2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics

Criminal Investigation Image Classification Based on Spatial CNN Features and ELM

DaXiang Li[1,2], Xin Qiu[1*], ZhiYu Zhu[1], Ying Liu[1,2]
1. Center for Image and Information Processing, Xi'an University of Posts and Telecommunications, Xi'an, China
2. Key Laboratory of Electronic Information Application Technology for Scene Investigation
Ministry of Public Security, Xi'an, China
Email: 35108809@qq.com, 1601210057@stu.xupt.edu.cn

***Abstract*—With the advent of the era of big data, the image of the criminal investigation has been explosively growing. To effectively help police officers classify suspect targets, a new classification method of criminal investigation image that combines spatial convolutional neural network (SCNN) and extreme learning machine (ELM) is proposed. Firstly, the image sub-blocks of different space are obtained by a five-block preprocessing on the criminal investigation image, which is part of the original image. Secondly, different image blocks are sent to a fine-tuned convolutional neural network to extract the features of the criminal investigation image. Finally, a criminal investigation image classification algorithm called SCNN-ELM is proposed in combination with extreme learning machine. The experimental results based on 2800 real criminal investigation image database and the standard Corel 1K image database show that the proposed method has an average accuracy improvement of 7.98% compared with the image classification method based on non-blocking CNN features, and the average classification accuracy is also superior than other similar methods.**

Keywords-spatial convolutional neural network; extreme learning machine; criminal investigation image classification

I. INTRODUCTION

With the advent of the era of big data, the number of images in the image database of the criminal investigation has been explosively increasing. If artificial methods are used to classify the images related to different cases, it is time-consuming and difficult to make mistakes because of visual fatigue. Therefore, using the image classification technology [1] to classify the criminal investigation image automatically can effectively help the criminal police to narrow down the scope of the investigation and improve the efficiency of the case, and it has important significance in the work of "Science and technology Strong police" [2].

Image classification technology usually includes two core links: image feature extraction and classifier training. According to the different methods of image feature extraction, image classification methods can be divided into two categories: *1)* the method based on artificial design feature, *2)* the method of learning feature automatically by machine. The method of artificial design feature is to design the bottom visual features such as color, texture and shape of image artificially, and then combine machine learning to realize image classification. For example: A regional growth method [1] of segmentation pretreatment has been proposed, through the HSI model extracted color features, and then used support vector machine (SVM) to classify the quality of tobacco. Literature [3] has fused the traditional Local Binary Pattern (LBP) and wavelet texture features, and used fuzzy KNN classifier for criminal investigation image scenes classification; the literature [4] combined the bag of words (BOW) and Histogram of Oriented Gradient (HOG) features for image text and scene classification; In literature [5], a novel method has been proposed to reduce the dimensionality of Scale Invariant Feature Transform (SIFT) features by using Principal Component Analysis (PCA) and then visual vocabulary was obtained through the BOW model. Finally, SVM was utilized to classify images. All these methods belong to the category of traditional image classification methods. The disadvantages are: the limitation of feature expression, the difficulty in comprehensive characterization, and the poor classification effect. The method of machine learning features is intended to facilitate the deep learning to automatically obtain the image features to achieve image classification. For example: LeCun [6] et al designed and trained the classic LeNet-5 network for handwriting fonts and computer-printed character classification based on previous research work; Krizhevsky [7] et al, proposed alexnet depth convolution network structure based on the disadvantages of LeNet-5 network, which won the championship in the 2012 imageNet large scale visual recogition challenge(ILSVRC). In the 2014 ILSVRC competition, the 22 layer GoogleNet structure [8], the VGGNet structure [9] and the SPP-Net model [10] were champions, runners-up and third runners, respectively. In the 2015 ILSVRC competition, ResNet [11] won the championship in image classification tasks to solve the problem of feature gradient disappearance and gradient explosion. Subsequently, for the problem of multi-label image classification, the HCP [12] was proposed in the absence of truth-valued training data. Literature [13] designed a method to measure the distance between classes quickly, optimized the convolutional neural network, and the image classification accuracy can reach 82.5% on the Caltech-256 database. Therefore, it has become a hot spot for many researchers to obtain image features through deep learning and then apply to image classification.

