Research on High Resolution Remote Sensing Image Classification Based on Convolution Neural Network

Wenwen Gong, Zhuqing Wang, Yong Liang, Xin Fan, Junmeng Hao

Research on high resolution remote sensing image classification based on convolution neural network

Wenwen Gong [1], Zhuqing Wang [1], Yong Liang [1(✉)], Xin Fan [2], Junmeng Hao [3]

[1]School of Information Science and Engineering, Shandong Agricultural University, Tai’ an 271018, China
[2]Tai’ an City intelligence research institute of science and technology, Tai’ an 271000, China
[3]School of information science and technology, Tai Shan University, Tai’ an 271018, China

Abstract: Traditional classification method based on machine learning algorithm has been widely adopted in very high resolution remote sensing image classification, yet the problem that could not effectively convey a higher level of abstract feature still need to be improved. This paper, relying on the convolution neural network algorithm, has conducted on the high-resolution remote sensing image classification method. Firstly, structure of convolution neural networks was analyzed. The prediction model of convolution neural networks was discussed, and the core of structure was the alternation of the convolution layer and the down sampling layer. Then, the training model of convolution neural networks was researched. By using weights sharing and local connection, convolution neural network, that image could directly entered into, avoids to a certain extent caused by image displacement, dimension change and so on. On this basis, basing on different phase GF-1 remote sensing data and MATLAB development environment under Windows10 operating system, then combining with object-oriented classification technology in image segmentation, this paper built the high resolution remote sensing image classification model based on convolution neural network. Finally, the parameters of the model were tested and analyzed repeatedly, and more accurate model parameters were obtained in this paper. Results show that the mode can effectively improve the classification accuracy, and provide technical support for improving remote sensing image interpretation and formulating sustainable development strategy.

Keywords: high resolution data; convolution neural network; abstract features; image classification

1 Introduction

Remote sensing image classification, which is one of the important ways of remote

Brief introduction of the author: Wenwen Gong(1993-), female, Graduate student, research into computer technology and application;

Date of reception: 2017-09-25;

Foundation item: State Key Research and Development Projects (The Integration of Multi-source Disaster Information and Multi-scale Spatial Information, 2016YFC0803104); National Natural Science Foundation of China (Study on Alimentarn Crop Ecological Compensation System Based on Carbon Sink Function, 71503148);

Corresponding author: Yong Liang, professor, E-mail: yongl@sdau.edu.cn.

sensing image analysis and interpretation, is a concrete application of pattern recognition in the field of remote sensing. Meanwhile, remote sensing image classification has been of great important value in some fields such as dynamic monitoring of geography national conditions, precision agriculture, urban management, target identification and accurate acquisition of remote sensing thematic information[1]. In recent years, with the continuous progress of remote sensing technology, the successful launch of all kinds of resource satellites and our high score series satellites has provided strong technical support for monitoring the dynamic defense of the earth's surface. The resolution of remote sensing images is getting higher and higher, and the spectral information of objects is also becoming more and more abundant. The full use of these features can effectively improve the classification quality [2]. Because the traditional classification methods of spectral statistical characteristics, such as maximum likelihood method and K nearest neighbor domain, only use spectral information of the image and not fully utilized rich details in the image, leading to lower classification accuracy. In order to solve this problem, a lot of space geometric information has been integrated into the process of high resolution remote sensing image classification; at the same time, Machine Learning (ML) algorithms are introduced, such as Support Vector Machine (SVM) [3], Random Forest(RF)[4], and Neural Networks(NN)[5]. Compared with traditional classification methods, multiple feature classification based ML has obtained higher accuracy and better classification results. Thus, the core of image classification is remote sensing image feature learning and feature extraction. However, because the traditional classification method based on spectral characteristics and multi-feature classification method based on ML algorithm belong to shallow learning algorithm, which only could extract the lower level features, such as the spectrum, texture and shape, and couldn’t effectively express the abstract characteristics of higher level. In addition, these methods are difficult to cross the semantic gap between low-level and high-level, and no longer applicable in current big data environment.

