ICIC Express Letters
Part B: Applications
Volume **7**, Number **7**, July **2016**

ICIC International ©2016 ISSN 2185-2766
pp. **1637–1643**

HANDWRITTEN SYMBOL RECOGNITION OF DATA FORM IMAGE BASED ON CONVOLUTIONAL NEURAL NETWORK

Shengnan Zhang[1], Qihui Wang[1], Lianqiang Niu[2] and Xiangzhen Chen[1]

[1]School of Information Science and Engineering
[2]School of Software
Shenyang University of Technology
No. 111, West Shenliao Road, Shenyang 110870, P. R. China
zsncjr@sina.com; niulq@sut.edu.cn; { 710939064; 921550065 }@qq.com

Received December 2015; accepted March 2016

Abstract. *Handwritten symbol recognition is the key question in the automatic processing of data form based on image recognition, but the background interference in an image and different handwriting habits make it difficult to extract features, and then affect the quality of recognition. This paper proposes a recognition method for handwritten symbol based on convolutional neural network, which does not need to remove the background elements such as table borders, and option numbers, and can automatically learn and classify features on the two-dimensional images directly; it is very good to solve the deficiency of the existing methods. The tests of a variety of handwritten symbol images captured from different data forms show that the recognition rate of the proposed method is superior to existing methods.*

Keywords: Handwritten symbol recognition, Convolutional neural network, Data form, Deep learning

1. Introduction.

Batch recognition and automatic statistics for data forms such as questionnaire, ballot, bank bills, cargo declaration and medical insurance by using image processing and pattern recognition are important to improve work efficiency and enhance processing accuracy. Data forms usually contain a large number of objective options that mainly rely on users to choose by writing '$\surd$', '$\times$' and '$\circ$' symbols. Therefore, correct recognition for handwritten symbol is the key issue in the automatic processing of form information based on image.

In general, most of handwritten symbols in the data form are written in a data table, check box, or directly on the number of corresponding options. Since the space of data table and check box is very limited and people's handwritten habits are not the same, it often causes the overlap between handwritten symbol and checkbox border, even position staggering. The main goal of existing methods is to solve the problem of character and form line separating [1-3], in which the general processing is to remove the table borders in the first place, and then to recognize symbols by manual feature extraction and classification. In fact, because of the overlapping of foreground and background, it is easy to cause information loss of foreground symbols, which is difficult to be restored. Once the background is complex, in which contains characters, we will even have less simple means to separate them intactly. In [4-7], some methods, such as run and structure analysis, are used to extract features, and for non-standard symbols, [8] continues to adopt non-negative matrix factorization method for special treatment. To a certain extent, they improve the precision of recognition. However, due to the difficulty of image preprocessing and non-standard writing of handwritten symbols, the methods based on the simple observation and manual extraction rely more on the pixels themselves, and it is difficult to get stable and complete features. Deep learning, by contrast, proposes a method for automatically learning pattern features by computer when more data are given, and combines feature learning with the modeling process. This learning from multi-samples can achieve the high level of abstraction of image, and get more essential shape features; furthermore, it may overcome incompleteness caused by manual design features, and its advantages have been confirmed in the handwritten numeral recognition.

This paper proposes a recognition method for handwritten symbol based on deep learning model of convolutional neural network, which does not need to remove the background elements such as table borders, and option numbers, and can automatically learn and classify features on the two-dimensional images directly and obtain the ideal recognition rate.

The rest of this paper is organized as follows. In Section 2, we discuss the existing problems in the current work. In Section 3, the building of CNN used for handwritten symbol recognition is presented. In Section 4, experimental results and analysis are given to show the recognition effects of our proposed method. Finally, conclusions are given in Section 5.

2. **Analysis of Current Work and Problems.** At present, there are two kinds of data forms in the actual application, one is ballot form with explicit specification limit, in which handwritten symbols primarily include '$\surd$', '$\times$' and '$\circ$'; the other is a limited form with rich elements, such as numbers, characters, and symbols.