In this paper, we propose a new classification method for criminal investigation image, namely spatial convolutional neural network and extreme learning machine (SCNN-ELM).The method consists of three parts, the first part of the

978-1-5386-5836-9/18/$31.00 ©2018 IEEE
DOI 10.1109/IHMSC.2018.10173

image is five-block preprocessing, the second part of the SCNN feature extraction, the third part of the different image features into the ELM classifier for classified label prediction. The rest of the paper is organized as following. In Section II, the SCNN-ELM method is explained in details. Section III introduce the image databases and our experimental results, and Section IV concludes this paper.

II. SCNN-ELM

At present, most of the main image classification methods are performed under the depth framework. In the aspect of feature extraction, it usually extracts only one CNN feature for an image without considering the spatial information of CNN. In the aspect of classifier, SVM is commonly used by many workers. Once a kernel function is selected in SVM, the mapping method is uniquely determined. In the era of big data, SVM can find the best mapping through different kernel functions, and the training speed is slow. The SCNN-ELM method proposed for solving the above problems is shown in Fig.1.

1) image preprocessing

As shown in Fig.2. First, the image is processed in five blocks (upper left, upper right, lower left, lower right, and center dotted area).Through the image of different locations, the image information is more abundant.

(a) S-1 (b) S-4 (c) S-5

Figure2. Five block diagram

2) Extracting SCNN Features

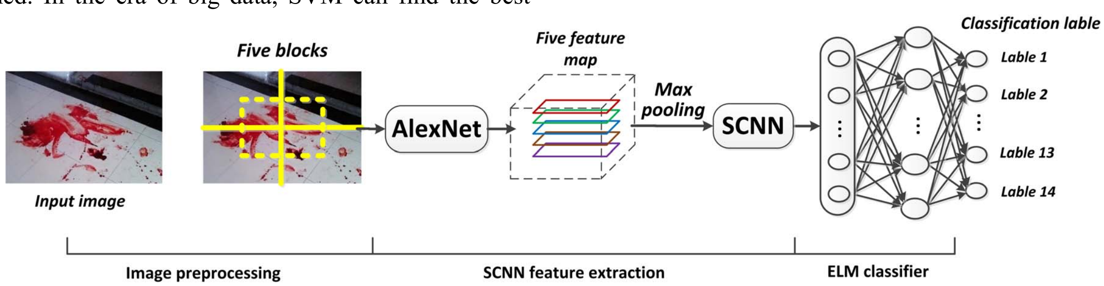

Figure1. SCNN-ELM for Criminal Investigation Image Classification

The sub blocks of each image are input into the AlexNet network respectively, and the 4096 parameters of the FC7 layer are extracted as the CNN features of each block, which are as:

$$\mathrm{CNN_i} = \{x_{i,j} \mid j = 1,...,4096\} \tag{1}$$

Where $i = 1,...,5$ .Finally, the maximum pooling of these 5 CNN features is done to get the spatial CNN features of the images, which are as follows :

$$\mathrm{SCNN} = \{I_j \mid j = 1,...,4096\} \tag{2}$$

Where $I_j = \max(x_{1,j}, x_{2,j},..., x_{5,j})$ represents the maximum value on the first *j* dimension of five CNN features.

3) Classifier design

If $T = \{(I_i, y_i) \mid i = 1,2,\ldots,N; k = 1,\ldots,5\}$ is a training sample, where $I_i = [I_{i,1}, I_{i,2},\ldots, I_{i,5}]^{\mathrm{T}} \in R^5$ is the training data, the label matrix $\boldsymbol{y}_i = [y_{i1}, y_{i2},\cdots, y_{im}]^{\mathrm{T}} \in \mathbf{R}^m$, there are L hidden layer nodes for a single layer feed-forward neural network, and the activation function is $g(\cdot)$, then the ELM classification model for a single hidden layer neural network can be expressed as:

$$\sum_{l=1}^{L} \beta_l g\left(\boldsymbol{C}_l * \boldsymbol{I}_i + b_l\right) = \boldsymbol{y}_i,\ i = 1,2,\cdots,N$$

$$g\left(\boldsymbol{C}_l * \boldsymbol{I}_i + b_l\right) = \frac{1}{1 + e^{-(C_l * I_i + b_i)}} \tag{3}$$