At present, Deep Learning (DL), as the most vigorous field of machine learning domain, comes from the well-informed research of neural network, which aims to construct a multilevel neural network through simulating the process of brain learning characteristics. By accepting data from the underlying layer (input layer), this technique extracts feature from bottom to top layer by layer, which could not only excavate the regularities of the data in time, but also set up complex mapping relations between low-level features and high-level semantics, so as to improve classification accuracy[6]. Especially, the depth learning algorithm represented by convolution neural networks (CNN), with the help of the convolution operation hierarchy, is more conducive to extract the features of two-dimensional images. Up to now, the convolution neural network model trained by depth learning algorithm has achieved remarkable results in the field of computer vision, etc. However, there are still relatively few researches on the classification of high resolution remote sensing images, and the related technologies are still not mature enough.

2 The Discussion of Convolution Neural Network Structure

Abroad, the research of convolution neural network started relatively earlier, and the whole development process can be generalized to the following three stages: the stage of principle forming, the stage of establishing a model and the stage of in-depth exploration. In 1962, Hubel launched a thorough research on cat visual cells, and then put forward for the first time a receptive field [7]. 2006, Geoffrey Hinton and his students published a paper on the top academic journals [8] concerning deep learning, mainly clarified that a deeper network model with powerful data representation capability and " layer-wise pre-training " achieved by unsupervised training in order to the optimal training. Subsequently, many different models of convolution neural network structure have been proposed. For example, Oxford’s VGG (Visual Geometry Group) [9], Google’s GoogLeNet[10], Microsoft’s ResNet [11], etc, these models were created remarkable achievements in the ILSVRC. In domestic, the convolution neural network developed rapidly although it started relatively late. In 2013, Baidu, as the world's largest Chinese search engine, took the lead in the launch of deep learning related research work and the creation of Baidu Deep Learning Institute. Microsoft, Alibaba, Tencent, IBM, as a high-tech companies with big data, have invested a great deal of resources to promote the theoretical study of DL. At the same time, with the use of GPU developed by NVIDIA, these corporations promoted that deep learning was widely used in the domestic expansion [12].

CNN, a typical deep learning algorithm, mainly includes two parts, which are prediction model and training model respectively. And then CNN establishes a mapping from input to output after a large number of data is used for supervised training [13].

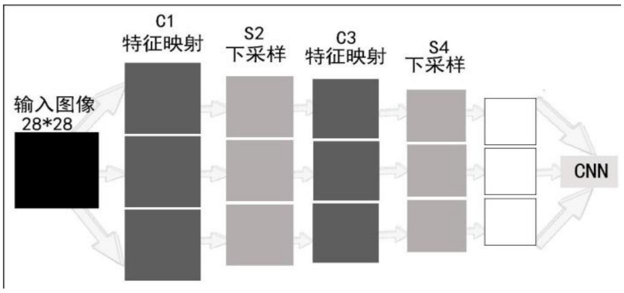


Fig.1 Schematic diagram of CNN

Classical convolution neural network is mainly composed of three types of neural network layer and an output layer, including convolution layer, down-sampling layer (pool operation), full connection layer, in which the convolution layer and the down-sampling layer alternately appear, as shown in Figure 1. Each neural network layer contains multiple two-dimensional planes, and each two-dimensional plane includes arbitrary number of independent neurons. Every single neuron in a two-dimensional plane links neural cell within local receptive field range of front layer, and which extracts partial features by convolution operation.

