Hyperspectral Image Classification Based on Deep Forest and Spectral-Spatial Cooperative Feature

Mingyang Li[1], Ning Zhang[2], Bin Pan[1], Shaobiao Xie[3], Xi Wu[1],
and Zhenwei Shi[1(✉)]

[1] Image Processing Center, School of Astronautics,
Beihang University, Beijing, China
shizhenwei@buaa.edu.cn
[2] Shanghai Aerospace Electronic Technology Institute, Shanghai, China
[3] School of Astronautics, Harbin Institute of Technology, Harbin, China

Abstract. Recently, deep-learning-based methods have displayed promising performance for hyperspectral image (HSI) classification. However, these methods usually require a large number of training samples, and the complex structure and time-consuming problem have restricted their applications. Deep forest, a decision tree ensemble approach with performance highly competitive to deep neural networks. Deep forest can work well and efficiently even when there are only small-scale training data. In this paper, a novel simplified deep framework is proposed, which achieves higher accuracy when the number of training samples is small. We propose the framework which employs local binary patterns (LBPS) and gabor filter to extract local-global image features. The extracted feature along with original spectral features will be stacked, which can achieve concatenation of multiple features. Finally, deep forest will extract deeper features and use strategy of layer-by-layer voting for HSI classification.

Keywords: Deep-learning · Deep forest
Local binary patterns (LBPS) · Gabor filter

1 Introduction

Depending on the development of sensor technology, hyperspectral sensors could provide images owning hundreds of bands with high spatial and spectral information. Hundreds of spectral band values which are continuous and narrow are recorded as a data cube, with the spectral resolution of nanometer level. Due to these advantages, the applications of hyperspectral data have been widely used in many fields such as spectral unmixing [15] and environmental monitoring [13]. Hyperspectral image classification is one of the most important technologies for these applications. However, hyperspectral image classification is still a challenge problem owing to its complex characteristic. The high dimensionality may

produce curse of dimensionality [5], and the abundant spectral bands may also bring noise to decrease performance of the classification. Therefore, we could not simply use spectral signals for hyperspectral sensing image classification.

During the last decade, many traditional methods based on feature extraction have been proposed to solve the problem. These methods use spectral and spatial information, and the classification algorithms [4,11,16].

In recent years, deep learning has been implemented in various fields such as image classification [7]. In [2], Chen et al. proposed applying deep-learning-based methods to handle HSI classification for the first time, where autoencoder (AE) is utilized to learn deep features of HSI. After that, several simplified deep learning methods are developed [12,14]. However, methods based on deep learning could suffer from the complex framework of neural networks, and the performance is limited by the small number of training samples. Furthermore, the methods also suffer from a time-consuming training process, and final experimental results are not easy to reproduce.

According to the motivations and problems mentioned above, we introduce a method based on deep forest [18] that can handle HSI classification with limited training samples. Compared to deep neural networks, deep forest achieves highly comparable performance efficiently. What is more, the hyper-parameters of deep forest is quite small, and the result is less sensitive to parameter setting. It will spend less time in training process and perform well on small-scale samples. In this letter, we propose a deep framework combining with spectral-spatial cooperative features for deeper HSI features extraction and classification, which achieves better performance with much less training samples than deep learning methods and traditional methods. To take fully into account the globality in the feature extraction, spectral-spatial cooperative feature combines local and global features with the original spectral feature. Furthermore, considering the feature learned from the last layer of deep forest may not be the best representative feature, we improve the framework of deep forest. We add a voting mechanism in the framework, and have a better experimental result.

The remainder of this paper is organized as follows. In Sect. 2, we first present a roughly process about the method, then we present a detailed description of spectral-spatial cooperative feature. At last, we introduce deep forest in detail, which also includes adding a voting mechanism in the framework. In Sect. 3, the experimental results and comparison experiments are displayed. Finally, we come to a conclusion in Sect. 4.