Where $\boldsymbol{\beta}_l = [\beta_{l1}, \beta_{l2}, \cdots, \beta_{lm}]^{\mathrm{T}}$ and $\boldsymbol{C}_l = [c_{l1}, c_{l2}, \cdots, c_{ln}]^{\mathrm{T}}$ are the output and input weights of the l hidden layer node, $\boldsymbol{C}_l * \boldsymbol{I}_i$ is the inner product of $\boldsymbol{C}_l$ & $\boldsymbol{I}_i$, and b_l is the offset of the l hidden layer node bias. Then the (4) model can be written as:

$$\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{Y} \tag{4}$$

$$\boldsymbol{H}(C_1,\cdots,C_L,b_1,\cdots,b_L,I_1,\cdots,I_N) = \begin{bmatrix} g(C_1 * I_1 + b_1) & \cdots & g(C_L * I_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(C_1 * I_N + b_1) & \cdots & g(C_L * I_N + b_L) \end{bmatrix}_{N\times L}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\beta}_L^{\mathrm{T}} \end{bmatrix}_{L\times m} ;\ \boldsymbol{Y} = \begin{bmatrix} y_1^{\mathrm{T}} \\ \vdots \\ y_N^{\mathrm{T}} \end{bmatrix}_{N\times m} \tag{5}$$

Where $\boldsymbol{H}$, $\boldsymbol{\beta}$ and $\boldsymbol{Y}$ are the output matrix of the hidden node bias, output weight and expected output respectively. When the input weight $\boldsymbol{C}_l$ and hidden node offset b_l are given at random and the activation function is sigmoid, the training hidden layer neural network is equivalent to the least square norm solution $\boldsymbol{\beta}$, that is, the output weight of the ELM network is:

$$\boldsymbol{\beta} = \boldsymbol{H}^{+}\boldsymbol{Y} \tag{6}$$

Where $\boldsymbol{H}^{+}$ is denoted as the generalized inverse of $\boldsymbol{H}$. In the training stage, the optimal value can be obtained, so that the model has a good generalization ability.

Assuming that there are m classes in the criminal investigation images, the membership degree of the i sample belongs to the $j \in (1,2,\cdots,\ m)$ class. When $y_{ij} = \{y_{i1}, y_{i2}, \cdots, y_{im}\}$ reaches the maximum value, the i sample belongs to the category the j label:

$$label(y_i) = arg_{j=1,\cdots,m} max\, y_{ij} \tag{7}$$

I. EXPERIMENTS

A. Image database

In order to verify the validity of the SCNN-ELM method in image classification application, the standard image collection Corel 1K database [15] and the real-world criminal investigation image database were selected for experiment. The self-built real criminal investigation image database contains 14 types of images, namely: current blood traces, accident vehicles, current prospect maps, current survey door frames, current survey fingerprints, current survey indoor maps, crime scene maps, interior floor plans, and current survey shoes , Suspected skin, tattoos of suspected suspects, weapon surveys, surveyed tires, and prospecting windows. Among them, 200 color images of each category total 2,800. The real sample image of the criminal investigation image database is shown in Fig.3.

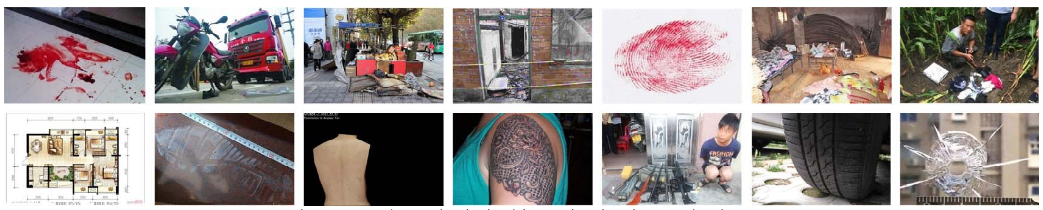

Figure 3. The real criminal investigation image database

B. Comparison of experimental results

1) In order to verify the validity of the Alexnet-finetune network model, the Corel 1K and the criminal investigation 2800 Image databases are compared with the Caffenet, Alexnet and VGG-16 network models respectively, and the experimental results are shown in Table 1.

