

Banknote Image Defect Recognition Method Based on Convolution Neural Network

Wang Ke^{1,2}, Wang Huiqin^{1,2}, Shu Yue³, Mao Li² and Qiu Fengyan⁴

¹*School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, P.R. China*

²*School of Information and Control Engineer, Xi'an University of Architecture and Technology, Xi'an 710055, P.R. China*

³*Chengdu Banknote Printing Ltd, Chengdu, 611103, P.R. China*

⁴*The People's Bank of China Business & Management Departments of Xi'an Branch, Xi'an 710002, P.R. China*

¹*wk1307@126.com,*

Abstract

There are shortcomings in the currently used traditional CCD imaging system which can automatically recognize banknote image defect, such as the need to manually extract the defect characteristics and low accuracy rate of detection results. This paper briefly introduced the advantage of convolution Neural Network (CNN) in image classification and designed a image defect identification method based on convolutional neural network (CNN). The experimental results on data sets show that the identification accuracy rate of this method is 95.6%, which is significantly better than traditional identification method.

Keywords: *Convolution Neural Network, Defect Recognition, Banknote Image, Deep-learning*

1. Introduction

Due to the imperfections of printing machine and printing material in banknote printing process and some influence of inevitably random factors, banknote shades distortion, ink stain, text blur, wrinkle, print missing, scratches, misregistration and other defects often appear in printing process. Identifying and detecting banknote image defect with computer vision is the current research focus and its key technology and ultimate goal is to identify image defects. Domestic and foreign scholars have done many researches on this area, but the used methods belong to traditional machine shallow learning structure, only converting original signal to specific problem space and forming a simple characteristics knot^[1]. Segmentally extract defected object, then manually select defect characteristics, manual characteristics description computation, statistical method and other shallow network identification method. Due to complicated and uneven background of banknote image, many types of defected object, defected object is tiny, it is very difficult to accurately segment banknote image defect, effectively describe and select manual characteristics, which need very specialized knowledge and not only subject to condition limits^[2-3], but also rely on accurate defect segmentation statistical model or shallow network model. It has high target pertinence and poor adaptability^[4].

Convolution neural network is one of the efficient methods in deep-learning and has become one of the hotspots of many scientific fields^[5-7], particularly in machine learning. Because this network can directly input original image and avoid the complex early pretreatment of image, it is widely used. This paper designed a banknote image defect identification method based on convolution neural network. This method uses convolution neural network to learn many training image samples and layer by layer to extract the

essential characteristics of defect area by convolution network, achieving the defect identification of samples and overcoming the needs to manually extract characteristics in traditional image classification. Experiment achieved high accuracy rate.

2. Convolution Neural Network

2.1. Convolution Neural Network Structure

Human brain is deep structure and recognizes information by filtering and sampling the received information. Deep learning is a kind of network structure which simulates human brain. It uses two learning method, which are unsupervised learning and supervised learning. Convolution neural network is a multi-layered supervised learning network which has input layer, hidden layer and output layer and hidden layer which containing convolution layer and low sample layer is the important part for extracting feature in deep convolution neural network. Overall network framework is as Figure 1 shows.

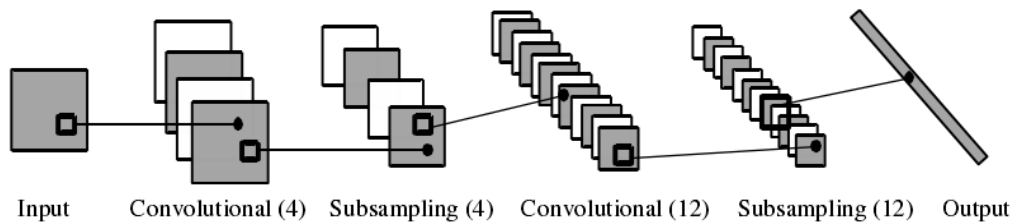


Figure 1. Convolution Neural Network Framework

Each layer in convolution neural network is composed of many two-dimensional planes and each plane contains many separate neurons. As feed forward neural network, convolution neural network can extract concrete topology structure from two-dimensional image and use back-propagation algorithm to adjust the structure and get optimized parameters which are adapted to network after making convolution and pooling operation on input image. Convolution layer and low sample layer in convolution neural network are spaced and this two times feature extraction structure makes network have higher tolerance for distortion when recognizing input samples.