2   The Proposed Method

The proposed features follow three parallel strategies: local feature extraction, global feature extraction and the original spectral feature. Figure 1 shows the overall implementation of the classification method. LBP operator is applied to the calculation of the entire band images to obtain LBP feature map, and the local feature is obtained by counting the histogram of each neighbourhood of the center pixel. The global feature is produced by using gabor filter which captures
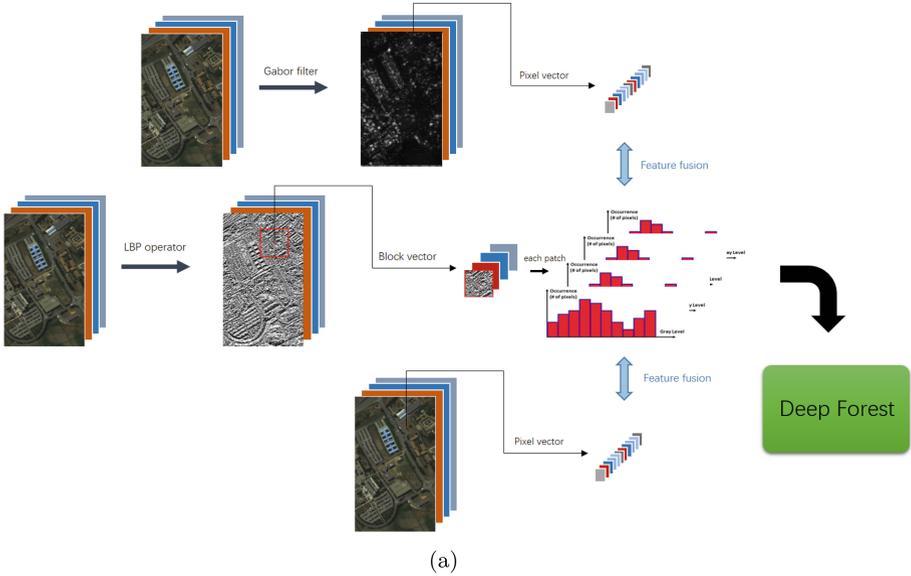
(a)

Fig. 1. Simple flowchart of DFCF method

texture features at different angles, and the original spectral feature of the central pixel is drawn out. Then we combine the features to get the spectral-spatial cooperative feature, while deep forest with a voting mechanism is employed to deeper feature extraction and the final classification predictions. Here, we present a brief introduction of Gabor filter and LBP operator at first, then, we describe feature fusion about spectral-spatial cooperative feature. Finally, the framework of deep forest is introduced.

2.1 Gabor Filter

A gabor filter, which can extract the relevant features in different scales and directions in the frequency domain, and its frequency and direction of expression are similar to the human visual system. The research shows that gabor filters are very suitable for texture expression and separation [3]. Therefore, it's widely used for computer vision and image processing. In the 2-D gabor filter model, the filter consists of real component and imaginary component. The mathematical expressions can be expressed as

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi \frac{x'}{\lambda} + \psi\right)\right) \quad (1)$$

real component:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right) \quad (2)$$

imaginary component:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \sin\left(2\pi\frac{x'}{\lambda} + \psi\right) \tag{3}$$

where:

$$x' = x\cos\theta + y\sin\theta; y' = -x\cos\theta + y\sin\theta \tag{4}$$

where λ is wavelength of the sinusoidal function, θ specifies the direction of the parallel stripes of the gabor function, which takes a value of 0 to 360°. ψ represents phase offset. γ, spatial aspect ratio, usually setting to 0.5, specifies the ellipticity of the gabor function support. σ is only determined by λ and b as

$$\sigma = \frac{\lambda}{\pi}\sqrt{\frac{\ln 2}{2} \cdot \frac{2^b + 1}{2^b - 1}} \tag{5}$$