2.1 Forward Propagation

The process that applying CNN to the classification of high resolution remote sensing images can be divided into two stages, the forward propagation phase and the backward propagation phase. The prediction model of CNN can be regarded as a forward propagation course. In which, the output of front layer is inputted into the next layer(current layer), meanwhile the incentive value obtained by the activation function is transmitted step by step to the next network layer. In other words, a sample (X, Yp) is selected randomly from the input data set, where X represents data, and Y represents labels. The sample data X is fed into the network, which is passed results calculated by each hidden layer from the input layer to the output layer for the sake of O, the actual operation output. The computations performed by the network can be represented by the following formula:

$$\mathrm{O} = \mathrm{F_n}(\ldots(\mathrm{F_2}(\mathrm{F_1}(\mathrm{XW_1})\mathrm{W_2})\ldots)\mathrm{Wn}) \qquad (1)$$

The output of the current neural network layer can be represented as:

$$\mathrm{X^l} = \mathrm{f}(\mathrm{W^l X^{l-1}} + \mathrm{b^l}) \qquad (2)$$

Among them, $\mathrm{F_i}$,i=1,2,…,n represents a linear transformation, and $\mathrm{W_i}$, i=1,2,…,n represents the weight parameters that have been trained, l=1,2,…,n, represents the number of layers, $\mathrm{b^l}$ is the bias of the current layer, f is the activation function, in which the common activation functions have the sigmoid function, the tanh function and the ReLu(Rectified Linear Units, ReLu) function.

When the network is processed and output the value of O, the network quality can be estimated by (Y, O), and the network of the ideal case satisfies Y==O.

The kernel structure of convolution neural network model is the convolution layer and the down sampling layer. The invariance of translation and deformation on this model is guaranteed by combining the local receptive field and weight sharing.

Convolutional layer. For extracting a variety of different characteristics, convolution, as a common algorithm in image recognition, selects different convolution kernel to conduct convolution operation with previous feature map in the process of feature extraction in convolution layer. Generally speaking, each convolution kernel is a square that is m * m, the m is the size of convolution kernels. The greater the value is, the better the effect of abstraction is, but more parameters are needed to train. On the contrary, the smaller the value is, the more refined the image can be, but more layers of convolution are needed to implement it. The results obtained by convolution kernel develop into the activation value after activation function computes, encouraging the formation of corresponding feature map of the next level and the increase of feature map number by each layer, which makes CNN getting more in-depth information. The schematic diagram of the convolution operation is shown in figure 2.

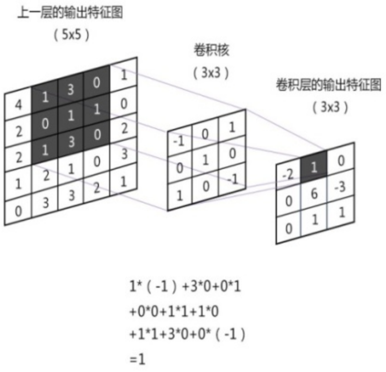


Fig.2 Schematic diagram of convolution operation

The convolution layer and the down sampling layer alternately appear as C1 and C3, and the following formulas are usually used for calculation:

$$X_j^l = \sigma(\sum_{m_j} X^{l-1} * \mathrm{kernelj}^l + b^l) \qquad (3)$$

In formula, l=1,2,…,n represents the number of network layer; j=1,2,…,n, m_j is an option for an input feature map on the previous layer; kernel represents the convolution kernel; b^l is the shared bias for each layer.

For example, an input image given, assuming the resolution of 3*96*96, is trained to get one hundred feature maps, the size of 3*11*11.These pictures adopt 100 different convolution kernels, in accordance with the size of 11*11 for convolution in the original image, which eventually will be able to get 100 feature maps of 86*86.

Lower-sampling layer. Each pixel in the input image processed by convolution layer and ReLu activation function contained a little of information in around area, which would cause information redundancy. The enormous cost of computational resources and the severe over-fitting problem would be cut down for the displacement robustness by pooling operation, which was realized by integrating the characteristics of each part of the image in different locations to reduce the amount of data. Common pooling methods are mean pooling, max pooling and stochastic pooling, etc [14]. As shown in Figure 3, a window with a size of 2*2 is used to perform averaging pooling operations on the misaligned sub-regions of any image after the convolution operation, which can obtain processed feature maps. Thus, the amount of data is significantly reduced.