2.2. Convolution Layer

Hidden layer in convolution neural network is the core of feature extraction. In convolution layer, the input of each neuron is connected to local receptive field in previous layer and extract the local feature. The format of convolution layer is:

$$x_j^l = f \left(\sum_{i \in M_j} x_i^{l-1} \times k_{ij}^l + b_j^l \right) \quad (1)$$

where l notates the number of layers, k notates convolution kernel, M_j notates receptive field of input layer, b notates bias. As for specific output mapping, feature mapping of input can be computed by applying convolution of different convolution kernel. $f(\cdot)$ is the nonlinear activation function used by neurons in convolution layer and commonly used nonlinear activation function are Sigmoid, Tanh function or softsign function. They may sometimes contribute to the gradient disappearance and in order to overcome this problem, revising linear unit ReLU function is used in many deep network structure^[8]. ReLU has faster convergence rate than traditional nonlinear function when training gradient declines. Formations of four nonlinear activation function are:

$$\text{Sigmoid: } R = \frac{1}{1+e^{-y}} \quad (2)$$

$$\text{tanh: } R = \frac{e^y - e^{-y}}{e^y + e^{-y}} \quad (3)$$

$$\text{softsign: } R = \frac{y}{1+|y|} \quad (4)$$

$$\text{ReLU: } R = \max(0, y) \quad (5)$$

2.2. Low Sample Layer

The function of low sample layer is to take sample of feature mapping of input and after taking sample, input feature and the number of features of output will not change, but the ratio of the size of output feature and input feature will decline significantly. Low sample layer is notated as:

$$x_j^l = f(\beta_j^l p(x_j^{l-1}) + b_j^l) \quad (6)$$

where $f(\cdot)$ is the activation function of low sample layer neuron and similar to convolution, b is weighting parameters, $p(\cdot)$ is pooling function.

2.3. Pooling

Using feature classification after convolution operation can train all the features to classify in theory but will increase the amount of calculation significantly. Therefore, use pooling to decrease the dimension of data after convolution operation, which divide input image into non-overlapping rectangles and calculate the average value (or maximum value) for each rectangle to describe the convolution feature after pooling. Maximum value calculation is called max pooling and average value calculation is called mean pooling. Output feature image resolution decreased after pooling operation, but can better maintain the characteristics of the high-resolution feature image. Some convolution neural networks are suitable in low sampling stage and set the convolution kernel window sliding step larger than 1 can decrease resolution.

2.4. All-Connection Layer

Each neuron on all-connection layer is connected to all neurons on previous layer feature image. Output of each neuron can be expressed as following formation:

$$h_{W,b}(x) = f(W^T x + b) \quad (7)$$

Where x notates the input of neuron, $h_{W,b}(x)$ notates the output of neuron, W notates connection weight, b notates bias, $f(\cdot)$ notates nonlinear activation function. Some recent convolution neural network introduced Dropout or DropConnect method in all-connection layer to decrease over-fitting of data^[9].

2.5. Training Method

Back-propagation rules and supervised training method are commonly used method in training convolution neural networks^[10-11]. Algorithm flowchart is as Figure 2 shows.