2.2   Local Binary Pattern

LBP (Local Binary Pattern)[10] is an operator used to describe the local texture features of an image, and it has significant advantages such as rotational invariance and gray-scale invariance. For each given center pixel v_c (scalar value), we find the corresponding pixel v_i on the specified radius r to compare with v_c, where r determines the distance from the center pixel. If the pixel value is higher than the center pixel value, set to 1, otherwise set to 0. After selecting p neighbors $\{v_0, v_1, \ldots, v_{p-1}\}$, the LBP calculating the center pixel v_c is as followed

$$LBP_{p,r}(v_c) = \sum_{i=0}^{p-1} S(v_p - v_c)2^p \tag{6}$$

where $S(v_p - v_c) = 1$ if $v_p > v_c$ and $S(v_p - v_c) = 0$ if $v_p < v_c$. Through the above formula, given an image, we can get the LBP value of each pixel (The direction of the calculation process is clockwise). While each pixel does not fall to an integer position, its gray value is estimated by bilinear interpolation based on the pixel gray value of the nearest two integer positions within the radius track. The LBP value reflects local texture information and smoothness. Then, an histogram is computed over a local patch, which represents LBP feature of the center pixel.

2.3   Spectral-Spatial Cooperative Feature

By utilizing gabor to filter the spectral image of each spectral band in a given HSI, we can extract global texture features in different directions and different scales. At the same time, the lbp feature map is obtained by computing the spectral image of each spectral band in a given HSI through the LBP operator. The histogram feature is calculated for the fixed size neighborhood of each central pixel, thus obtaining local texture features. In this paper, before performing the above calculations, we use PCA to reduce the spectral dimension due to

many of spectral bands containing redundant information. We stack the local texture and the global texture feature with its original spectral feature to form the initial extracted feature. The feature contains local and global features, also including spectral and spatial information. Therefore, it is called Spectral-Spatial Cooperative Feature.
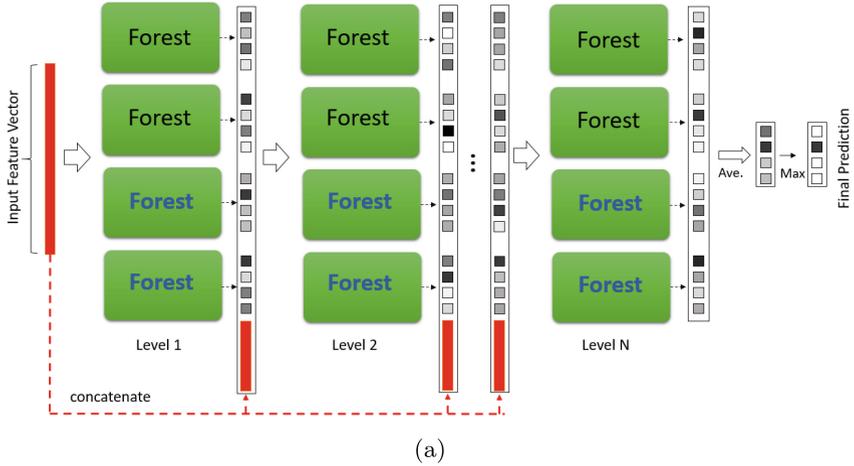

(a)

Fig. 2. The structure of Deep Forest. Each level outputs a distribution vector, then, the vector is concatenated with input feature vector, which is formed as a new feature to input the next level. Finally, output of the last level is averaged to get a probability distribution, and set the prediction label as the highest probability of the label.

2.4 Deep Forest

Cascade Forest. The structure of the network in deep neural networks mostly bases on the layer-by-layer stacking, and it is utilized to process features. In view of this, Zhou and Feng [18]. proposed deep forest, a novel decision tree ensemble method, which employs a cascade structure. The illustration is shown in Fig. 2, where the input received by each level is obtained by the preceding level, and the processing result of this level is outputted to the next level.

