3D Convolutional Neural Networks with Image Fusion for Hyperspectral Image Classification

Cheng Shi(✉), Jie Zhang, Zhenzhen You, and Zhiyong Lv

School of Computer Science and Engineering, Xi'an University of Technology,
Xi'an, China
Chengc_s@163.com

Abstract. Image fusion can extract redundant information of multiple images into one image, and the goal of image fusion is to better apply to classification tasks. Convolutional neural networks have proved to be an effective way for accuracy of classification. However, fusion and classification usually considered separately. In this paper, we design a 'fusion-classification networks', and introduce image fusion technology and 3D convolutional neural networks (3D CNNs) into HSI classification. In the proposed method, the fusion process is guided by the classification result, and the classification accuracy is improved by the fusion process. Image fusion technology is performed on spectral bands to exploit the redundancy information of HSI, and 3D CNNs are applied on the fused image to extract more robust spectral-spatial features. The proposed method is tested on two datasets. Its outstanding performance is validated in comparison with other state-of-the-art approaches.

Keywords: Hyperspectral image classification · 3D convolutional neural networks · Image fusion

1 Introduction

Hyperspectral image (HSI) has hundreds of spectral bands for each image pixel. This rich information can provide useful information for image classification and recognition [1]. Furthermore, the increasing spatial resolution makes it more vital to analyze the spatial information. HSI classification has developed a lot in remote sensing image analysis. How to learn the effective features from the rich spatial and spectral information is a challenging task.

The typical feature extraction method is subspace-projection method, such as principal component analysis (PCA) [2], independent component analysis (ICA) [3], and linear discriminant analysis [4]. Moreover, wavelet transform has been investigated into the hyperspectral image classification [5, 6]. The limitation of these methods is that they can extract the spectral information only. But the spatial information is neglected, which also plays an important part in HSI classification. Through the combination of the spatial and spectral information, some spatial-spectral classification methods were proposed, such as Extended Multi-Attribute Profiles (EMAPs) [7], active learning [8], decision fusion [9, 10], edge-preserving filtering [11, 12].

Q. Liu et al. (Eds.): CENet 2020, AISC 1274, pp. 419–428, 2021.
https://doi.org/10.1007/978-981-15-8462-6_47

Recently, deep neural networks are demonstrated a useful tool for image feature extraction. It can learn more complex features from the image than the traditional neural networks (less than three layers). Now the commonly used deep neural networks include stacked autoencoder networks (SAEs) [13], Deep Belief Networks (DBNs) [14], deep Boltzmann machines (DBMs) [15], and deep convolutional neural networks (CNNs) [16–18]. In [19, 20], SAEs and DBNs were firstly introduced into the HSI classification problem, and achieved ideal results. Among all of these networks, CNNs are the two-dimension (2D) neural networks, which make it more suitable for exploiting the deep structure information of the image. Because of the superior performance on 2D image, CNNs have been introduced into handwritten characters recognition [17], document image classification [16, 18], and HSI classification [21, 22]. In [21], CNNs were introduced into the HSI classification in spectral domain. The convolutional and subsampling processes were considered as the one-dimension operator. In [22], the spatial and spectral information was extracted separately to form some two-dimension (2D) feature maps. And then the CNNs were trained based on each feature map to get useful high-level features. Due to the high spectral dimension of HSI, [23] and [24] proposed HSI classification method by combining dimension reduction method and CNNs.

Traditional CNNs are designed to process the 2D data. So, dealing with the 3D data, we should convert it to the 2D model firstly. HSI is a 3D cube, the spatial and spectral information are all useful for classification. And it is believed that the spectral and spatial information should be considered simultaneously to further improve the classification accuracy [25]. To capture the effective space features, some 3D feature learning method was proposed, such as 3D Gabor wavelet [25] and 3D DWT [26] based HSI classification methods. These methods have proved to obtain higher classification accuracy, but they have a problem that 3D features extraction method need to determine the number of features, and it will bring large number features which need to be selected. The feature extraction process and the feature selection process have great influence on the classification accuracy. Deep neural networks can adaptively learn the features of the image, and therefore, some researchers extended the 2D CNNs into 3D mode and applied on video data, 3-D object recognition, and HSI classification. 3D convolutional neural networks (3D CNNs) can automatically learn the features from low-level to high-level, and the feature extraction and the feature selection are fused to learn the appropriate features. Hence 3D CNNs can learn more discriminative features.