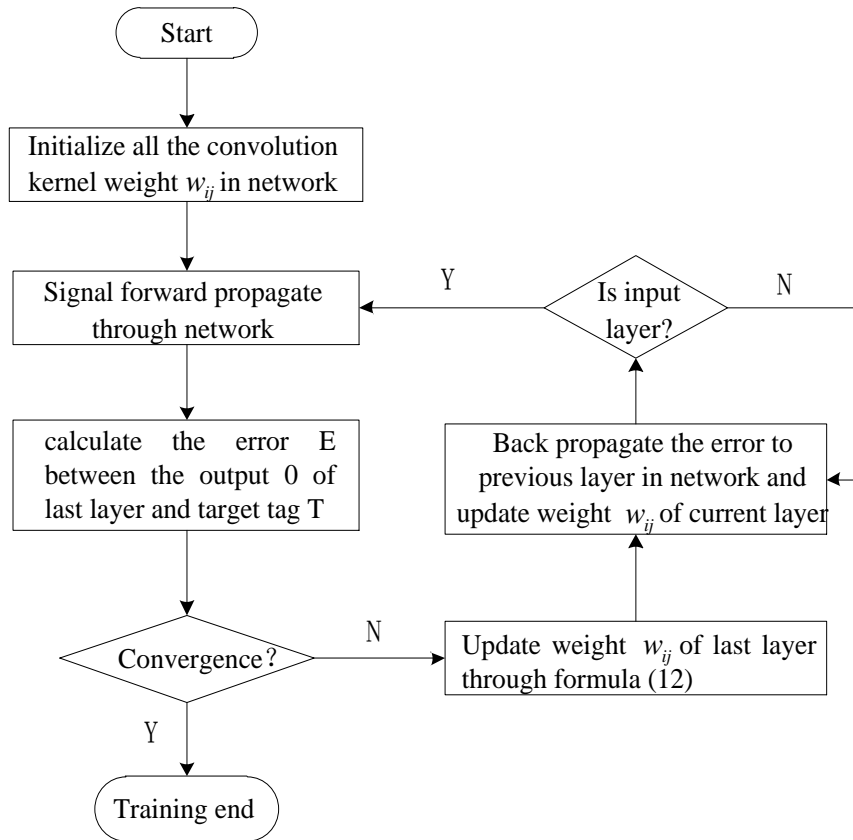


Figure 2. Training Process of Convolution Neural Network

Network signal forward propagate, which means propagate from input feature to out feature. Output X of first layer change into output image O of last layer through multiple convolution neural network layers. Compare output feature image O to expect tag T can generate error E . Propagate error to each node through each layer by traversing the reverse path of network and update corresponding convolution kernel weight w_{ij} according to weight update formula (12). In training process, initial weight value in network is usually randomly initialized (or pre-trained by unsupervised method) and network error decreases with increasing the number of iteration times. This process converges to a stable weight value set and extra training times have litter influence. For any layer L in convolution network, update formula for weight between i th input feature X_i and j th output Y_j feature w_{ij} is^[11]:

$$\Delta w_{ij} = \alpha \delta_j X_i \quad (12)$$

When L layer is the last layer in convolution network, as Figure 3(a) shows:

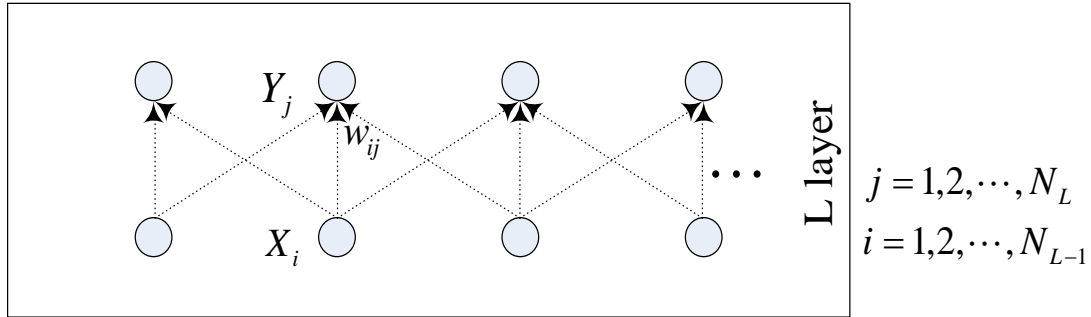


Figure 3. (a) L Layer is the Last Layer

δ_j is:

$$\delta_j = (T_j - Y_j)h'_L(X_i) \tag{13}$$

where T_j is j th expected tag, $h'_L(x)$ is derivative of nonlinear mapping function, $j = 1, 2, 3, \dots$, when L layer is not the last layer, as figure 4(b) shows, L + 1 is its next layer, then δ_j is:

$$\delta_j = h'_L(X_i) \sum_{m=1}^{N_{L+1}} \delta_m w_{jm} \tag{14}$$

where N_{L+1} is the number of output features in L + 1 layer, $m = 1, 2, 3, \dots, N_{L+1}$, w_{jm} is the weight between j th output of L layer (j th input of L + 1 layer) and m th output of L + 1 layer, as Figure 3(b) shows:

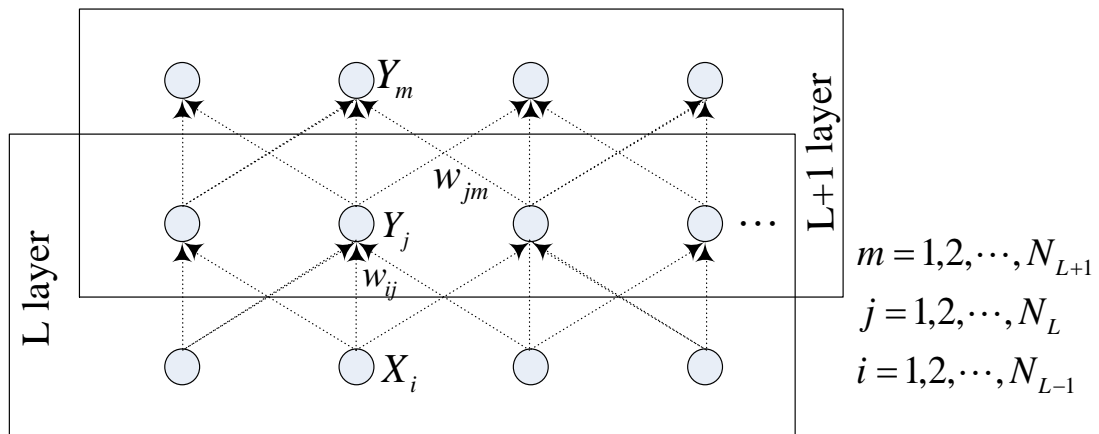


Figure 3. (b) L layer is not the last layer

3. Banknote Image Defect Recognition Method based on Convolution Neural Network

Convolution neural network achieved satisfactory results in terms of image classification and recognition, such as handwritten character recognition, traffic sign recognition. Settings of convolution layer and low sampling layer in convolution neural network are very flexible and different structure setting will get different results. Therefore, applied network structure should be different for different types of input samples. One paper^[12] applied this network structure settings for ImageNet image library

and input image size is 224*224: 96C-5S-256C-3S-384C-384S-256C-3S, including four convolution layers, four low sampling layer and the size of convolution kernel is 5*5. One paper [13] applied this network structure settings for LFW image library and the size of input image is 39*39: 20C-2S-40C-3S-60C-2S-80C, including four convolution layers, three low sampling layers and the size of convolution kernel is 4*4 and 3*3. The convolution neural network framework of banknote image defect recognition method designed by this paper is as follows:

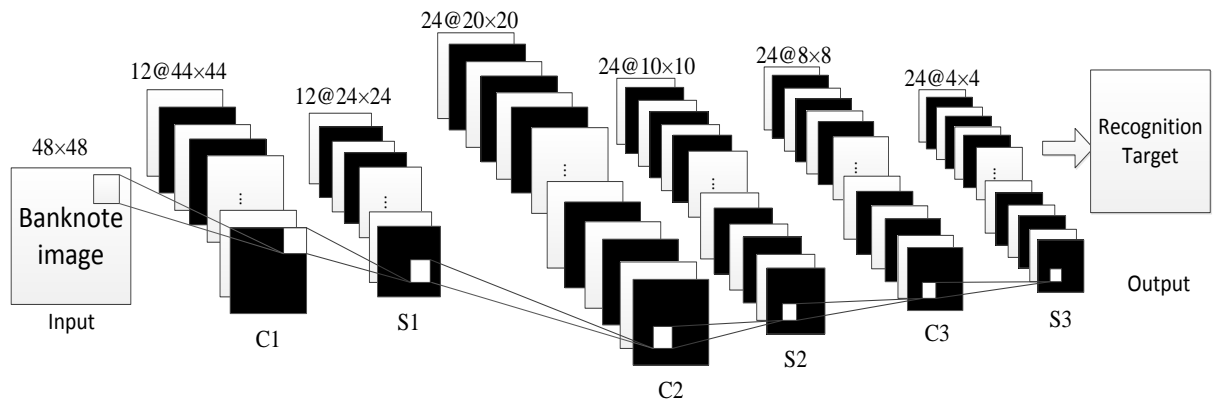


Figure 4. CNN Framework of Banknote Image Defect Recognition

CNN is composed of three convolution layers, three low sampling layers and one all-connection layer, and the size of convolution kernel for each layer is 5*5, 5*5 and 3*3. Convolution layer uses Sigmoid function as activation function and low sampling layer uses max pooling. They both use all-connection neural network.

3.1. C1 Layer

C1 is the first convolution layer and get twelve two-dimension feature images which size are 48*48, which is achieved by a convolution kernel which size is 5*5. The size of convolution kernel determines a neuron's receptive field size. If convolution kernel is too small, effective local feature cannot be extracted. If convolution kernel is too large, the complexity of the extracted features may far exceed the representation capacity of the convolution kernel. C1 uses 5*5 convolution kernel and 48*48 image to get a $(48-5+1) * (48-5+1) = 44*44$ feature image. Results achieved by convolution is not directly stored in C1 layer, but was first computed by Sigmoid activation function and then as the eigenvalues of certain neuron of C1 layer. The number of parameters need to be trained by C1 is $12 * (5*5+1) = 312$ and the number of connections between input layer and C1 layer is $312 * (44*44) = 604032$.

3.2. S1 Layer

S1 is low sampling layer and gets twelve 24*24 feature images by summing all the non-overlapping 2*2 sub-blocks, multiplied by a weight and plus a obtained bias term. Sub-sampling computation process is:

$$y = \text{Sigmoid}[w \cdot \text{sum}(x_i) + b] \quad (15)$$

Because the size of feature map in C1 is 44*44, the final sub-sampling result is a 24*24 feature sub-graph. Scaling factor used by every low sampling layer is 2 for controlling zoom decline rate. Because the scaling is exponential scaling, too fast reduced rate also means more rough extraction of image feature and more image feature details will lose. In CNN, setting scaling factor as 2 can generally meet the requirements. Each sub-sampling

feature maps need train two parameters, and then there are totally $12 \times 2 = 24$ parameters need to be trained in S1.

3.3. C2 Layer

C2 is also a feature extraction layer, which has a similar place with C1 layer and also have some differences. The size of convolution kernel used by C2 is 5×5 , so the size of feature map achieved is $(24-5+1) \times (24-5+1) = 20 \times 20$. Receptive field of each neuron in S1 is actually 10×10 of original image through processing of C1 and S1. The size of convolution kernel of C1 is 5×5 and the size of sampling sub-block of S2 is 2×2 . Now receptive field of C2 further expands and is 50×50 of original image after a 5×5 convolution kernel extract the feature of S1. C2 has 24 feature maps and the number is two times as C1's. C1 obtains twelve mapping planes through one picture in input layer, and now C2 need to map 24 feature maps from S1's 12 feature maps, during which settings are done. When each feature map in C2 is making convolution, some of all feature maps in S1 is as input and get convolution. The reason why all the feature maps in S1 is always as input is that incomplete connection mechanism makes the number of connections within reasonable limits. In the meantime, the most important is that it destroys the symmetry of the network and different combinations extract different characteristics. Method used in this paper is that if the feature map number of S1 is the factor of feature map number of C2, they have connections. For example, feature map 1 of S1 is connected to all feature maps of C2, feature map 2 of S1 is connected to all the even number feature maps of C2 and feature map 3 of S1 is connected to the feature maps of C2 which number is divisible by 3 and so on. For the j th feature map C2 $j \in \{1, 2, \dots, 24\}$, we have:

$$C_{2j} = \text{Sigmoid}(\sum_{i \in M} S_i + b) \quad (16)$$

where $\in \{i, j \text{ mod } i == 0\}$, $i \in \{1, 2, \dots, 12\}$. In the following sub-sampling layer, use the same combination strategy.

3.4. Other Convolution Layers and Sub-sampling Layers

Working principle of these layers is similar to previous layers, but as the depth increases, extracted features are more abstract and more expressive. After convolution and sampling of C1, S1, C2, S2, extracted features are expressive enough, but they are still not abstractive enough. In the experiment, when using convolution neural network which only have C1, S1, C2, S2, achieved classification accuracy is only about 60%. However, if using the convolution neural network framework in Figure 5, classification accuracy increased significantly, which means depth has great influence on performance of convolution neural network and inadequate depth will weaken feature extraction capabilities of convolution neural network.

3.5. Output Layer

Output layer is an all-connection layer of S3. S3 has $24 \times 4 \times 4 = 384$ neuron. Every neuron is connected to an output neuron and output layer has 15 neurons, so there are $384 \times 15 = 5760$ connections. This paper regards S3 as a 384-dimensional linear vector and mapping from S3 to output layer is equal to using this vector to classification. Output of each neuron represents the possibility of classification result. This classification machine has 5760 parameters, so it has ability to describe.

4. Experiment Results and Analysis

4.1. Dataset Collection

The original experimental data mainly comes from defective banknote images collected by CCD imaging detection system in production line. Experiment use ink stain defect recognition effect to verify the construction method based on convolution neural network proposed by this paper. Firstly, filter the image data sets according to the following main rules: picture which has incorrect information (such as snow point, a large area of black and so on); picture which source is not surface inspection points, such as detection point picture (such picture is black and white); picture that ink dot defects are too far from the center of the picture. When screening defect image, in order to verify the suitability of convolution neural network, we did not make image partition operation. Defect pictures contain gum gravure pattern, paper area, watermark, number, safety lines and all the other information. Tested dot-defect also contains many other defects that appearance is similar to dot, such as oil dirty, rub dirty and so on. These defects are retained, so defective points not only have deep and shallow black points, but also have red, yellow, blue and other defective points. The smallest ink point has less than three pixels, and the largest defective area has about twenty to thirty pixels. Dot defective banknote image is as Figure 5 shows.

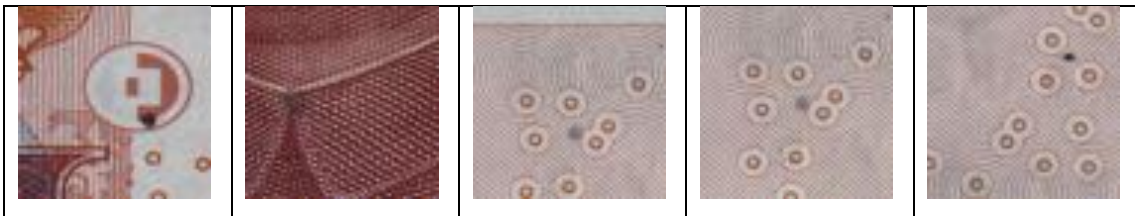


Figure 5. Dot Defective Banknote Image Sample

According to the previous design, the size of picture of input convolution neural network is 48×48 , so make uniform scaling on the original image in dataset. Finally we obtained 405 image samples including dot defect and randomly choose 100 images as testing samples and the others are put into training set. Defect-free image data sets come from the same imaging detection system, including the front and back. Scaling the banknote image area into 48×48 to get 420 defect-free sample images. Randomly choose 100 images as testing samples and the other images are put into training set. Defect-free banknote image sample is as Figure 6 shows.