Each level represents an ensemble of decision tree forests. To encourage the diversity, each level has different decision tree forest. In Fig. 2, complete-random tree forests and random forests [1] is used in the structure. In this paper, we simply use three complete-random tree forests to structure each level. Each tree forest contains 900 complete-random trees. With randomly selecting a feature for split at each node of the tree, the tree forest is generated. Then grow the tree until each leaf node contains only the same class of instances or no more than 10 instances. For each forest, it will estimate the distribution of the class by counting the percentage of different class of the training samples at the leaf node where the instance concerned falls into. Finally, the distribution of the classes will be averaged across all trees, and each forest will output distribution of each class.

Each forest outputs a distribution of each class, which forms a class vector. Then, the original vector is concatenated with the class vector to form a new feature vector that inputs to the next level of deep forest. For instance, based on a classification task with 16 classes, each of the three forests will output a sixteen-dimensional class vector. Thus, the input feature of the next level will augment $3 \times 16 = 48$ dimensions.

In contrast to most deep neural networks, deep forest can handle different scales of training data, including small-scale ones. The reason is that, for a validation set, the model complexity of deep neural networks is fixed. However, after expanding a new level, the training process of deep forest will be terminated when the performance is not significantly improved. Therefore, the number of levels will be automatically determined. Furthermore, in order to reduce the risk of over-fitting, the output of each cascade level will be generated by a number of cross-validations to ensure that the output is sufficient to characterize the input.
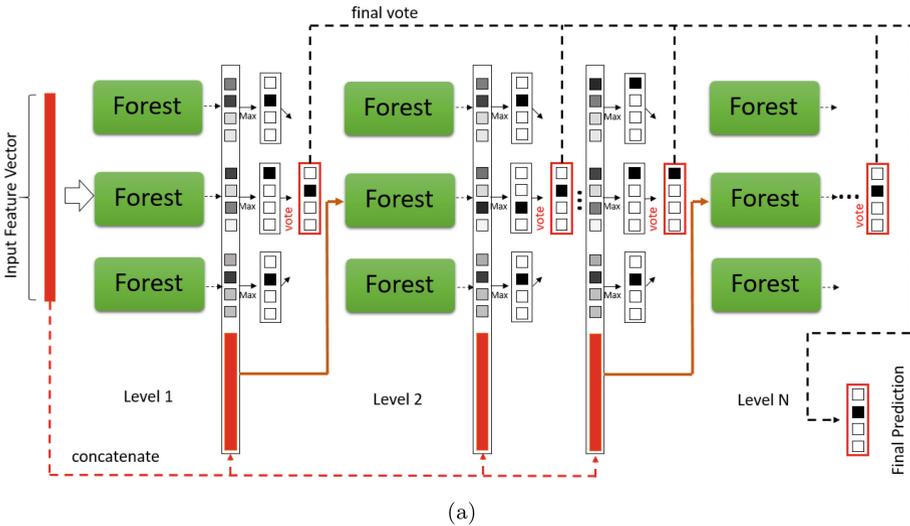

(a)

Fig. 3. Cascade forest with the voting mechanism. The gray scale in a small box represents a probability value of the class. When the box is black, the probability is 1.

A Voting Mechanism. In this paper, we structure four levels deep forest to handle HSI classification task whose number of training data is poor. There are three complete-random tree forests in each level. Each tree forest outputs a distribution of each class and a prediction label. For HSI classification task, the distribution each level outputs and the prediction labels are produced by studying the feature this level inputs. In each layer, the feature of input is not identical so that it will lead to the prediction of each layer is also diverse, as it is well known that diversity [17] is crucial for ensemble construction. Therefore, we employ each layer of the prediction results, joining a voting mechanism in

deep forest model. The illustration is shown in Fig. 3. Given a feature as input, we can get three predictions in each level. Then, the final prediction of this level is obtained by using the three predictions to vote. Finally, the final predictions of each level will vote to form a result prediction. Compared with the original model, the model not only relies on the output distribution of the previous level, but also uses the prediction label to affect the next level, which increases the diversity of the structure and the final classification can get more accurate label results.

