Scene classification of remote sensing image based on deep convolutional neural network

Zhou Yang, Xiao-dong Mu, Feng-an Zhao

Zhou Yang, Xiao-dong Mu, Feng-an Zhao, "Scene classification of remote sensing image based on deep convolutional neural network ," Proc. SPIE 10806, Tenth International Conference on Digital Image Processing (ICDIP 2018), 108064V (9 August 2018); doi: 10.1117/12.2502942

Event: Tenth International Conference on Digital Image Processing (ICDIP 2018), 2018, Shanghai, China

Scene Classification of Remote Sensing Image Based on Deep Convolutional Neural Network

Zhou Yang *, Xiao-dong Mu, Feng-an Zhao
Xi'an High Technology Research Institute, Xi'an, Shaanxi, China 710025

ABSTRACT

Aiming at low precision of remote sensing image scene classification, a classification method DCNN_FF based on deep convolutional neural network (DCNN) feature fusion (FF) is proposed. This method utilizes the existing pre-trained network models CaffeNet and GoogLeNet, and extracts the features of the classified remote sensing images by fine tuning on the target dataset. After dimension reduction by principle component analysis (PCA), the features extracted from the two network models are combined. Finally, the support vector machine (SVM) is used for classification of the combined features. The experimental results on the commonly used and latest datasets show that, this method can utilize the existing network models and combine with the structural advantages of different models, and its average classification accuracy is higher than that of single network model by more than 1.68%. Thus it improves the accuracy of remote sensing image scene classification.

Keywords: remote sensing image; scene classification; deep convolutional neural network; principle component analysis; support vector machine

1. INTRODUCTION

Scene classification of remote sensing images plays an important role in many fields, such as land management, urban planning, environmental exploration and monitoring, and natural disaster detection. Over the past decades, researchers have done a great deal of experiments in the scene classification for satellites and aerial photographs, and have developed many taxonomies[1]. However, with the continuous improvement of science and technology, the spatial resolution of remote sensing images is getting higher and higher, and their spatial and structural patterns are becoming more and more abundant. The phenomena of the same objects with different spectrum and the foreign ones with common spectrum are more common. But the most of the classical methods are based on artificial or shallow learning algorithms, and the low middle-level semantic features extracted are limited in the description ability, which makes it difficult to improve the classification accuracy further.

In recent years, the method of deep learning, as the most advanced technique in computer vision recognition[11], has been successfully applied to many recognition problems and obtained great improvements over the past, such as object detection[13], face and speech recognition[14], behavior recognition[16], semantic segmentation[17], natural image classification[18] and remote sensing scene classification[19], etc. Many recent papers[23] show that it is an effective method to apply a deep convolutional neural network(DCNN) model pre-trained on a large dataset (such as ImageNet) to a remote sensing image dataset which is relatively small. Among them, the best result is obtained by fine tuning the pre-trained CNN on a specific remote sensing dataset and using it as a feature generator of the support vector machine when classification. However, in the above classification methods, it is just the comparison of the CNN models with different strategies or the comparison between different CNN models. It is seldom considered that different CNN models should be combined and they can complement each other with different structure advantages. In the meantime, although many remote sensing datasets have been selected for testing in various methods, and UC Merced Land-Use is one of the most widely used datasets and has achieved good performance. But there are still some common and obvious deficiencies. For example, the number of scene categories, the number of images in all kinds of scenes, and the total number of labeled images are relatively few, and the differences and diversity of scenes are both small. This limits the improvements of the classification methods to some extent, especially the ones' based on convolutional network model, because it requires a lot of tagged data for training.

*yzmailbox2015@163.com; phone 18710493241;

Tenth International Conference on Digital Image Processing (ICDIP 2018), edited by Xudong Jiang, Jenq-Neng Hwang
Proc. of SPIE Vol. 10806, 108064V · © 2018 SPIE · CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2502942

In view of the above problems, a classification method DCNN_FF based on DCNN and feature fusion(FF) is proposed in this paper, which is used to improve the accuracy of remote sensing image scene classification. Its structure is shown in Figure 1. A comparative study was conducted on a more informative NWPU-RESISC45 dataset[27] to analyze the effect of dataset size on different classification methods and classification accuracy.