Table 1 Comparison of classification accuracy of different model

Model	Classification Accuracy (%)	
	Corel 1K	Criminal investigation image
CaffeNet	80.29	74.14
AlexNet	81.78	74.58
VGG-16	87.54	81.63
AlexNet-finetune	84.32	80.94

The experimental results show that the features of the same batch of images obtained by different depth network models are used for image classification. The classification accuracy of CaffeNet, AlexNet, VGG-16 and AlexNet-finetune is gradually increased. VGG-16 is 1 percentage points higher than AlexNet-finetune in classification accuracy, but because of its long time consuming, this paper uses AlexNet-finetune network model for feature extraction.

2) In order to verify which layer of the AlexNet-finetune model is more expressive, the features of each layer respectively were extracted in this part of the experiment, and the experimental results are shown in Fig.4.

The experimental results show that with the increase of the convolution layer, the feature expression image is better, and the feature expression tends to be smooth to the whole connecting layer, but the training time increases with each

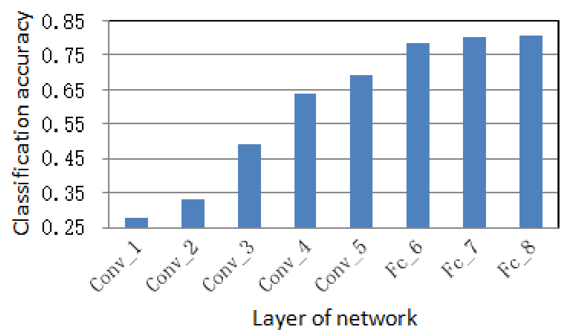

Figure 4. Contrast of classification accuracy for different layers

additional layer, so the FC7 layer data of Alexnet-finetune model is chosen as the feature.

3) In order to verify the validity of the spatial semantic features in this paper, the image is divided into blocks as shown in Fig.2, then the SCNN feature is extracted by the Alexnet-finetune network selected by the experiment, and the image is classified by the SCNN-ELM method, and the experimental results are shown in Table 2.

The experimental results show that by performing feature extraction on different image sub-blocks, the more sub-blocks, the richer the image feature information and the higher the classification accuracy.The classification accuracy of five blocks is nearly 8 percentage points higher than the non-blocks and the classification rate of five blocks is nearly 3 percentage points higher than the 4 blocks.Which is greatly improved.Therefore, this paper uses five blocks to preprocess the image.

TABLE 2 SPATIAL SEMANTIC COMPARISON TEST RESULTS

Method	Classification Accuracy (%)	
	Corel 1K	Criminal investigation image
S-1	77.03	72.96
S-4	82.54	76.88
S-5	84.32	80.94

4) For the SCNN-ELM algorithm, the number of hidden layers is an important parameter that affects the time and precision of the classification. Fig.5 and Fig.6 are the relationship between the number of hidden layers and the classification accuracy and calculation time respectively.

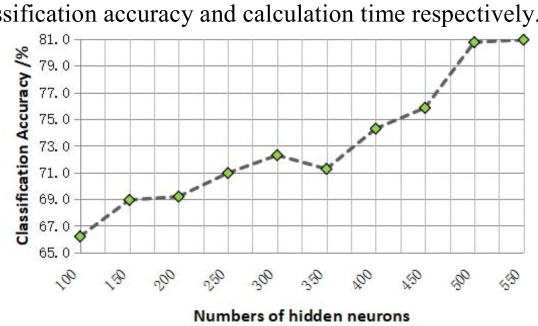

Figure 5. The relation diagram of the number of hidden layers and the classification rate

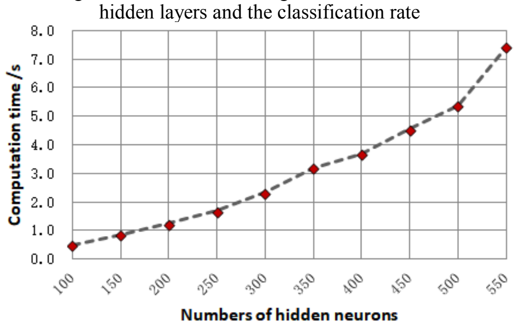

Figure 6. The relation diagram of the number of hidden layers and the computation time

Experimental results are shown that with the increase of the number of hidden layer units, the image classification accuracy fluctuates, but the overall trend is increasing. When the number of neurons reaches 500, the classification accuracy is generally stable, and the computation time then increases rapidly. Therefore, we selects 500 neurons in the hidden layer.