To inherit the advantages of 3D CNNs, we propose 3D convolutional neural networks with image fusion (3D CNNs_IF) model for HSI classification in this paper. 3D CNNs_IF consists of three parts: band fusion, 3D feature extraction, and classification. Hence, 3D CNNs_IF can be considered as a ‘fusion-classification networks’. 3D CNNs require high spatial and time complexity, so it is necessary to reduce dimension of the HSI first. The reduced image affects the performance of the following feature extraction. Image fusion is to retain the most useful information of multiple images into one image. According to the idea of image fusion, we design a band fusion layer to reduce the dimension of the HSI, and maintain the important spectral and spatial information in the image after dimension reduction. Band fusion layer is to fusion the adjacent several bands into one, which can remove the noisy pixels and reduce the redundant information of HSI. We note in [11], the combination of band fusion and recursive filtering were proposed for HSI classification. Compared with [11] and other traditional dimension reduction methods, the band fusion layer is connected with the 3D feature extraction layer, and the reduced image can not only reduce the computational complexity of 3D CNNs, but also can extract more effective spectral and spatial information under the guidance of the 3D deep features. Based on the proposed 3D CNNs_IF model, the proposed HSI classification method can capture more spectral and spatial features with less loss of information.

The proposed method is tested on two well-known datasets: Pavia University and Salinas images. Experimental results show that the proposed 3D CNNs_IF method can improve the classification accuracy obviously. This paper is organized in four sections. Section 2 describes the structure of 3D convolutional neural networks with image fusion. Section 3 consists of the experimental results and analysis, and Sect. 4 presents the concluding remarks.

2 The Structure of 3D Convolutional Neural Networks with Image Fusion

In this section, we propose 3D convolutional neural networks with band fusion method for hyperspectral image classification. We first construct the 3D cube sample by taking a 3D neighborhood window on the HSI, and then feed each 3D sample into 3D CNNs_BF model for training. The proposed 3D CNNs_BF model consists of three parts: band fusion, 3D feature extraction, and classification, which is shown in Fig. 1.

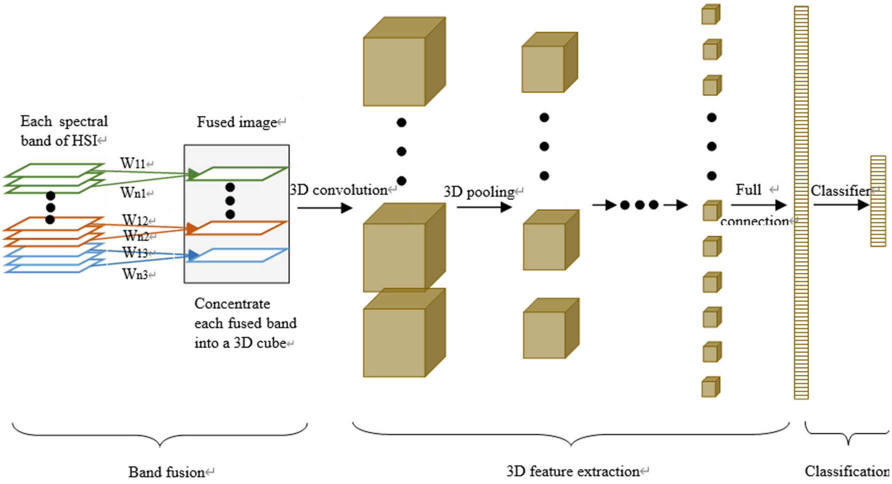


Fig. 1. Framework of 3D convolutional neural networks with image fusion.

2.1 Image Fusion

HSI usually consists hundreds of spectral bands, which contain a lot of redundant information. In order to effectively use the redundant information and reduce the complexity of the subsequent feature extraction, we first generate a band fusion layer to extract valid information.

Firstly, we divided the spectral bands into M groups, and each group contains N spectral bands $(X_1,\ldots,X_N)$. The spectral bands of each group are fused into one band by weighted fusion rule. And then pass the fused band into a nonlinear activation function.

$$\mathbf{F}_j = s(\sum_{i=1}^{N} \mathbf{W}_{ij} X_i) \quad j = 1,\ldots,M \tag{1}$$

where $\mathbf{W}_{ij}$ is the fusion weight connecting $i-th$ input and $j-th$ output, $\mathbf{F}_j$ means the $j-th$ output band, and $s(\bullet)$ is the sigmoid function.