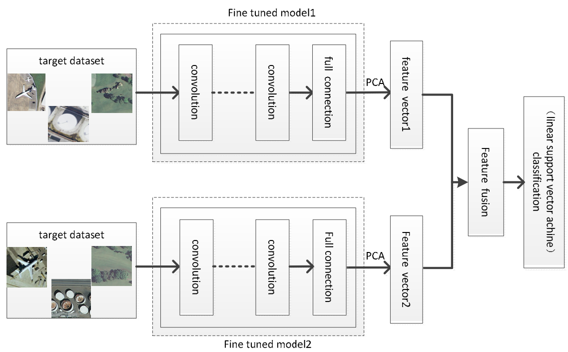

Figure 1. The structure of the proposed method

2. CNNS

Compared with the traditional classification methods, CNNs have achieved quite good performance in recent remote sensing scene classification[28], because CNNs, especially deep CNNs, can extract the different level features from images, including the abstract low-level features (point, line, angle, etc.) in the first few layers, the middle-level features (semantic information) and the later high-level features (category information). Thus they effectively link the low-level abstract features with the high-level semantic tags, which ensures the accuracy of classification.

There are three types of deep learning methods for remote sensing image scene classification: one is to train a new model of CNN[31]; another is to use the pre-trained CNN as a feature exporter and then classify them with other classifiers (eg SVM). The third one is to fine tune already trained CNNs[30]. Through the comparison of the experiments on different datasets by the researchers, the first method is not good as the latter two ones because of the high computational cost, easy over-fitting and the need for a large amount of training data. And the third one gets the best classification.

However, most of the above researches focus on the fine tuning or partial structure adjustment and improvement of the existing CNN models, as well as the comparison of different network models with conventional shallow methods in remote sensing image scene classification, and the structural complementarity of different CNN models is not realized. For example, some models focus on the reasonable collocation of convolutional layers and pooling layers in the structures and the selection of appropriate convolutional kernel sizes, but the model parameters are more (such as CaffeNet[35]); while some models reduce the parameters by cutting down the network depth in the unit modules, which increases the overall depth of the networks (eg GoogLeNet[36]). And the features extracted by these different network structures are bound to vary. If the different level features are fused, the abstraction ability of the features will be enriched, which is useful for improving the scene classification of the remote sensing images.

3. METHOD

In this paper, a classification method based on DCNN_FF is proposed, in which two used CNN models are CaffeNet and GoogLeNet with complementary advantages and good classification effects. After fine tuned on remote sensing datasets, the two models are taken as feature extractors, then the feature vectors obtained by them are fused and classified as the input of SVM.

CaffeNet, the CNN model used in this paper, is obtained by Caffe. Among them, Caffe is one of the most popular libraries for deep learning, especially for CNNs. It is the research and development by Berkeley Visual Learning Center (BVLC), and improved by other researchers. Caffe can be customized with profiles and expanded with new layer types easily, and C ++, Python and MATLAB programmatic interfaces are provided to quickly get a new convolutional network model. Figure 2 shows the network architecture of CaffeNet. And paper [37] can be referred to for details. It consists of 5 convolutional layers with a pooling one behind each of them, and three fully connected layers in final. At last, a 4096-dimensional eigenvector is obtained.

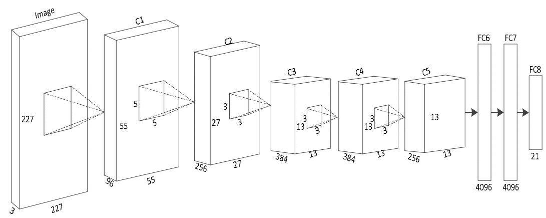

Figure 2. CaffeNet structure (The outer box represents the size and feature number of each feature layer, and the square represents the size of the convolutional kernel. There are five convolutional layers and three full connections)

The GoogLeNet proposed in paper [29] is the second deep model used in this paper. It is the CNN structure that wins the ILSVRC14 competition. Its main feature is the use of "initial modules" (Figure 3) and is inspired by the "net nesting" idea of [38] and the theoretical results of [39]. In short, the initial module reduces the complexity of convolutional filtering through the initial phase of network depth reduction, allowing multiple filters to be used in parallel for filtering at different resolutions. The initial module has two main advantages: First, more spatial information is retained at each layer by using filters with different sizes; second, it significantly reduces the number of free parameters in the network and makes it difficult to overfit, which can increase the depth of the network structure, and it is very important for the classification effect[40]. In the 22-layer GoogLeNet structure, there are up to 50 convolutional layers but the total parameters are much less than those of the 8-layer CaffeNet structure. The detailed structure and related parameters of the GoogLeNet can be referred to [29], and finally a 1024-dimensional eigenvector can be obtained.

