktsson, J.A., 2015. Intrinsic image decomposition for feature extraction of hyperspectral images. *IEEE Trans. Geosci. Rem. Sens.* 53 (4), 2241–2253.
- Khodadadzadeh, M., Li, J., Plaza, A., Ghassemian, H., Bioucas-Dias, J.M., Li, X., 2014. Spectral–spatial classification of hyperspectral data using local and global probabilities for mixed pixel characterization. *IEEE Trans. Geosci. Rem. Sens.* 52 (10), 6298–6314.
- Li, W., Du, Q., 2014. Gabor-filtering-based nearest regularized subspace for hyperspectral image classification. *IEEE J. Select. Top. Appl. Earth Observ. Rem. Sens.* 7 (4), 1012–1022.
- Liao, W., Pizurica, A., Bellens, R., Gautama, S., Philips, W., 2015. Generalized graph-based fusion of hyperspectral and lidar data using morphological features. *IEEE Geosci. Rem. Sens. Lett.* 12 (3), 552–556.
- Liao, W., Huang, X., Coillie, F.V., Gautama, S., Pizurica, A., Philips, W., Liu, H., Zhu, T., Shimoni, M., Moser, G., Tuia, D., 2015. Processing of multiresolution thermal hyperspectral and digital color data: outcome of the 2014 IEEE GRSS data fusion contest. *IEEE J. Select. Top. Appl. Earth Observ. Rem. Sens.* 8 (6), 2984–2996.
- Li, J., Reddy Marpu, P., Plaza, A., Bioucas-Dias, J.M., Atli Benediktsson, J., 2013. Generalized composite kernel framework for hyperspectral image classification. *IEEE Trans. Geosci. Rem. Sens.* 51 (9), 4816–4829.
- Li, J., Huang, X., Gamba, P., Bioucas-Dias, J.M., Zhang, L., Benediktsson, J.A., Plaza, A., 2015. Multiple feature learning for hyperspectral image classification. *IEEE Trans. Geosci. Rem. Sens.* 53 (3), 1592–1606.
- Li, W., Wu, G., Zhang, F., Du, Q., 2017. Hyperspectral image classification using deep pixel-pair features. *IEEE Trans. Geosci. Rem. Sens.* 55 (2), 844–853.
- Liu, Y., Cao, G., Sun, Q., Siegel, M., 2015. Hyperspectral classification via deep networks and superpixel segmentation. *Int. J. Rem. Sens.* 36 (13), 3459–3482.
- Lu, X., Wu, H., Yuan, Y., Yan, P., Li, X., 2013. Manifold regularized sparse NMF for hyperspectral unmixing. *IEEE Trans. Geosci. Rem. Sens.* 51 (5), 2815–2826.
- Ma, A., Zhong, Y., Zhao, B., Jiao, H., Zhang, L., 2016. Semisupervised subspace-based DNA encoding and matching classifier for hyperspectral remote sensing imagery. *IEEE Trans. Geosci. Rem. Sens.* 54 (8), 4402–4418.
- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. *ISPRS J. Photogram. Rem. Sens.* 66 (3), 247–259.
- Pal, M., Foody, G.M., 2010. Feature selection for classification of hyperspectral data by SVM. *IEEE Trans. Geosci. Rem. Sens.* 48 (5), 2297–2307.
- Pan, B., Shi, Z., Zhang, N., Xie, S., 2016. Hyperspectral image classification based on nonlinear spectral–spatial network. *IEEE Geosci. Rem. Sens. Lett.* 13 (12), 1782–1786.
- Pan, B., Shi, Z., An, Z., Jiang, Z., Ma, Y., 2017a. A novel spectral-unmixing-based green algae area estimation method for GOCI data. *IEEE J. Select. Top. Appl. Earth Observ. Rem. Sens.* 10 (2), 437–449.
- Pan, B., Shi, Z., Xu, X., 2017b. R-VCANet: a new deep-learning-based hyperspectral image classification method. *IEEE J. Select. Top. Appl. Earth Observ. Rem. Sens.* 10 (5), 1975–1986.
- Pan, B., Shi, Z., Xu, X., 2017c. Hierarchical guidance filtering-based ensemble classification for hyperspectral images. *IEEE Trans. Geosci. Rem. Sens.* 55 (7), 4177–4189.