According to Eq. (1), the valid information of multiple bands can be fused into a band through the fusion process. Different from traditional fusion method, the weight $\mathbf{W}_{ij}$ is not a pre-fixed value, but is updated by the supervised back propagation algorithm. Hence the fused image can be more effective on the extraction of redundant spectral information.

Through Eq. (1), we can get M fused spectral bands. And then the obtained bands $\mathbf{F}_j, j = 1,\ldots,M$ are concentrated into a 3D cube for the following process of feature extraction.

2.2 3D Feature Extraction

In Sect. 2.1, the spectral dimension of each sample is reduced and these samples are concentrated into a 3D cube for spectral-spatial feature learning. To extract more discriminative spectral-spatial feature, 3D convolutional neural networks are applied to extract the spectral and spatial information simultaneously.

Similar to 2D CNNs, the forward process for 3D CNNs is also a layer-by-layer based calculation. The features can be obtained according to the feature maps in the previous layer. The calculation process is described following.

Convolution layer: At the convolution layer, the input 3D feature maps are convolved with the 3-D filters and then passed through a nonlinear function to obtain the 3D output feature maps. The process is formulated as Eq. (2).

$$\mathbf{F}_j^l = s(\sum_{i \in M} \mathbf{F}_i^{l-1} * \mathbf{K}_{ij}^l + b_j^l) \quad i = 1,\ldots,p, j = 1,\ldots,q \tag{2}$$

where $\mathbf{F}_j^l$ means the output 3-D feature map in l layer, q is the number of the output feature maps. $\mathbf{F}_i^{l-1}$ is the input 3-D feature map in $l-1$ layer, and p is the number of the input feature maps. $\mathbf{K}_{ij}^l$ is a 3-D filter, and b_j^l is the bias for $j-th$ output feature map.

Pooling layer: The max-pooling method is adopted in this paper. The purpose of the pooling layer is to sub-sample the previous feature map and make it size reduced. The formula is shown in Eq. (3).

$$\mathbf{F}_j^l = s(g(\mathbf{F}_i^{l-1}) + b_j^l) \quad i = 1, \ldots, p, j = 1, \ldots, q \tag{3}$$

where $g(x)$ represents a 3-D sub-sampling operation. The previous feature map F_i^{l-1} is divided into several small 3-D cubes. For each cube, map its maximum value into the output feature map.

Full connection layer: If each feature map in previse layer is a 3-D cube, it should be flatted into a vector first. And then each value of the vector is connected to all the units of full connection layer. The formula is shown in Eq. (4).

$$f^l = s(\mathbf{W}^l g^{l-1} + b^l) \tag{4}$$

where g^{l-1} is the flatted vector in the previous layer, and $\mathbf{W}^l$ is the connection weight matrix.

2.3 Classification

Softmax classifier is used for classification in 3-D convolutional neural networks. The softmax model is shown in Eq. (5).

$$y^L = \frac{1}{\sum_{j=1}^{numclass} e^{w_j^T f^l}} \begin{bmatrix} e^{w_1^T f^l} \\ e^{w_2^T f^l} \\ \vdots \\ e^{w_{numclass}^T f^l} \end{bmatrix} \tag{5}$$

where *numclass* means the class number in the classification problem. The output $\mathbf{y}^L$ is a vector for each sample. The value in vector $\mathbf{y}^L$ represents the probability of all the classes. The position where it has a maximum value in vector $\mathbf{y}^L$ is the label for the sample. The mini-batch gradient descent method is used for training the proposed model.

3 Experimental Results

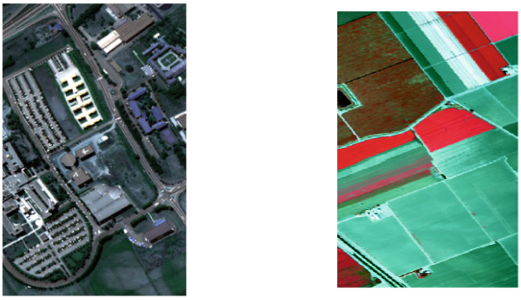
3.1 Datasets

Two different datasets: The Pavia University and Salinas images are used in the following experiments to verify the effectiveness of the proposed 3D_CNN_IF method.