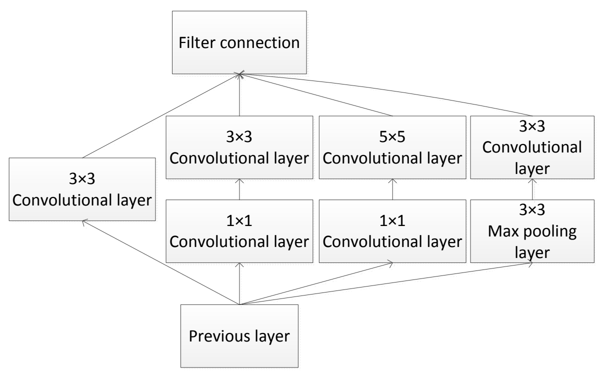

Figure 3. The inception module in GoogLeNet (convolutional kernels of different sizes can get features at different spatial scales. They are then aggregated and fed to the next layer. The 1x1 convolutional kernel is used for dimension reduction before the 3x3 and 5x5 convolutional ones)

Firstly, the two models of CaffeNet and GoogLeNet are respectively trained and fine tuned on the remote sensing dataset which will be classified, and the training data are extended by using the method mentioned in reference [41] while training, that is, the five parts of four corners and the middle are segmented from the image with the horizontal inversion together, and a dataset with ten times of the original images can be obtained, which alleviates the shortcoming of fewer labeled images in the remote sensing dataset.

In [33], it is pointed out that the first fully connected layer has better expression than the latter ones. So this paper selects the feature extracted from the first fully connected layer in CaffeNet as the feature output of the model. Because of the characteristics of the GoogLeNet strcture, we choose its final feature output which are more representative. In order to reduce the dimensionality of fusion features, improve the overall efficiency and reduce the over-fitting, we first reduce the output dimensions of the two models by principal component analysis(PCA) (as shown in Equation 1).

$$X_1 = \Theta_1 X_0 \tag{1}$$

Among them, X_0 is the eigenvectors of the model output, Θ_1 represents the coefficient matrix, and X_1 is the feature vector after dimension reduction. Then the two eigenvectors are fused. Finally, the fusion features are classified by SVM.

4. RESULTS AND ANALYSIS

4.1 Test datasets and parameter settings

The experimental dataset 1 is UC Merced Land-Use dataset, which includes 21 types of remote sensing images consist of aircraft, forest and expressway, etc. And each type contains 100 images of 256 × 256 pixels with a spatial resolution of 0.3 meter per pixel. Figure 4 shows an example of various kinds of images.