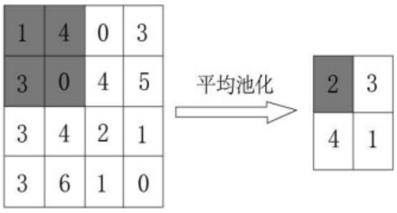


Fig.3 Schematic diagram of the maximum pool layer

The calculation form of the sample layer is shown in equation (4).

$$X_j^l = \sigma(g_{m_j}(X^{i-1}) + b^l) \quad （4）$$

In formula: g(z) represents the sub sampling function defined on the data Z, which can be defined as the average value, the maximum value or the random value of the sub region, b^l is the bias, and σ is the activation function; j=1, 2, … , n, m_j represents an option for the input feature map.

Fully connected layer, through the adoption of the structure that the neurons in the current network layer interconnect with the neurons in former layers but no connection between each neuron in the layer, realizes the conversion of the image feature vector from two-dimensions to one-dimensions that have higher order invariance. These one-dimensional could be fed to the output layer of network and classified by classifier. Ultimately, this completes a learning curse of forward propagation, and results are outputted to the outside users by the output layer.

2.2 Backward Propagation

The training model of convolution neural network is a process of backward propagation. In the last layer of model, the error function (loss function) is obtained through comparing the predicted results of forward propagation with the objective function (the labels of training samples). Error updating values are calculated further and transmitted back to layer by layer. The error function can be expressed as:

$$E_p = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{c}(y_k^n - O_k^n)^2 \quad (5)$$

In formula, k=1,2, … ,n, y_k^n denotes the kth dimensions of labels that the nth samples corresponds to;O_k^n signify the kth output of network prediction that the nth samples corresponds to; When carries on the back propagation, the following formula was used ：

$$\delta^l = (W^{l+1})^T\delta^{l+1} * f(\acute{u^l}),u^l = W^l x^{l-1}+b^l \quad (6)$$

In which, l=1,2 , … ,n, δ^l represents the error function of the previous layer; δ^{l+1} is on behalf of the error function of the current layer (equivalent to the E_p in the output layer); W^{l+1} is the weight matrix; u^l represents the output of the previous layer.

After the error back-propagation, the error function of each network layer is obtained. Then, the convolution parameters and offsets are renovated through adopting Stochastic Gradient Descent algorithm, moreover are iterated and adjusted repeatedly, which stop iteration until the network reaches the convergence condition to make training model optimization [15].

3 Design of Classification Method Based on Convolution Neural Network

In order to verify the validity and accuracy of the convolution neural network classification method, basing on different phase panchromatic and multi-spectral data of GF-1 and MATLAB development environment under Windows10 operating system, then combining with object-oriented classification technology in image segmentation, this paper establishes the model based on Deep learning toolbox through programming, which is used for classification of various objects in images.

3.1 Construction of Experimental Data Set

An experimental study area in Beijing acquired at March 24, 2016 is selected, including buildings, roads, water areas, overpasses and other types of objects. Before selecting training samples, some preprocessing operations on raw data, such as radiometric calibration, FLAASH atmospheric correction and ortho-rectification, are carried out, including the selection of ground control points (GCP), establishing of correction model and image resampling. Then, the panchromatic and multispectral data of pre-processed images are fused to improve the accuracy of the classification. The fused remote sensing image is shown in figure 4.