The Pavia University image was recorded by Reflective Optics System Imaging Spectrometer. It has 115 bands with size of 610 × 340. Only 103 bands are used for experiments due to water absorption. Its spectral range is from 0.43 to 0.86 μm. 42776 labeled pixels with 9 classes are contained in the ground truth map.

The Salinas image was recorded by the AVIRIS sensor over Salinas Valley, CA, USA. It has 224 bands with size 512 × 217. Remove water absorption, 204 bands are used for the experiments. Its spectral range is from 0.4 to 2.5 μm. 54156 labeled pixels with 16 classes are from the ground truth map.

In our experiments, the parameter setting for the proposed method is shown as follows: The spatial size of the input sample is 17 × 17, the number of group is set as 20, the first convolution layer has 12 filters with size 4 × 4 × 5, the second convolution layer has 36 filter with size 3 × 3 × 3. Max-pooling is used after the first convolution layer. The full connection layer has 100 units. And the end of the 3-D CNNs is the softmax classifier. The iteration number is set as 500 epochs for training, the learning rate is set as 0.05. 10 independent replications of all the experiments are conducted. The average values of overall accuracy (OA), average accuracy (AA), and kappa coefficient are used to evaluate the classification accuracy. For each replication, 10% of the labeled samples are randomly selected to generate the training set, and the rest samples for testing. The false color image of two datasets are shown in Fig. 2.



(a) (b)

Fig. 2. Experiment datasets (a) False color image of Pavia University; (b) False color image of Salinas.

3.2 Comparisons to Other Methods

In this section, PCA, EMAPs [7], 1-D CNNs [21], SAEs [19], and 3D DWT [26] are used to make a comparison. For EMAPs method, the threshold value of standard deviation attribute is [2.5%, 5%, 7.5%, 10%], and the threshold value of area attribute is set as 200, 500, and 1000. For 3D DWT method, haar wavelet is utilized for feature extraction. The other parameters are set as the fault in their papers.

Figures 3, 4 show the RGB composition map for classification results and Tables. 1–2 show the classification accuries. PCA and 1D CNNs methods only use the spectral information for classification, and EMAPs and SAEs are the spectral-spatial based classification methods. 3D DWT and the proposed method are the 3D based methods, which can capture the spectral and spatial information simultaneously. We can observe from Figs. 3 and 4 that PCA method is sensitive to noise, and the classification maps present the noisy classification results. Although 1D CNNs belong to the spectral-based classification method, but it can capture the high-level features, and suppress the noise in classification map. Obviously, the classification results of spectral-spatial methods are better than the spectral-based method. Some misclassifications of the mixed spectral pixels are reduced in classification map, especially for class 3 of Pavia University and class 8 of Salinas. Of all the compared methods, 3D based methods shows a competitive classification result. Since 3D DWT cannot represent the image adaptively, which affects its performance. The proposed method can capture the high-level features of the HSI, and obtains much higher classification accuracy than other methods do Tables 1 and 2.

Table 1. Classification accuracies of Pavia University.

Classes	PCA	EMAPs	1D CNNs	SAEs	3D DWT	3D CNNs_IF
OA/total	82.06	96.36	93.28	95.18	95.99	97.77
AA	76.31	95.10	90.38	93.18	94.28	97.13
Kappa	74.96	95.17	91.04	93.59	94.67	97.04

Table 2. Classification accuracies of Salinas.

Classes	PCA	EMAPs	1D CNNs	SAEs	3D DWT	3D CNNs_IF
OA/total	96.95	97.45	91.73	92.01	94.25	97.97
AA	70.86	97.52	94.68	96.56	97.54	98.15
Kappa	74.19	97.16	90.77	91.10	93.60	97.75