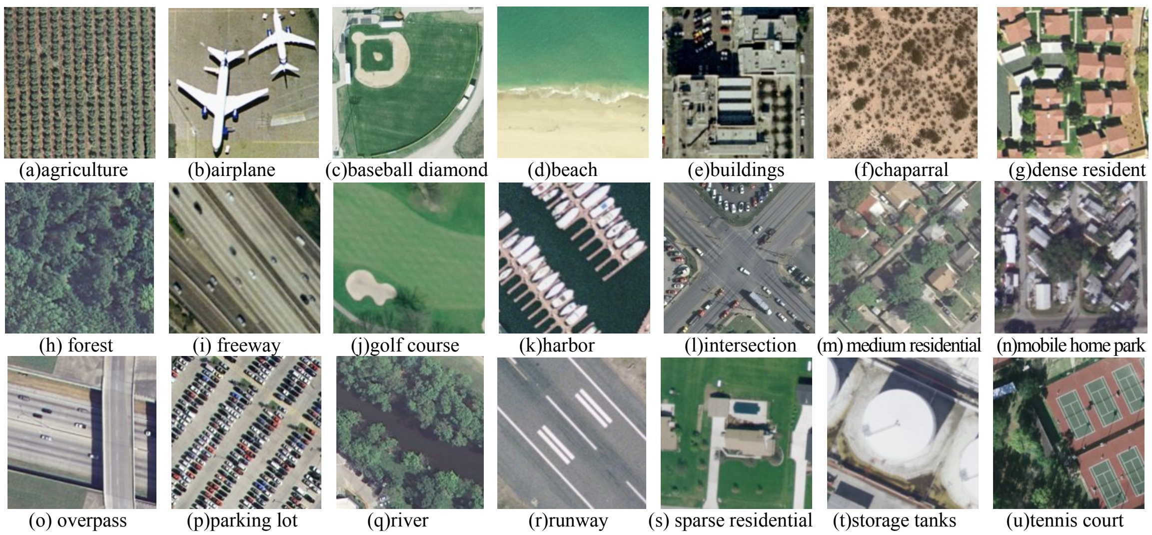

Figure 4. Sample charts of UC Merced Land-Use dataset

The experimental dataset 2 is NWPU-RESISC45 dataset (Remote Sensing Image Scene Classification_Benchmark and State of the Art), and it is a new dataset including 45 scene categories which is composed of 30 widely used scenarios selected from various datasets by Cheng Gong, Han Junwei, etc, and 15 scene categories selected from Science Net and Sogou academic website with church, desert, basketball court, etc. Each category contains 700 images with 256 × 256 pixels, and most of the images have a spatial resolution from 0.2 to 30 meters per pixel. Figure 5 shows an example of a partial category images.

Figure 5. Some sample charts of NWPU-RESISC45 dataset

In the experiment, we take the average of 100 test results as the final classification result(training data and test data are randomly selected for each experiment), and use the public source library Caffe and VLFeat[42] to extract the features of CNNs. The pre-trained models used in this paper can be found in the Caffe model library[43], in which the methods in paper [44] are used to fine-tune the two models. Although a variety of fusion methods have been proposed in literatures, the direct combination method is used for feature fusion in order to simplify the processing, and the linear kernel function is used as the one of SVM classifier, and training and testing were performed in the public LIBLINEAR library[46]. After many tests, it is best for classification to reduce the eigenvectors of the two models to 512 dimensions and then perform fusion.

4.2 Analysis for the classification ability

When testing in the UCM Land-Use dataset, 80 ones were randomly selected as the training data and the remaining were used as the test data in each type of images. In order to verify the improvement of the proposed method, we compare it with other classification methods and the improved strategies of the two models respectively. And the results of the experiment are shown in Table 1. As can be seen from the test results in Table 1, no matter what kind of CNN model was used, the fine tuned model got the better classification result than the one used as a feature extractor with the SVM for classification, and the best classification results were got by combination of the two strategies. At the same time, it can be seen that the GoogLeNet model with a deeper network structure is more advantageous than the CaffeNet model with a "shallower" structure due to the limited number of images in the dataset, regardless of the improvement strategies. Finally, the proposed method achieves better classification results than both of the two model structures.

Table 1 classification accuracy of UCM Land-Use dataset by different methods and their improved strategies

Algorithm	Improved strategy	Classification accuracy
SIFT+SC[4]		81.67
FK-S[47]		91.63
VLAD[48]		92.50
VLAT[48]		94.30
OverFeat[19]		90.91
MS-DCNN[49]		91.34
CaffeNet[34]		93.42
	feature extractor[44]	94.28
	fine tuned[44]	95.12
	fine tuned+feature extractor	95.69
GoogLeNet	feature extractor[44]	94.38
	fine tuned[44]	96.48
	fine tuned+feature extractor	96.85
DCNN_FF	feature extractor	95.89
	fine tuned	97.97
	fine tuned+feature extractor	98.53

The classification recognition rate of 21 remote sensing scenes is shown in Figure 6 using this method, and the classification recognition rates of the other two methods (that is, the pre-trained models are used as feature extractors after being fine tuned) are also listed for contrast. From the test results, in the different scenes classification, the accuracy of the ones with small differences in the vein is generally high, such as forest, farmland, runway, beach, etc; while the accuracy of complex scenes containing multiple objects is generally low, such as building, mobile home park, tennis court, etc; Because GoogLeNet uses different size filters to retain more accurate spatial information in its initial module, its classification effect is better than Caffenet on gulf, intersection, rivers and other scenes which relationship of the target locations is clear. The proposed method achieves a good performance in all kinds of scenes classification because of combining the advantages of two model structures.