3 Experimental Results

3.1 Data Sets and Experimental Setup

Two popular HSI datasets are used in our experiments: Indian Pines and Pavia University.

Indian Pines: The Indian Pine data set is one of the most common data sets in the HSI classification and is obtained from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) in Northwestern Indiana, India. The image size is 145×145, with wavelengths ranging from 0.4 to $2.5\,\mu m$. The data set originally had 220 spectral bands. In order to reduce the effect of the water absorption, 20 water absorption bands were removed, and the remaining 200 bands were used in the experiments. There are 16 classes of interests in the data set, with totally 10249 pixels which are labeled.

Pavia University: The data of Pavia University were collected in the city, Italy Pavia, through the reflective optical system imaging spectrometer (ROSIS-3) sensor. The data set originally had 115 spectral bands. Due to the effect of noise interference, 12 bands were removed and 103 bands is used in the experiment. The image size is 610×340, with wavelengths range from 0.43 to $0.86\,\mu m$. Spatial resolution is approximately $1.3\,m$. In this data set, a total of 42776 pixels are labeled, which are classified into 9 classes.
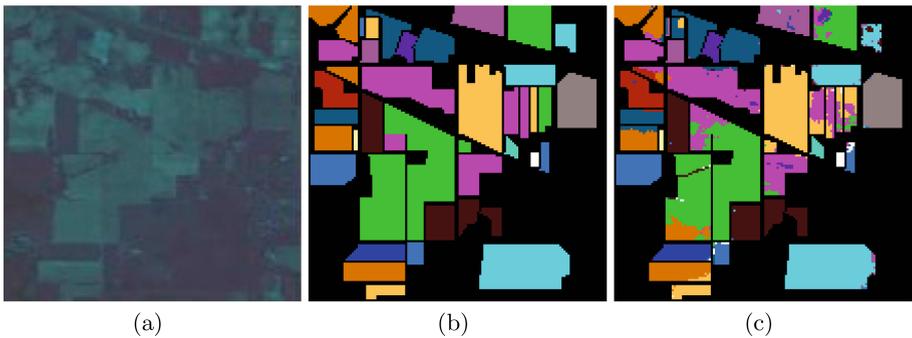

Fig. 4. Indian Pines. (a) False color composite image choosing R-G-B=bands 15-25-59. (b)Ground truth. (c)The prediction of DFCF.
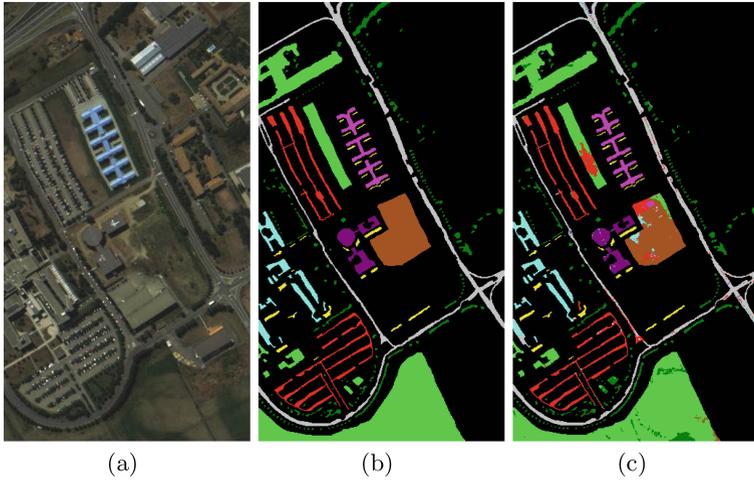
Fig. 5. Pavia University. (a) False color composite image choosing R-G-B=bands 15-40-65. (b)Ground truth. (c)The prediction of DFCF.