(1) Symbol recognition of the first kind of form is simple, the main methods of which include template matching method [9], structural features and direction tracking method [7], run features method [4], and so on. Among them, the template matching method has certain limitation to the writing position, and its anti-interference and changing ability are poor; what is more, it needs to match each symbol with template, which brings a large amount of calculation and the slower recognition speed. The recognition rate of structural features method is higher, but it is easily affected by noise, so the selected features must be stable enough. In addition, due to the different writing habits, the same symbol may have very different local contours, which will result in the difficulty of feature extraction and mis-recognition. Run features method describes the overall geometric features of handwritten symbols through run and run region. Compared with the former two methods, this method has a higher stability and recognition rate and it also puts forward some irregular judgment methods for the processing of handwritten symbols with large deformation of local contour and noise segments shown in Figure 1, but it is hard to enumerate all the exception as possible.

(2) The ballot is a simple limited form. The complicated form, such as questionnaires, papers, and notes, contains more elements and has no strict limits like ballot either. In [1-3], some problems and corresponding solutions are discussed, but the complexity of the problem not only lies in non-standard of handwritten symbols, but lies in the inaccuracy and unintegrity of target segmentation as well; the examples of some symbols with background interference are shown in Figure 2. At the same time, the images contain noises that are hard to be eliminated, which are from image processing, the background characters or strips, such as '$\surd$' on 'B' shown in Figure 2. Existing methods can simply

(a) (b) (c) (d)

FIGURE 1. Non-standard handwritten symbols: (a) a symbol contains ambiguity symbol on the right; (b) a symbol contains ambiguity symbol on the left; (c) a symbol with internal interference line; (d) a symbol with cross interference line

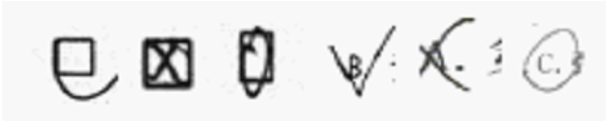

Figure 2. Handwritten symbols with background interference

eliminate the strips, but it will easily cause pattern damages. In fact, it is difficult to separate a clear and continuous strip pattern due to the influence of various noises and deformations, and this is just the essence that the manual methods cannot extract the stable features easily. Now, deep learning that learns from the image directly and automatically extracts the features makes it possible to solve this problem.

Deep learning simulates hierarchical working mode of vision system. It builds a deep network model with hierarchical structure on the basis of artificial neural network, and combines low-level features to form the more abstract ones so as to discover the distributed characteristic presentation of data. Convolution neural network (CNN) is a kind of typical deep learning model, in which the image is directly used as the input of the network and the complex feature extraction and data reconstruction in traditional recognition algorithms can be avoided; furthermore, the extracted features are invariant to transition, rotation and scaling and they also have higher tolerance for the input samples with distortion [10].

3. Building of CNN Used for Handwritten Symbol Recognition.

3.1. **Structure analysis and parameter optimization.** CNN is a multi-layer neural network, which is made of an input layer, an output layer, multiple convolution and sampling layers that are alternated. Input layer receives two-dimensional image directly. Convolution layer is used for feature extraction, each convolution layer has multiple feature maps and each map has a convolution kernel that is responsible for extracting a kind of feature. In addition, each map has independent neurons. When extracting features, the input of each neuron is linked to the local receptive field of previous layer and it does the convolution operation and nonlinear transformation accordingly.

All neurons in the same feature map of a convolution layer share weights, so this not only reduces the number of network parameters that need to be adjusted, and then leads to decrease their complexity of selection, but also makes the features have displacement and rotation invariant. At the same time, each convolution layer is linked to a sampling layer that is responsible for subsampling the obtained features from the previous convolution layer, and then the scaling invariant can be guaranteed. With twice feature extractions, the network has higher tolerance ability for the input distorted sample. In general, a fully connected layer is set at the end of CNN, and the number of its output nodes is just the number of the classified targets.