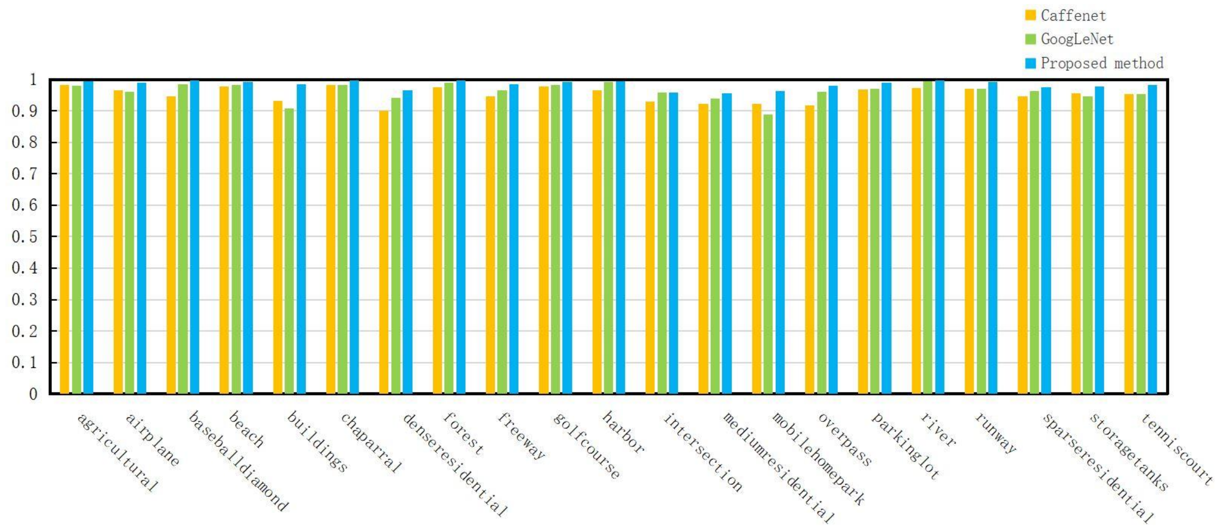

Figure 6. Classification accuracy of the 21 classes remote sensing scenes

4.3 The impact analysis of different datasets on the classification methods and the accuracy

In order to compare the impact of different datasets on the classification methods and accuracy, NWPU-RESISC45 dataset is used as the second experimental dataset. Because of its variety of images and the abundance of similar scenarios, the intra-class diversities and the inter-class similarities are greatly enhanced, which provides more space for the testing of the classification algorithms and reduces the deficiencies of previous datasets such as over-fitting due to a small number of images.

In this experiment, 600 images of each type were randomly selected as the training data and the remaining 100 ones were used as the test data. The classification accuracy and confusion matrix are shown in Table 2 and Figure 7, respectively.

Table 2 classification accuracy of different methods and their improved strategies on NWPU-RESISC45 dataset

Algorithm	Improved strategy	Classification accuracy
CaffeNet	feature extractor	94.55
	fine tuned	95.43
	fine tuned+ feature extractor	96.10
GoogLeNet	feature extractor	94.69
	fine tuned	96.71
	fine tuned+ feature extractor	97.18
DCNN_FF	feature extractor	96.41
	fine tuned	98.26
	fine tuned+ feature extractor	98.97