In comparison experiments, limited by space limitations, we refer three methods to compare the method which proposed in this paper, including GCK, NRS and EPF-G. GCK [8] is a composite-kernel-framework based method. NRS [9], based on gabor filter, extract global texture feature for HSI classification, while EPF-G [6] takes advantage of edge-preserving filtering to classify hyperspectral images. In order to be able to evaluate the performance of the methods more comprehensively, overall accuracy (OA), average accuracy (AA) and Kappa coefficient (κ) are employed in the experiments. In the experiment, some important parameters are set in the way of cross-validation. Due to space limitations, this article will not elaborate.

Indian Pines: At first, 30 principal components are used for operations. we use $b = 1$ and 8 orientations to set gabor filter, and patch size of LBP operator is 21×21.

Pavia University: At first, 30 principal components are used for operations. we use $b = 5$ and 8 orientations to set gabor filter, and patch size of LBP operator is 21×21.

Deep Forest: When we apply deep forest, the results are not sensitive to parameter settings. We build 3 complete-random forests on each layer of cascade forest and 4 levels are employed to structure deep forest for both of Pavia University and Indian Pines.

3.2 Classification Results

This section presents the results of each method, and also shows the superiority of the method we proposed. Each data set is randomly selected for 20 samples

Table 1. INDIAN PINES: the classification accuracy of different methods (%).

Class	Samples		Methods			
	Train	Test	GCK	NRS	EPF-G	DFCF
Alfalfa	20	26	96.94 ± 3.71	$\mathbf{99.92 \pm 0.55}$	62.07 ± 20.5	98.00 ± 1.92
Corn-notill	20	1408	$\mathbf{75.67 \pm 5.80}$	70.28 ± 7.43	74.83 ± 9.59	74.94 ± 6.64
Corn-mintill	20	810	81.59 ± 4.47	75.47 ± 7.71	74.68 ± 10.96	$\mathbf{86.23 \pm 5.87}$
Corn	20	217	93.17 ± 5.04	$\mathbf{94.58 \pm 4.59}$	42.07 ± 11.72	93.88 ± 4.67
Grass-pasture	20	463	89.70 ± 4.29	87.70 ± 5.18	$\mathbf{94.83 \pm 5.18}$	91.34 ± 4.06
Grass-trees	20	710	$\mathbf{97.57 \pm 1.29}$	92.55 ± 4.19	94.39 ± 3.90	96.89 ± 3.02
Grass-pasture-mowed	14	14	97.54 ± 3.94	$\mathbf{100 \pm 0.00}$	91.01 ± 12.62	99.57 ± 1.70
Hay-windrowed	20	458	99.42 ± 0.23	98.21 ± 2.31	99.85 ± 0.41	$\mathbf{99.95 \pm 0.10}$
Oats	10	10	$\mathbf{100.00 \pm 0.00}$	$\mathbf{100.00 \pm 0.00}$	78.18 ± 19.56	99.00 ± 7.00
Soybean-notill	20	952	80.98 ± 4.11	72.46 ± 8.17	69.21 ± 6.70	$\mathbf{85.55 \pm 5.50}$
Soybean-mintill	20	2435	79.87 ± 4.58	71.76 ± 8.39	84.40 ± 6.03	$\mathbf{85.70 \pm 5.36}$
Soybean-clean	20	573	84.07 ± 6.37	81.25 ± 7.01	59.43 ± 13.22	$\mathbf{89.61 \pm 5.36}$
Wheat	20	185	$\mathbf{99.56 \pm 0.29}$	99.18 ± 1.28	99.50 ± 1.33	99.06 ± 0.80
Woods	20	1245	93.74 ± 2.88	87.09 ± 5.12	97.27 ± 2.53	$\mathbf{97.47 \pm 1.53}$
Buildings-Grass-Trees-Drives	20	366	93.16 ± 3.66	90.43 ± 6.10	71.08 ± 15.00	$\mathbf{95.20 \pm 5.32}$
Stone-Steel-Towers	20	73	95.71 ± 4.69	$\mathbf{98.97 \pm 2.03}$	81.16 ± 7.79	98.16 ± 1.39
OA			85.54 ± 1.42	79.87 ± 1.82	78.45 ± 2.75	$\mathbf{88.56 \pm 1.82}$
AA			91.17 ± 0.85	88.74 ± 1.03	79.62 ± 2.35	$\mathbf{93.16 \pm 1.09}$
$\kappa \times 100$			83.61 ± 1.58	77.14 ± 2.03	75.64 ± 3.07	$\mathbf{86.99 \pm 2.04}$

per class for training and the rest for testing. All methods are performed 50 times, and the average value and standard deviation are presented in Tables 1 and 2.