According to the principle of CNN, the CNN's structure designed for handwritten symbol recognition is shown in Figure 3.

(1) Input layer

The handwritten symbols discussed in this paper refer to three types of '$\surd$', '$\times$' and '$\circ$'. In form recognition system, in order to extract the user's handwritten information rapidly, we need to determine their handwritten areas on the basis of predefined layout structure, and then extract the corresponding image to recognize. Due to the limited space for writing, the size of input image is assumed to be 28$\times$28 in this paper. Before the image is inputted into the network, in order to meet the requirements of the network, the original image is processed by binaryzation and normalization.

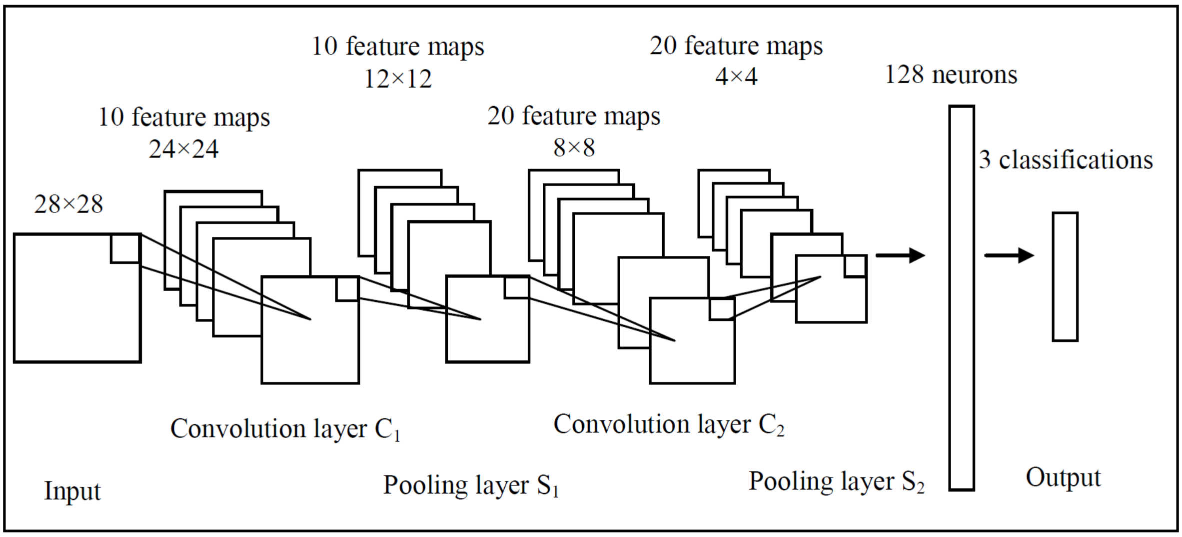

Figure 3. CNN framework for handwritten symbol recognition

(2) Convolution layer

In Figure 3, convolution layer C_1 and C_2 are also known as feature extraction layers, whose function is to process the feature maps of the previous layer so as to get feature maps of current layer. The design parameters of convolution layer include the number of layers, the number of feature maps in each layer, and the size of the convolution kernel. Among them, the last one is also the most important. From the standpoint of feature extraction, the size of the convolution kernel determines the scope of the local receptive field of neurons. Convolution kernel is too small to effectively extract the local feature, on the contrary, the complexity of extracted feature may be beyond the express ability of convolution kernel. Considering the network training time and recognition effect, convolution kernel with large size will lead to adjust too many parameters in the network and the training time will get longer; otherwise, the best result of recognition cannot be obtained without enough parameters. Through repeated contrast experiments, we set two convolution layers, the numbers of feature maps of C_1 and C_2 are set to 10 and 20 respectively, and the size of convolution kernels is 5×5.