5) The classification method of this paper is compared with other methods, as shown in Table 3.In the above experiments, each class of images is constructed with 60% images, and the remaining 40% is used as the test set. The confusion matrix of 20 repeated experiments based on the SCNN-ELM method in the Corel 1K image database is shown in Table 4 (P_1 to P_{10} respectively represent: Africa, Beach, Building, Bus, Dinosaur, Elephant, Snow Mountain, Flower, Horse, and Food);The confusion matrix of 20 repeated experiments based on the SCNN-ELM algorithm in the self-built real criminal investigation image database is shown in Table 5(C_1 to C_{14} respectively represent: current survey indoor maps, current blood traces, current survey door frames, accident vehicles, current prospect maps, current survey fingerprints, interior floor plans, crime scene maps, current survey shoes, suspected skin, tattoos of suspected suspects, weapon surveys, surveyed tires, and prospecting windows). Where the diagonal represents the percentage of the correct classification of each type of image, the non diagonal represents the percentage of the error classification of each type of image.

TABLE 3 COMPARISON OF EXPERIMENTAL RESULTS BY VARIOUS CLASSIFICATION METHOD (%)

Method	Classification Accuracy (%)	
	Corel 1K	Criminal investigation image
ANN	70.84	67.71
SVM	78.45	75.83
BP	76.37	73.45
Fisher	79.21	74.31
PCA+BP	76.98	72.92
HOG+SVM	81.75	77.84
ELM	79.89	75.63
LBP-ELM	80.43	76.78
SCNN-ELM	84.32	80.94

Experimental results show that the proposed SCNN-ELM method is superior to other method. When the Corel1K image database was classified, the beaches and snow mountains are misjudged more, because the color of the snow mountain is mainly blue and white, which is very close to the color of the water and sky of the beaches, and it is easy to confuse. The current survey door frames, crime scene maps and prospecting windows are misjudged more, The reason is that at the time of image collection, there may be door frames and prospecting windows in the current survey door frames, so the classification is prone to confusion.

II. CONCLUSION

In order to solve the problem of poor classification accuracy of criminal investigation image, this paper proposes a new image classification method of SCNN-ELM. The advantages of this method are as follows: Firstly, for single image feature extraction, considering the spatial location information of criminal investigation image, it can improve the classification accuracy. Secondly, for the limitation of traditional features, this paper uses the features of convolutional neural network to better represent images. Finally, the advantage of the ultimate learning machine is higher than that of traditional support vector machines. Experimental results show that this algorithm is superior to other traditional classification method and it is an effective classification method for criminal investigation.

TABLE 4 SCNN-ELM ALGORITHM IN COREL1K IMAGE CLASSIFICATION CONFUSION MATRIX (%)

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	C_8	C_9	C_{10}
P_1	**70.0**	9.0	0.0	2.0	12.0	0.0	4.0	1.0	2.0	0.0
P_2	9.0	**66.0**	6.0	5.0	1.0	1.0	7.0	0.0	5.0	0.0
P_3	5.0	1.0	**72.0**	15.0	0.0	0.0	6.0	0.0	1.0	0.0
P_4	8.0	5.0	5.0	**70.0**	4.0	2.0	3.0	0.0	3.0	0.0
P_5	7.0	0.0	1.0	5.0	**78.0**	0.0	8.0	1.0	0.0	0.0
P_6	0.0	0.0	0.0	0.0	0.0	**96.0**	2.0	0.0	2.0	0.0
P_7	1.0	8.0	12.0	2.0	5.0	0.0	**67.0**	0.0	5.0	0.0
P_8	0.0	1.0	0.0	1.0	0.0	0.0	2.0	**96.0**	0.0	0.0
P_9	0.0	10.0	4.0	0.0	0.0	1.0	1.0	2.0	**82.0**	0.0
P_{10}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**100**