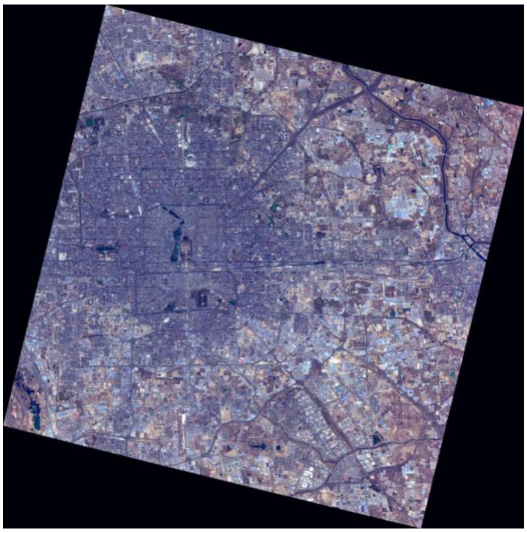
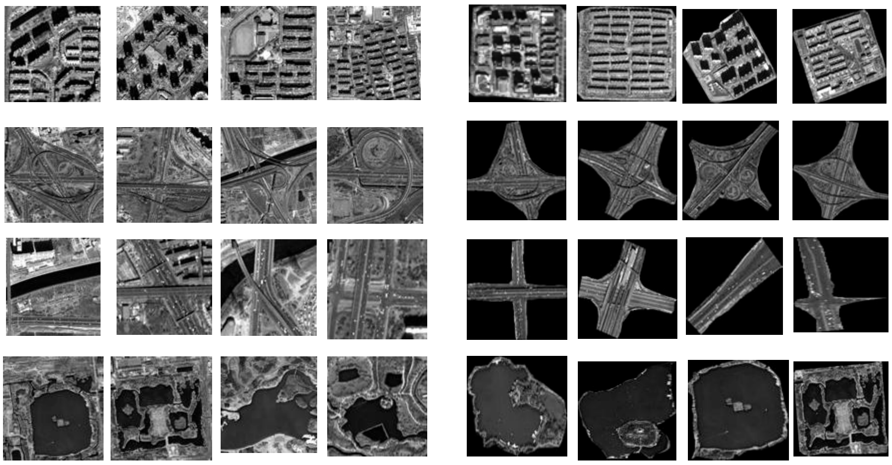


Fig.4 Fused image

For color image fused, the sample data is collected, imported and saved, and then used to train the network by applying the object-oriented classification technique and "the largest number of objects" used in image segmentation in this experiment. Finally, the best model for remote sensing image classification is obtained. The specific process is as follows.

Sampling of remote sensing images by means of two kinds of artificial visual methods, this test selected uniformly and randomly from the fused image for each object about 1200 images, and totaled up to 4800 images. In which, 4000 images were used as the training samples and 800 images were used as test samples. In order to validate the effect of the algorithm effectively, the test samples were not completely equal to the training samples. Each sample image that was uniformly normalized (100 × 100 pixels) × 3 channels, was unified gray level by MATLAB software and generated as the mat file, a image data format. Then, the label matrix corresponding to the image data was defined, and also saved as a mat file. For example, the first category of training image was the buildings, then the first column of label matrix was 1000; the second category was the road image that use 0100 to express, and so on. Eventually, the training image, the test image and its labels were saved as train_data.mat and test_data.mat files respectively, which were inputted to the CNN. Simultaneously, in the case of other same operations, the sample images obtained in two different ways were used as input samples to compare the results. A part of the training samples were, respectively, indicated in Figure 5. In which, figure 5(a) showed the ground object image without a clear boundary by means of MATLAB software, figure 5(b) indicated the ground object image with a clear boundary by means of ENVI software.



(a)Partially sample sets that don't have specific boundary

(b)Partially sample sets that have specific boundary

Fig.5 Partial training sample set

3.2 Design of Network Structure

3.2.1 Selection of Network Parameters

In view of the influence arising from the number of network layers and the size of convolution kernel in the CNN architecture on the training process and classification results of the network model, repeated experiments were carried out to elect the learning rate of network, the number of network layers and the size of convolution kernel when the average error rate was minimum, which could obtain the optimal network parameters to improve the accuracy of object recognition. For example, the effect of learning rate was to control the scope of each weight adjustment. If the value was too large, the learning rate was faster, but easy to cause divergence. By contraries, if the value was too small, convergence could be considerable but slower.