(1) *The result of Indian Pines*: Indian data set is randomly selected 20 samples per class for training and the rest for testing. If half number of the samples is less than 20, we randomly select half number of the sample for training and the rest for testing. In Table 1, quantitative experimental results are listed. In all methods, the DFCF achieve the highest OA, AA and κ, and the standard deviation is relatively small, which indicates that the method has excellent classification accuracy and good stability for Indian Pines data set. In addition, in a total of 16 classes, there are 7 best class accuracies obtained by using DFCF, and 11 class accuracies of all classes are higher than 90%. Figure 4 shows the classification result obtained by utilizing DFCF method.

Table 2. PAVIA UNIVERSITY: the classification accuracy of different methods (%).

Class	Samples		Methods			
	Train	Test	GCK	NRS	EPF-G	DFCF
Asphalt	20	6611	80.91 ± 8.55	84.17 ± 5.64	$\mathbf{96.54 \pm 2.82}$	87.24 ± 6.49
Meadows	20	18629	$\mathbf{97.78 \pm 1.29}$	84.09 ± 7.35	94.67 ± 3.05	84.77 ± 5.65
Gravel	20	2079	74.60 ± 8.30	83.69 ± 7.89	80.75 ± 13.03	$\mathbf{90.30 \pm 5.82}$
Trees	20	3044	80.00 ± 12.53	91.69 ± 4.54	73.78 ± 14.59	$\mathbf{92.56 \pm 4.62}$
Painted metal sheets	20	1325	99.73 ± 0.54	$\mathbf{99.99 \pm 0.05}$	94.61 ± 3.90	99.66 ± 0.21
Bare	20	5009	87.69 ± 6.87	86.45 ± 4.89	60.08 ± 12.21	$\mathbf{94.60 \pm 4.09}$
Bitumen	20	1310	79.03 ± 11.13	86.30 ± 6.68	76.99 ± 11.96	$\mathbf{96.28 \pm 3.43}$
Self-Blocking Bricks	20	3662	70.02 ± 7.54	77.74 ± 8.29	84.96 ± 6.57	$\mathbf{94.39 \pm 4.20}$
Shadows	20	927	62.36 ± 10.52	92.91 ± 2.87	98.39 ± 1.30	$\mathbf{100.00 \pm 0.00}$
OA			86.36 ± 3.08	85.11 ± 3.08	83.54 ± 5.33	$\mathbf{89.12 \pm 2.64}$
AA			81.35 ± 3.38	87.45 ± 1.26	84.53 ± 3.82	$\mathbf{93.31 \pm 1.26}$
$\kappa \times 100$			82.30 ± 3.85	80.85 ± 3.67	79.12 ± 6.32	$\mathbf{85.99 \pm 3.25}$

(2) *The result of Pavia University*: Figure 5 and Table 2, respectively, show intu-
itional and quantitative experiment results on Pavia University data set. In
the case of the equal training numbles, each method achieves a better result
than the result achieved on Indian Pine data set. The reason may be that
the spatial resolution of Pavia University data set so that it achieves higher
classification result. In all methods, the DFCF still achieve the highest OA,
AA and κ, and the smallest standard deviation indicates the stability of the
method. When the task is limited to the number of training samples, AA
will become an important indicator. We have almost 7% advantages com-
pared to the highest result of the contrast methods. In addition, in all the 9
classes, DFCF work best in 6 classes, and exceeds 99% in 2 classes.