TABLE 5 SCNN-ELM ALGORITHM IN REAL CRIMINAL INVESTIGATION IMAGE CLASSIFICATION CONFUSION MATRIX (%)

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}
C_1	**81.0**	0.0	3.0	0.0	0.0	0.0	7.0	3.0	0.0	0.0	0.0	0.0	0.0	6.0
C_2	0.0	**80.0**	4.0	0.0	0.0	1.0	1.0	6.0	0.0	1.0	0.0	2.0	0.0	5.0
C_3	2.0	4.0	**55.0**	3.0	3.0	3.0	5.0	11.0	4.0	1.0	0.0	3.0	0.0	6.0
C_4	1.0	1.0	4.0	**88.0**	2.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	2.0
C_5	0.0	4.0	2.0	3.0	**85.0**	0.0	3.0	2.0	0.0	0.0	0.0	0.0	0.0	1.0
C_6	5.0	0.0	6.0	2.0	1.0	**82.0**	0.0	2.0	0.0	0.0	0.0	1.0	0.0	1.0
C_7	4.0	2.0	6.0	0.0	0.0	1.0	**80.0**	4.0	0.0	0.0	0.0	2.0	0.0	1.0
C_8	6.0	1.0	5.0	0.0	3.0	5.0	3.0	**64.0**	0.0	0.0	0.0	0.0	0.0	13.0
C_9	0.0	0.0	4.0	2.0	0.0	2.0	0.0	0.0	**89.0**	0.0	0.0	1.0	0.0	2.0
C_{10}	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	**98.0**	0.0	0.0	0.0	0.0
C_{11}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**100**	0.0	0.0	0.0
C_{12}	0.0	3.0	4.0	0.0	0.0	2.0	0.0	1.0	0.0	0.0	0.0	**90.0**	0.0	0.0
C_{13}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**100**	0.0
C_{14}	0.0	5.0	7.0	2.0	6.0	2.0	1.0	6.0	7.0	0.0	0.0	1.0	0.0	**63.0**

ACKNOWLEDGMENT

This work was supported by Shaanxi Province International Cooperation and Exchange Project (No.2017KW-013), Xi'an University of Posts and Telecommunications Graduate Innovation Fund Project (No.CXJJ 2017007).

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] Pan Z L, Meng Q I, Wei C Y, et al. Color Region Classification of Flue-cured Tobacco Leaves Based on the Image Processing and Support Vector Machine[J]. Acta Agronomica Sinica, 2012, 38(2):374-379.

[3] HAN De-ming. Essential Genealogy of the Right of Investigation in the Context of Information[J]. Journal of People's Public Security University of China (Social Science Edition), 2016, 182 (4): 66-72.

[4] Zhao Y D, Wang Q, Fan J L, et al. Fuzzy KNN classification for criminal investigation image scenes[J]. Application Research of Computers, 2014, 14(14):1-7.

[5] Gao H, Dou L, Chen W, et al. Image classification with Bag-of-Words model based on improved SIFT algorithm[C]// Control Conference. IEEE, 2013:1-6.

[6] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.

[7] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the Acm, 2012, 60(2):2012.

[8] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[J]. 2014:1-9.

[9] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[10] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9):1904.

[11] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015:770-778.

[12] Wei Y, Xia W, Lin M, et al. HCP: A Flexible CNN Framework for Multi-label Image Classification.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 38(9):1901-1907.

[13] Qu Y, Li L, Shen F, et al. Joint Hierarchical Category Structure Learning and Large-Scale Image Classification[J]. IEEE Transactions on Image Processing, 2017, PP(99):1-1.

[14] Chen Y, Jiang H, Li C, et al. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks[J]. IEEE Transactions on Geoscience & Remote Sensing, 2016, 54(10):6232-6251.

[15] ANDREWS S, TSOCHANTARIDIS I, HOFMANN T. Support vector machines for multiple-instance learning [C]//In: Proceedings of the 15th Neural Information Processing Systems, Cambridge, 2003: 561-568.