- Rodriguez-Galiano, V., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogram. Rem. Sens.* 67, 93–104.
- Romaszewski, M., Głomb, P., Cholewa, M., 2016. Semi-supervised hyperspectral classification from a small number of training samples using a co-training approach. *ISPRS J. Photogram. Rem. Sens.* 121, 60–76.
- Santos, A.B., De, A., Araujo, Albuquerque, Menotti, D., 2013. Combining multiple classification methods for hyperspectral data interpretation. *IEEE J. Select. Top. Appl. Earth Observ. Rem. Sens.* 6 (3), 1450–1459.
- Tuia, D., Camps-Valls, G., 2009. Semisupervised remote sensing image classification with cluster kernels. *IEEE Geosci. Rem. Sens. Lett.* 6 (2), 224–228.
- Wan, L., Tang, K., Li, M., Zhong, Y., Qin, A.K., 2015. Collaborative active and semisupervised learning for hyperspectral remote sensing image classification. *IEEE Trans. Geosci. Rem. Sens.* 53 (5), 2384–2396.
- Wang, Z., Nasrabadi, N.M., Huang, T.S., 2015. Semisupervised hyperspectral classification using task-driven dictionary learning with laplacian regularization. *IEEE Trans. Geosci. Rem. Sens.* 53 (3), 1161–1173.
- Wang, Q., Meng, Z., Li, X., 2017. Locality adaptive discriminant analysis for spectral–spatial classification of hyperspectral images. *IEEE Geosci. Rem. Sens. Lett.* 14 (11), 2077–2081.
- Xia, J., Chansussot, J., Du, P., He, X., 2016a. Rotation-based support vector machine ensemble in classification of hyperspectral data with limited training samples. *IEEE Trans. Geosci. Rem. Sens.* 54 (3), 1519–1531.
- Xia, J., Bombrun, L., Adali, T., Berthoumieu, Y., Germain, C., 2016b. Spectral–spatial classification of hyperspectral images using ICA and edge-preserving filter via an ensemble strategy. *IEEE Trans. Geosci. Rem. Sens.* 54 (8), 4971–4982.
- Xu, X., Shi, Z., 2017. Multi-objective based spectral unmixing for hyperspectral images. *ISPRS J. Photogram. Rem. Sens.* 124, 54–69.
- Xu, X., Li, J., Huang, X., Mura, M.D., Plaza, A., 2016. Multiple morphological component analysis based decomposition for remote sensing image classification. *IEEE Trans. Geosci. Rem. Sens.* 54 (5), 3083–3102.
- Yang, L., Yang, S., Jin, P., Zhang, R., 2014. Semi-supervised hyperspectral image classification using spatio-spectral Laplacian support vector machine. *IEEE Geosci. Rem. Sens. Lett.* 11 (3), 651–655.
- Yao, X., Han, J., Cheng, G., Qian, X., Guo, L., 2016. Semantic annotation of high-resolution satellite images via weakly supervised learning. *IEEE Trans. Geosci. Rem. Sens.* 54 (6), 3660–3671.
- Zhang, Q., Shen, X., Xu, L., Jia, J., 2014. Rolling guidance filter. In: *European Conference on Computer Vision (ECCV)*. Springer, pp. 815–830.
- Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensing data: a technical tutorial on the state of the art. *IEEE Geosci. Rem. Sens. Mag.* 4 (2), 22–40.
- Zhao, W., Guo, Z., Yue, J., Zhang, X., Luo, L., 2015. On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery. *Int. J. Rem. Sens.* 36 (13), 3368–3379.
- Zhong, Z., Fan, B., Ding, K., Li, H., Xiang, S., Pan, C., 2016. Efficient multiple feature fusion with hashing for hyperspectral imagery classification: a comparative study. *IEEE Trans. Geosci. Rem. Sens.* 54 (8), 4461–4478.
- Zhong, P., Gong, Z., Li, S., Schonlieb, C.B., 2017. Learning to diversify deep belief networks for hyperspectral image classification. *IEEE Trans. Geosci. Rem. Sens.* PP (99), 1–15.
- Zhou, Z.H., 2012. *Ensemble Methods: Foundations and Algorithms*, 8(1). Taylor & Francis, pp. 77–79.
- Zhou, Z.-H., Feng, J., 2017. Deep Forest: Towards an Alternative to Deep Neural Networks.