4　Conclusion

In this letter, we have introduced deep forest combining with spectral-spatial
cooperative features (DFCF) for deeper HSI feature extraction and classifica-
tion. DFCF is based on deep framework, thus, it can be considered as a simple
deep learning method. Through deep forest, the features are extracted into more
representative features, which further increase the accuracy of the classification.
Furthermore, in order to develop deep forest that is more suitable for HSI clas-
sification tasks, we have added a voting mechanism in its framework to get a
significant classification result. Finally, we have used some of the most advanced
methods on two popular datasets for comparison experiments, and the result
has indicated that DFCF method works well and outperforms other methods.

Further research has the following points: (1) Further study of the application of deep forest in HSI classification. (2) Reduce the number of training samples and further improve classification accuracy.

References

1. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
2. Chen, Y., Lin, Z., Zhao, X., Wang, G., Yanfeng, G.: Deep learning-based classification of hyperspectral data. IEEE J. Sel. Top. Appl. Earth Observations Remote Sens. **7**(6), 2094–2107 (2014)
3. Clausi, D.A., Ed Jernigan, M.: Designing Gabor filters for optimal texture separability. Pattern Recogn. **33**(11), 1835–1849 (2000)
4. Ghamisi, P., Mura, M.D., Benediktsson, J.A.: A survey on spectral-spatial classification techniques based on attribute profiles. IEEE Trans. Geosci. Remote Sens. **53**(5), 2335–2353 (2015)
5. Hughes, G.: On the mean accuracy of statistical pattern recognizers. IEEE Trans. Inf. Theory **5**(1), 55–63 (1968)
6. Kang, X., Li, S., Benediktsson, J.A.: Spectral-spatial hyperspectral image classification with edge-preserving filtering. IEEE Trans. Geosci. Remote Sens. **52**(5), 2666–2677 (2014)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems, pp. 1097–1105 (2012)
8. Li, J., Marpu, P.R., Plaza, A., Bioucas-Dias, J.M., Benediktsson, J.A.: Generalized composite kernel framework for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **51**(9), 4816–4829 (2013)
9. Li, W., Qian, D.: Gabor-filtering-based nearest regularized subspace for hyperspectral image classification. IEEE J. Sel. Top. Appl. Earth Observations Remote Sens. **7**(4), 1012–1022 (2014)
10. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
11. Pan, B., Shi, Z., Xu, X.: Hierarchical guidance filtering-based ensemble classification for hyperspectral images. IEEE Trans. Geosci. Remote Sens. **55**(7), 4177–4189 (2017)
12. Pan, B., Shi, Z., Xu, X.: R-VCANet: a new deep-learning-based hyperspectral image classification method. IEEE J. Sel. Top. Appl. Earth Observations Remote Sens. **10**(5), 1975–1986 (2017)
13. Pan, B., Shi, Z., An, Z., Jiang, Z., Ma, Y.: A novel spectral-unmixing-based green algae area estimation method for GOCI data. IEEE J. Sel. Top. Appl. Earth Observations Remote Sens. **10**(2), 437–449 (2017)
14. Pan, B., Shi, Z., Zhang, N., Xie, S.: Hyperspectral image classification based on nonlinear spectral-spatial network. IEEE Geosci. Remote Sens. Lett. **13**(12), 1782–1786 (2016)

15. Xu, X., Shi, Z.: Multi-objective based spectral unmixing for hyperspectral images. ISPRS J. Photogrammetry Remote Sens. **124**, 54–69 (2017)
16. Zhong, Z., Fan, B., Ding, K., Li, H., Xiang, S., Pan, C.: Efficient multiple feature fusion with hashing for hyperspectral imagery classification: a comparative study. IEEE Trans. Geosci. Remote Sens. **54**(8), 4461–4478 (2016)
17. Zhou, Z.H: Ensemble Methods: Foundations and Algorithms. Taylor and Francis (2012)
18. Zhou, Z.H., Feng, J.: Deep forest: Towards an alternative to deep neural networks (2017)