The network with a size of 11 * 11 convolution kernels, requiring a relatively small number of network layers but more training parameters, could abstract images effectively, which gave rise to an appearance that the overall performance of the algorithm is low[16]. The higher the levels of the network were, the higher the accuracy was usually, however at the same time, the performance of the algorithm would be limited to some extent. Therefore, in order to improve the performance of the algorithm, this study employed the model that possessed the convolution kernel with the size of 5×5, the down-sampling area with the size of 2×2. At the end of the model, the traditional full connection layer and Softmax regression of neural network were adopted. From the experimental results, a conclusion could be drew that the network structure could not only extract the eigenvalues of the image, but also avoid the problem caused by the excessive training parameters and too long training time, which could effectively reduce the error rate of the classification.

3.2.2 Selection of Activation Function

Comparing ReLu with traditional activation function, ReLu function was able to be sparse expression, more in line with biological point of view. Traditional sigmoid and tanh function belonged to the saturation nonlinear activation function, its training speed was slower than that of the Relu function, an unsaturated and nonlinear activation function. It was concluded that ReLu could reduce training time and improve algorithm performance. Therefore, the nonlinear ReLu function was used as the activation function after each convolution layer. The value of activation was obtained to continue forward propagation.

4 Experimental Results and Analysis

4.1 The Index of Classification Accuracy

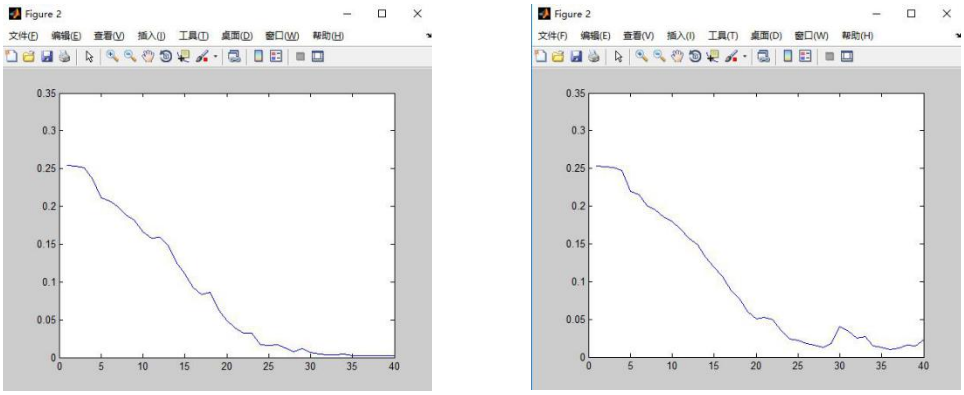
At the end of the classification, the error rate of classification was obtained by comparing the classification results with the sample labels, that is to say, the error in classification:

$$\text{the error in classification} = \frac{\text{The number of samples that have been misjudged}}{\text{the total of samples}}$$

In order to evaluate the quality of network trained, the mean square error (MSE) was chosen to evaluate the quality of the network in this experiment. MSE, as a convenient way to measure "average error", was used to evaluate the degree of data change. Generally speaking, the smaller the MSE was, the better the quality of network was. In the meantime, the better the quality of the network and the accuracy of the prediction model was. Nevertheless, the quality of the network is poor. In the course of training network, the each training value of cost function was preserved. Finally, the process that the MSE change with the increase of training times is intuitively displayed by the curve graph.