Figure 6. Defect-Free Banknote Image Samples

Finally we obtained 625 training set images, including 49% positive samples and 51% negative samples. Testing set images are totally 200, including 50% positive images and 50% negative images.

4.2. Experiment Results Analysis

Convolution neural network solves the weights by iteration and obtain the ideal parameters after many iterations. Experiment results of applying different iteration times are as shown in Table 1.

Table 1. Experiment Results of Different Iteration Times

Iteration times	10	50	100	200	300	500
Average accuracy rate/%	53.5	86.2	90.1	95.6	95.6	95.0
Training time/s	64.3	342.5	711.8	1428.9	2231.5	4065.1

Network learning is not sufficient enough when iteration times are small and trained model is not ideal, so recognition effect is poor. With the increase of the number of iterations, network parameters continue to be optimized and recognition accuracy also raise, however, increase rate of accuracy rate of classification gradually reduced. When the number of training times is enough, there will be not big change about network parameters, which means convolution network has achieved convergence state and recognition classification has achieved optimal performance.

Because designed input of convolution neural network is 48*48 images and collected picture in banknote defect dataset is 185*160, firstly we need to zoom the picture size and then put into convolution neural network. It also takes into account that CNN is not sensitive to scale and rotation. So there is omitted the excessive pretreatment steps. In the processing, a small proportion element of the image could also be able to express its message to the senior feature through the multi convolution which benefits from the structure used by CNN. In order to extract the information of different scales, it is only necessary to provide adequate training samples instead of providing too complex multi-scale strategy.

In comparison with experiment results of the defect recognition method designed by this paper, we used two additional methods: one method is direct multi class support vector machine (DMSVM) ^[14] and the other one is pyramid histogram of oriented gradients (PHOGSVM) ^[15]. Experiment results of different methods are as shown in table 2. Table 2 shows that the recognition rate of the method provided by this paper is higher than DMSVM method and MLPNN method. It is equally important that the method provided by this paper directly recognizes defects, avoiding the drawbacks for the traditional methods of DMSVM and MLPNN needing to manually extract the dominant feature and has a better robustness.

Table 2. Experiment Results Comparison

Method	DMSVM	PHOGSVM	Method used in this paper
Recognition Rate	90.6%	92.5%	95.6%

5. Conclusion

Traditional banknote image defect identification extract features from defective image based on experience and then make classification on these features. There is blindness, complicated operation and low accuracy rate of classification and other defects in this method. This paper proposed a banknote image defect identification method based on convolution neural network, which is based on thorough research on convolution neural network. There is no above defect in this method and experiment results shows that this

method has pretty high accuracy rate of banknote image defect identification. However, the paper proposed by this paper also has following shortcomings:

(1) This paper just studied ink dots recognition in banknote image defect and did not study other defects such as rub dirty, print missing, overprint and other defects. More types of defective image classification and recognition remains to be studied. Therefore, types and quantity still need to be supplemented and image datasets still need to be improved.

(2) This paper just applied traditional convolution neural network structure, revised convolution neural network, such as tower convolution neural network and other methods still needs to be studied. In the meantime, we are considering combine SVM with convolution neural network to make more detailed defect classification.

Acknowledgements

This work was financially supported by the special scientific research plan project of Shaanxi Province Natural Science Foundation (2016JM6079), basic research fund project for Xi'an University of Architecture and Technology (JC1514).