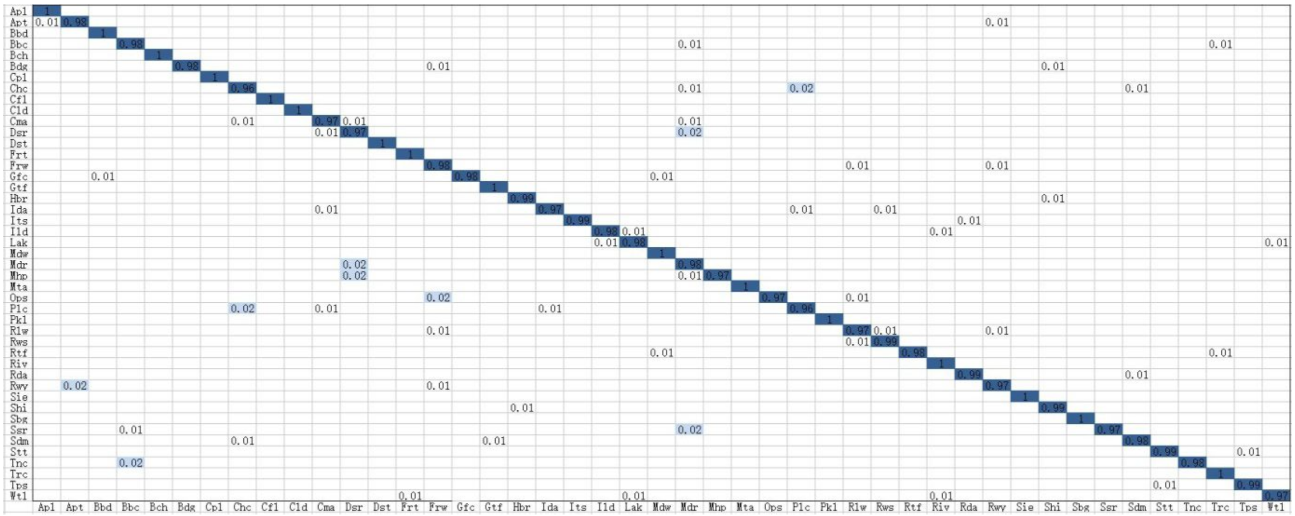

Figure 7. The confusion matrix of the scene classification on NWPU-RESISC45 dataset

As can be seen from the test results in Table 2, the method still yields the best classification accuracy which is similar to the one in Table 1, except that the classification accuracy of the NWPU-RESISC45 dataset is higher than that of the UCM Land-Use one. This is because the abundant images in the NWPU-RESISC45 dataset are useful for training the deep CNNs sufficiently, which taps the potential advantages of deep networks. It is also reflected in the improvement of the correctness of different classification methods on the two datasets. And the GoogLeNet with a deeper network level is better than the relatively "shallow" Caffenet.

It can be seen from Fig. 7 that the accuracy and the misclassification rate of each scene are similar to those in Fig. 6 when using this method. That is the recognition accuracy of the objects is higher which contain a single scene, such as forest, snowberg and grassland, etc. However, the misclassification rate is higher when the scenes are with large differences in the texture, or including more objects, such as the church, the dense residential and the railway, etc. And they are easily misclassified especially for ambiguous scenes such as the church and the palace, the dense residential and the mobile home park as well as the tennis court and the basketball court.

5. CONCLUSION

The method proposed in this paper used pre-trained CaffeNet and GoogLeNet to extract the features of remote sensing images by fine tuning on the target dataset. And the features extracted from two models were fused after dimension reduction, finally, the SVM classifier was used to classify the scenes. The experimental results in the datasets UCM Land-Use used commonly and NWPU-RESISC45 with more reference values, show that the classification accuracy is improved at least 1.68% by the proposed method compared with the single network models. And it is possible to compute in parallel and reduce the entire time complexity because of the parallel network structure, so it can be predicted that the running time of the proposed method will not be much more than that of using a single network model when the hardware conditions are permitted, and it’s the study work in the next step. Therefore, it is an effective way to improve the classification accuracy of remote sensing images by combining the pre-trained network models with different structures, which can not only develop their own structural features but also complement each other’s advantages. At the same time, The NWPU-RESISC45 dataset also provides a good data support for the further study of remote sensing image methods.