4.2 The Analysis of Experimental Results

In the process of classification of high resolution remote sensing image, due to the diversity and complexity of the convolution neural network was used for the extraction and learning of image feature. Through a series of experimental comparative analysis, a multi-class classifier was trained combining label data in the output layer of the network, basing on determining the optimal network parameters. Through the verification of test samples, the error rate of image classification is shown in figure 6. There is a linear negative correlation between the number of experimental iterations and the error rate of image classification. The results of color images differed from grayscale images when CNN is used to extract features. Fig. 6 (a) is the result of a grayscale image, and Fig. 6 (b) is the result of a color image. As a whole, the effect of gray image is better. The error rate of network training gradually converges to 0.002 with the number of iterations, and the classification accuracy reaches 99.8%. This is because that the texture features of the image have a large proportion of in the process of training. At the same time, it fully shows that CNN has powerful representation ability in dealing with task of the image texture feature classification.



（a）Classification error curve of image that don’t have specific boundary

（b）Classification error curve of image that have specific boundary

Fig.6 CNN error rate convergence curve

5 Concluding remarks

Considering the existing classification methods in the expression of complex functions would have certain limitations. This study mainly discussed how to apply the convolution neural network algorithm in high resolution remote sensing image containing a large amount of information data in detail. Basing on the effective algorithm and deep learning toolbox in MATLAB software, the CNN model for high resolution remote sensing image classification was built. Furthermore, this study analyzed and compared repeatedly the selection of parameters in the model and the gray of the input sample image, considering the object-oriented classification technology and the relationship between the objects and pixels. Results showed that the model could effectively improve the classification accuracy and possess broad prospects in the research. Therefore, this method was able to provide the strong technical support for the analysis and interpretation of remote sensing image and sustainable development strategies.

References

1. Lü Q, Dou Y, Niu X, et al. Remote sensing image classification based on DBN model [J]. Journal of Computer Research & Development, 2014, 51(9): 1911-1918(2014)
2. Guo Yubao, Chi Tianhe, Peng Ling, et al. Urban land classification of high resolution remote sensing image using random forests [J]. Bulletin of Surveying and Mapping, 2016(5):73-76(2016)
3. Yang Changkun, Wang Chongchang, Zhang Dingkai, et al. Classification of high resolution satellite images based on SVM [J]. Mapping and spatial geographic information, 2015, (9):142-144(2015)
4. Liu Haijuan, Zhang Ting, Shi Hao, et al. Classification and evaluation of high resolution remote sensing images based on RF model [J]. Journal of Nanjing Forestry University: Natural Science Edition, 2015, 39(1): 99-103(2015)
5. Lu Jingguo. Research on remote sensing image classification and modeling based on Neural Network Ensemble [J]. Bulletin of Surveying and Mapping, 2014(3):17-20(2014)
6. Liu Dawei, Han Ling, Han Xiaoyong. Classification of high resolution remote sensing images based on depth learning [J]. Journal of Optics, 2016, 36(4):1-8(2016)
7. Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex [J]. Journal of Physiology, 1962, 160(1):106(1962)
8. Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks [J]. Science, 2006, 313(5786):504-7(2006)
9. Simonyan K, Zisserman A. Very Deep Convolution Networks for Large-Scale Image Recognition [J]. Computer Science(2014)
10. Mollahosseini A, Chan D, Mahoor M H, et al. Going Deeper with Convolutions[J]. 2014:1-9(2014)
11. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition [J]. Computer Science (2015)
12. Zhao Yongke. Deep learning -21 days, actual combat Caffe[M]. Beijing: Publishing House of Electronics Industry. 2016.7 (2016)
13. Liu Jianwei, Liu Yuan, Luo Xionglin. Advances in depth learning [J]. Application research of

computer, 2014, 31(7):1921-1930(2014)

14. Wang Zhen, Gao Maoting. Design and implementation of image recognition algorithm based on convolution neural network [J]. Modern computers: popular edition, 2015(20):61-66(2015)
15. Zheng-Ping H U, Chen J L, Wang M, et al. Recent progress on convolutional neural network in pattern recognition [J]. Journal of Yanshan University (2015)
16. Nair V, Hinton G E. Rectified Linear Units Improve Restricted Boltzmann Machines[C]// International Conference on Machine Learning. DBLP, 2010:807-814(2010