References

- [1] Z. J. Sun, L. Xue, Y. M. Xu, "Overview of Deep Learning", *Application Research of Computers*, vol. 29, no. 8, (2012), pp. 2806-2810.
- [2] J. Zapata, R. Vilar and R. Ruiz, "Performance evaluation of an automatic inspection system of weld defects in radiographic images based on neuron classifiers", *Expert Systems with Applications*, vol. 38, no. 7, (2011), pp. 8812-8824.
- [3] J. Mirapeix, P. B. Garcia-Allende and A. Cobo, "Real-time arc-welding defect detection and classification with principal component analysis and artificial neural networks", *NDT & E International*, vol. 40, no. 4, (2007), pp. 315-323.
- [4] Alaknanda, R. S. Anand and P. Kumap, "Flaw detection in radiographic well-meant images using morph-logical watershed segmentation technique", *NDT&EInternational*, vol. 42, no. 1, (2009), pp. 2-8.
- [5] Q. V. Le, J. Ndiam and A. Coates, "On optimization methods for deep learning", *The 28th International Conference on Machine Learning, Bellevue, Washington*, (2011) June 28-July, pp. 2:4-7.
- [6] J. Mairal, P. Koniusz and Z. Harchaoui, "Convolutional kernel network" [DB/OL].[2014-06-12][2014-09-24] <http://arxiv.org/abs/1406.3332>
- [7] C. Szegedy, W. Liu, Y. Jia, "Going deeper with Convolution" [DB/OL].[2014-09-17][2014-09-24]<http://txiv.org/abs/1409.4842>.
- [8] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feed forward neural networks", *International Conference on Artificial Intelligence and Statistics, Brookline, MA*: vol. 30, no. 12, (2010), pp. 249-256.
- [9] Wan L., M. Zeiler, S. Zhang, "Regularization of neural networks using dropconnect", *Proceedings of the 30th International Conference on Machine Learning, Brookline, MA: Microtome Publishing*, vol. 28, no.3, (2013), pp. 1058-1066.
- [10] Y. LeCun, L. Bottou and Y. Bengio, "Grandient-based learning applied to document recognition", *Proceedings of the IEEE*, vol. 86, no. 11, (1998), pp. 2278-2324.
- [11] G. E. Hinton, "How neural networks learn from experience", *Scientific American*, vol. 267, no. 3, (1992), pp. 145-151.
- [12] A. Krizhevsky, I. Sutskevcr and G. E. Hinton, "Image Net classification with deep convolutional neural networks", *Advances in Neural Information Processing Systems. Annual Conference, Lake Tahoe, Nevada*, (2012) Dec. 3-6, pp. 1097-1105.
- [13] Y. Sun, X. Wang and X. Tang, "Hybrid deep for face verification", *IEEE International Conference on Computer Vision* (2013), <http://www.cs.utexas.edu/gagman/paper/suyog/jain-iccv2013.pdf>
- [14] Q. M. Shen, J. M. Gao and C. H. Li, "Recognition of weld defect types", *Journal of Xian Jiao tong University*, vol. 267. 44, no. 7, (2010), pp. 100-103.
- [15] T. Y. Lim, M. M. Ratnam and M. A. Khalid, "Automatic classification of weld defects using simulated data and an MLP neural network", *Insight*, vol. 49, no.3, (2007), pp. 154-159.
- [16] J. I. Arribas, J. Cidsueiro and T. Adali, "Neural architectures for parametric estimation of a posteriori probabilities by constrained conditional density functions", *Neural Networks for Signal Processing IX*, vol. 4, no. 33, (1999), pp. 263-272.
- [17] R. Vilar, J. Zapata and R. Ruiz, "An automatic system of classification of weld defects in radiographic images", *NDT and E International*, vol. 42, no. 5, (2009), pp. 467-476.

- [18] D. M. Tsai, S. C. Wu and W. C. Li, "Defect Detection of Solar Cells in E-electroluminescence Images Using Fourier Image Reconstruction", *Solar Energy Materials and Solar Cells*, vol. 99, no. 2, (2012), pp. 250-262.

Authors



Wang Ke, Received his Master degree from Southwest Jiaotong University. Lecturer and Ph. D candidate. His main research interest is digital image processing and machine learning.



Wang Huiqin, Received her PhD degree from Xi'an Jiaotong University. Professor. Her main research interest is digital image processing and signal processing.