REFERENCES

[1] Cheng G, Han J, Guo L, et al. “Effective and efficient midlevel visual elements-oriented land-use classification using VHR remote sensing images,” IEEE Transactions on Geoscience and Remote Sensing, 53(8): 4238-4249(2015).
[2] Cheng G, Han J, Zhou P, et al. “Multi-class geospatial object detection and geographic image classification based

on collection of part detectors," ISPRS Journal of Photogrammetry and Remote Sensing, 98: 119-132(2014).
[3] Cheng G, Zhou P, Han J, et al. "Auto-encoder-based shared mid-level visual dictionary learning for scene classification using very high resolution remote sensing images," IET Computer Vision, 9(5): 639-647(2015).
[4] Cheriyadat A M. "Unsupervised feature learning for aerial scene classification,"IEEE Transactions on Geoscience and Remote Sensing, 52(1): 439-451(2014).
[5] Yang Y, Newsam S. "Spatial pyramid co-occurrence for image classification," Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE: 1465-1472(2011).
[6] Yang Y, Newsam S. "Bag-of-visual-words and spatial extensions for land-use classification," Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. ACM, 270-279(2010).
[7] Hu F, Xia G S, Wang Z, et al. "Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(5)(2015).
[8] Luo B, Zhang L. "Robust autodual morphological profiles for the classification of high-resolution satellite images,"IEEE Transactions on Geoscience and Remote Sensing, 52(2): 1451-1462(2014).
[9] Pacifici F, Chini M, Emery W J. "A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification," Remote Sensing of Environment, 113(6): 1276-1292(2009).
[10] Huang X, Zhang L. "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery,"IEEE transactions on geoscience and remote sensing, 51(1): 257-272(2013).
[11] Liu C, Hong L, Chen J, et al. "Fusion of pixel-based and multi-scale region-based features for the classification of high-resolution remote sensing image" Journal of Remote Sensing. 19(2): 228-239(2015).
[12] Schmidhuber J. "Deep learning in neural networks: An overview," Neural networks, 61: 85-117(2015).
[13] Zhang L, Zhang L, Du B. "Deep learning for remote sensing data: A technical tutorial on the state of the art," IEEE Geoscience and Remote Sensing Magazine, 4(2): 22-40(2016).
[14] S. Balaban, "Deep learning and face recognition: The state of theart," In Proc. SPIE Int. Soc. Opt. Eng., Baltimore, MD, USA, Art. no. 94570B(2015).
[15] Abdel-Hamid O, Mohamed A, Jiang H, et al. "Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition," Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE, 4277-4280(2012).
[16] Karpathy A, Toderici G, Shetty S, et al. "Large-scale video classification with convolutional neural networks," Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 1725-1732(2014).
[17] Zheng S, Jayasumana S, Romera-Paredes B, et al. "Conditional random fields as recurrent neural networks,"Proceedings of the IEEE International Conference on Computer Vision, 1529-1537(2015).
[18] He K, Zhang X, Ren S, et al. "Deep residual learning for image recognition" Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778(2016).
[19] Penatti O A B, Nogueira K, dos Santos J A. "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 44-51(2015).
[20] Midhun M E, Nair S R, Prabhakar V T, et al. "Deep model for classification of hyperspectral image using restricted boltzmann machine," Proceedings of the 2014 International Conference on Interdisciplinary Advances in Applied Computing. ACM, 35(2014).
[21] Chen Y, Zhao X, Jia X. "Spectral - spatial classification of hyperspectral data based on deep belief network," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(6): 2381-2392(2015).
[22] Li T, Zhang J, Zhang Y. "Classification of hyperspectral image based on deep belief networks," Image Processing (ICIP), 2014 IEEE International Conference on. IEEE, 5132-5136(2014).
[23] Sharif Razavian A, Azizpour H, Sullivan J, et al. "CNN features off-the-shelf: an astounding baseline for recognition," Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 806-813(2014).
[24] Oquab M, Bottou L, Laptev I, et al. "Learning and transferring mid-level image representations using convolutional neural networks," Proceedings of the IEEE conference on computer vision and pattern recognition, 1717-1724(2014).
[25] Donahue J, Jia Y, Vinyals O, et al. "Decaf: A deep convolutional activation feature for generic visual recognition," International conference on machine learning, 647-655(2014).