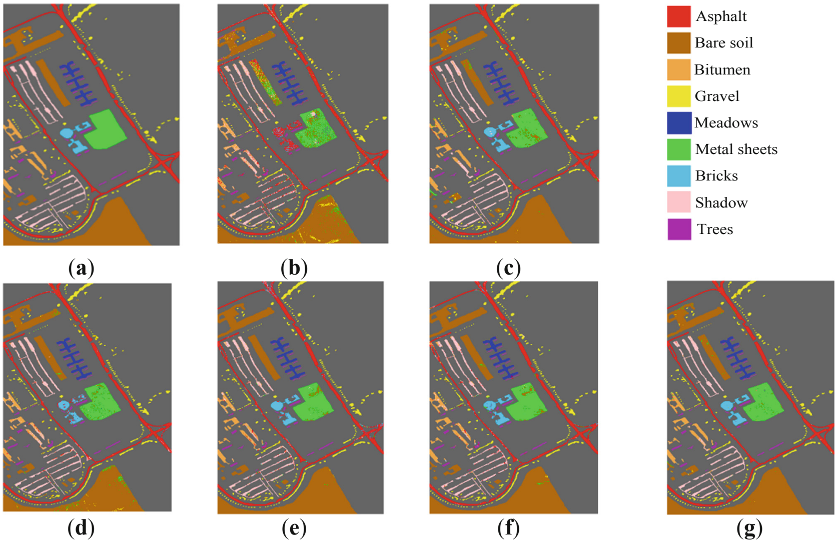


Fig. 3. Classification maps for Pavia University dataset; (a) Ground truth map; (b) PCA; (c) EMAPs [7]; (d) 1D CNNs [21]; (e) SAEs [19]; (f) 3D DWT [26]; (g) 3D CNNs_IF

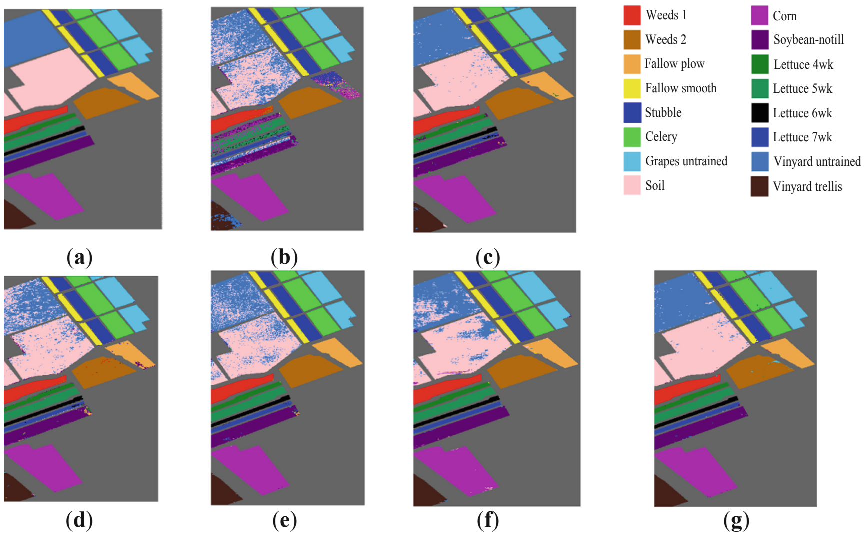


Fig. 4. Classification maps for Pavia University dataset; (a) Ground truth map; (b) PCA; (c) EMAPs [7]; (d) 1D CNNs [21]; (e) SAEs [19]; (f) 3D DWT [26]; (g) 3D CNNs_IF.

4 Conclusions

In this paper, a novel framework for HSI classification via 3D CNNs_IF is proposed. Image fusion technology is applied on spectral band of HSI to exploit the effective information, and 3D CNNs are used for exploiting the deep features of HSI. In our method, the "fusion-classification" networks can obtain more robustness and discrimination features, and the classification accuracy is improved obviously. For hyperspectral images, reducing spectral redundancy is very important for classification. In our further work, we will study more effective "fusion-classification" networks to further improve the classification accuracy.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (61902313, 61701396, 61702409), Natural Science Foundation of Shaanxi Province [2020JQ-652], Research Project of Shaanxi Provincial Education Department.

References

1. Fauvel, M., Tarabalka, Y., Benediktsson, J.A., Chanussot, J., Tilton, J.C.: Advances in spectral-spatial classification of hyperspectral images. Proc. IEEE **101**, 652–675 (2013)
2. Prasad, S., Mann Bruce, L.: Limitations of principal components analysis for hyperspectral target recognition. IEEE Geosci. Remote Sens. Lett. **5**, 625–629 (2008)
3. Villa, A., Benediktsson, J.A., Chanussot, J.E.F.: Hyperspectral image classification with independent component discriminant analysis. IEEE Trans. Geosci. Remote Sens. **49**, 4865–4876 (2011)