[26] Zeiler M D, Fergus R. "Visualizing and understanding convolutional networks," European conference on computer vision. Springer, Cham, 818-833(2014).
[27] Li H, Tao C, Wu Z, et al. "RSI-CB: A Large Scale Remote Sensing Image Classification Benchmark via Crowdsource Data," arXiv preprint arXiv:1705.10450(2017).
[28] Vakalopoulou M, Karantzalos K, Komodakis N, et al. "Building detection in very high resolution multispectral data with deep learning features," Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International. IEEE, 1873-1876(2015).
[29] Szegedy C, Liu W, Jia Y, et al. "Going deeper with convolutions," Proceedings of the IEEE conference on computer vision and pattern recognition, 1-9(2015).
[30] Nogueira K, Penatti O A B, dos Santos J A. "Towards better exploiting convolutional neural networks for remote sensing scene classification," Pattern Recognition, 61: 539-556(2017).
[31] Zhang F, Du B, Zhang L. "Scene classification via a gradient boosting random convolutional network framework," IEEE Transactions on Geoscience and Remote Sensing, 54(3): 1793-1802(2016).
[32] Zhao W, Du S. "Scene classification using multi-scale deeply described visual words,"International Journal of Remote Sensing, 37(17): 4119-4131(2016).
[33] Hu F, Xia G S, Hu J, et al. "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," Remote Sensing, 7(11): 14680-14707(2015).
[34] Othman E, Bazi Y, Alajlan N, et al. "Using convolutional features and a sparse autoencoder for land-use scene classification," International Journal of Remote Sensing, 37(10): 2149-2167(2016).
[35] Lowe D G. "Distinctive image features from scale-invariant keypoints" International journal of computer vision, 60(2): 91-110(2004).
[36] Dalal N, Triggs B. "Histograms of oriented gradients for human detection," Computer Vision and Pattern Recognition, CVPR 2005. IEEE Computer Society Conference on. IEEE, 1: 886-893(2005).
[37] Jia Y, Shelhamer E, Donahue J, et al. "Caffe: Convolutional architecture for fast feature embedding," Proceedings of the 22nd ACM international conference on Multimedia. ACM, 675-678(2014).
[38] Lin M, Chen Q, Yan S. "Network in network," arXiv preprint arXiv:1312.4400(2013).
[39] Arora S, Bhaskara A, Ge R, et al. "Provable bounds for learning some deep representations," International Conference on Machine Learning, 584-592(2014).
[40] Simonyan K, Vedaldi A, Zisserman A. "Learning local feature descriptors using convex optimization," IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(8): 1573-1585(2014).
[41] Krizhevsky A, Sutskever I, Hinton G E. "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, 1097-1105(2012).
[42] Vedaldi A, Fulkerson B. "VLFeat: An open and portable library of computer vision algorithms," Proceedings of the 18th ACM international conference on Multimedia. ACM, 1469-1472(2010).
[43] Caffe Model Zoo. Available online: https://github.com/BVLC/caffe/wiki/Model-Zoo(14 August 2015).
[44] Castelluccio M, Poggi G, Sansone C, et al. "Land use classification in remote sensing images by convolutional neural networks," arXiv preprint arXiv:1508.00092(2015).
[45] Sivic, J.; Zisserman, "A. Video Google: A text retrieval approach to object matching in videos," Processings of the IEEE International Conference on Computer Vision. Nice, France, 1470 − 1477(2003).
[46] Fan R E, Chang K W, Hsieh C J, et al. "LIBLINEAR: A library for large linear classification," Journal of machine learning research, 9(Aug): 1871-1874(2008) .
[47] Zhao B, Zhong Y, Zhang L, et al. "The Fisher kernel coding framework for high spatial resolution scene classification," Remote Sensing, 8(2): 157(2016).
[48] Negrel R, Picard D, Gosselin P H. "Evaluation of second-order visual features for land-use classification," Content-Based Multimedia Indexing (CBMI), 2014 12th International Workshop on. IEEE, 1-5(2014).
[49] XU Suhui, MU Xiaodong, ZHAO Peng, et al. "Scene Classification of Remote Sensing Image Based on Multi-scale Feature and Deep Neural Network," Acta Geodaetica et Cartographica Sinica, 45(7): 834-840(2016).