4. Li, W., Prasad, S., Fowler, J.E.E.F.: Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. IEEE Trans. Geosci. Remote Sens. **50**, 1185–1198 (2012)
5. Schmidt, F., Doute, S., Schmitt, B.: Wavanglet: an efficient supervised classifier for hyperspectral images. IEEE Trans. Geosci. Remote Sens. **45**, 1374–1385 (2007)
6. Daamouche, A., Melgani, F.: Swarm intelligence approach to wavelet design for hyperspectral image classification. IEEE Geosci. Remote Sens. Lett. **6**, 825–829 (2009)
7. Dalla Mura, M., Benediktsson, J.A., Waske, B.E.F.: Morphological attribute profiles for the analysis of very high resolution images. IEEE Trans. Geosci. Remote Sens. **48**, 3747–3762 (2010)
8. Jun, L., Bioucas-Dias, J.M., Plaza, A.: Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning. IEEE Trans. Geosci. Remote Sens. **51**, 844–856 (2013)
9. Li, J., Huang, X., Gamba, P.E.F.: Multiple feature learning for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **53**, 1592–1606 (2015)
10. Tuia, D., Flamary, R., Courty, D.: Multiclass feature learning for hyperspectral image classification: Sparse and hierarchical solutions. J. Photogrammetry Remote Sens. **105**, 272–285 (2015)
11. Kang, X.D., Li, S.T., Benediktsson, J.A.: Feature extraction of hyperspectral images with image fusion and recursive filtering. IEEE Trans. Geosci. Remote Sens. **52**, 3742–3752 (2014)
12. Kang, X.D., Li, S.T., Benediktsson, J.A.: Spectral-spatial hyperspectral image classification with edge-preserving filtering. IEEE Trans. Geosci. Remote Sens. **52**, 2666–2677 (2014)
13. Bengio, Y., Lamblin, P., Popovici, D.E.F.: Greedy layer-wise training of deep networks. In: Proceedings of the Twentieth Conference on Neural Information Processing Systems. Advances in Neural Information Processing Systems 19, Vancouver, British Columbia, Canada. DBLP, pp. 153–160, December 2007
14. Hinton, G.E., Osindero, S., Teh, Y.: A fast learning algorithm for deep belief nets. Neural Comput. **18**, 1527–1554 (2006)
15. Salakhutdinov, R., Hinton, G.E.: Deep Boltzmann machines. J. Mach. Learn. Res. **5**, 1967–2006 (2009)
16. Lecun, Y.L., Bottou, L., Bengio, Y.E.F.: Gradient-based learning applied to document recognition. Proc. IEEE **86**, 2278–2324 (1998)
17. Lecun, Y.L., Kavukcuoglu, K., Farabet, C.: Convolutional networks and applications in vision. In: Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 253–256 (2010)
18. Kang, L., Kumar, J., Ye, P.E.F.: Convolutional neural networks for document image classification. In: IEEE 22nd International Conference on Pattern Recognition (ICPR), pp. 3168–3172 (2014)
19. Chen, Y., Lin, Z., Zhao, X.: Deep learning-based classification of hyperspectral data. IEEE J. Sel. Top. Appl. Earth Observations Remote Sens. **7**, 2094–2107 (2014)
20. Chen, Y., Zhao, X., Jia, X.: Spectral-spatial classification of hyperspectral data based on deep belief network. IEEE J. Sel. Top. Appl. Earth Observations Remote Sens. **8**, 1–12 (2015)
21. Hu, W., Huang, Y.Y., Wei, L.E.F.: Deep convolutional neural networks for hyperspectral image classification. J. Sens. **6**, 468–477 (2015)
22. Yue, J., Zhao, W.Z., Mao, S.J.E.F.: Spectral-spatial classification of hyperspectral images using deep convolutional neural networks. Remote Sens. Lett. **6**, 468–477 (2015)

23. Zhao, W.Z., Du, S.H.: Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. IEEE Trans. Geosci. Remote Sens. **4**, 4544–4554 (2016)
24. Zhao, W.Z., Du, S.H.: Learning multiscale and deep representations for classifying remotely sensed imagery. J. Photogrammetry Remote Sens. **113**, 155–165 (2016)
25. Shen, L., Zhu, Z.: Discriminative Gabor feature selection for hyperspectral image classification. IEEE Geosci. Remote Sens. Lett. **10**, 29–33 (2013)
26. Qian, Y., Ye, M.: Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. IEEE Trans. Geosci. Remote Sens. **51**, 2276–2291 (2013)