The core work of the convolution layer is to convolute the precious feature map using a trainable kernel, and then form the corresponding neurons under the action of activation function; at last, the feature map of current layer is generated. The output of each neuron can be calculated as follows:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * Kernel_{ij}^l + B^l\right) \tag{1}$$

where f is the activation function. In CNN, the common functions include sigmoid, hyperbolic tangent, and ReLU. Because the convergence speed of ReLU is faster than other functions and it can greatly improve the training speed of network when training gradient descent, we select ReLU as the activation function, and its functional formula is $R = \max(0, y)$.

$Kernel$ is convolution kernel, M_j is a group of the selected input features, l is the number of current network layer, and B is bias. In the convolution network shown in Figure 3, the convolution kernel with 5×5 is used on the input image with 28×28, and then we get ten 24×24 feature maps of C_1 layer.

Before starting training, all the weights should be initialized with some small different random numbers so as to ensure that the network will not enter the saturated state for the larger weights and result in training failure.

(3) Pooling layer

S_1 and S_2 in Figure 3 are pooling layers, whose main function is to reduce the resolution of the output feature maps of convolution layer through the maximum pooling operation, and the migration and the distortion of the images can be eliminated while maintaining the features in the feature maps with higher resolution. The neuron in pooling layer is calculated as follows, which is the input of neurons in the next layer.

$$x_j^l = \max_{i \in n \times n} y_i^{l-1} u(n, n) \tag{2}$$

where n is the size of the sampling window, also known as the scaling factor, which is an important parameter. The larger n is, the faster reduction speed of image is, and it means that the more image details will be lost. Here, we adopt the general 2×2 setting. After sampling, the size of ten feature maps in C_1 layer is turned to be 12×12, in turn, the size of twenty feature maps in C_2 layer is 8×8, and it will be 4×4 by S_2 sampling.

(4) Output layer

Output layer is connected to the last pooling layer, which adopts the connecting way of dropout. This way randomly lets weights of some nodes in hidden layer not work during the model training. The above update mode is no longer dependent on the combined action of the implicit nodes with fixed relationship and it may avoid such situation that some features have effects only under other specific features. Through repeated contrast experiments, we find that dropout can reduce the error rate of test samples, decrease over-fitting and prove generalization ability.

In Figure 3, S_2 has 20×5×5 neurons, and each neuron is linked to the full connection layer with 128 neurons, so there are 128×500 parameters between pooling layer and full connection layer. The output of the last layer is 3, thus there are 128×3 parameters between the connection layer and output layer.

3.2. **Training of CNN.** Although CNN uses the hierarchical structure similar to that of BP neural network, CNN adopts the training mechanism named initialization layer by layer, namely training step by step and tuning step by step, and its purpose is to prevent the gradient-diffusion in the process of back propagation.

Gradient descent includes batch gradient descent and stochastic gradient descent. Among them, the former is a mode that accumulates multiple data and batches updating them which is commonly used when the whole training samples are known, but if the number of the samples is too large, the training time will last much longer and it is less used. The latter is to update the parameters one by one. Although often used in large training set, it tends to converge to the local optimal solution. Since CNN is conducted on GPU in this paper, which can improve the operation speed more than ten times, we adopt batch gradient descent method so as to obtain the global optimal solution with higher accuracy, and then get the most accurate results.

4. Experiment and Analysis.

4.1. **Dataset.** Since there is no formal image library of handwritten symbol '$\surd$', '×' and '○' up to now, we established a library with 6600 samples of handwritten symbols, which adopts the same format as MNIST handwritten digital library. The library contains 2200 images of the above three kinds of symbols; furthermore, each kind of symbol images is divided into three types: 'no background', 'with checkbox background', and 'with text background'. The images with background are subdivided into two types: symbol and form line overlapping, staggering. In addition, the library also includes some images of incomplete symbols. Thus it covers a variety of situations of handwritten symbols. Some examples are shown in Figure 4, and the size of every image is 28×28.

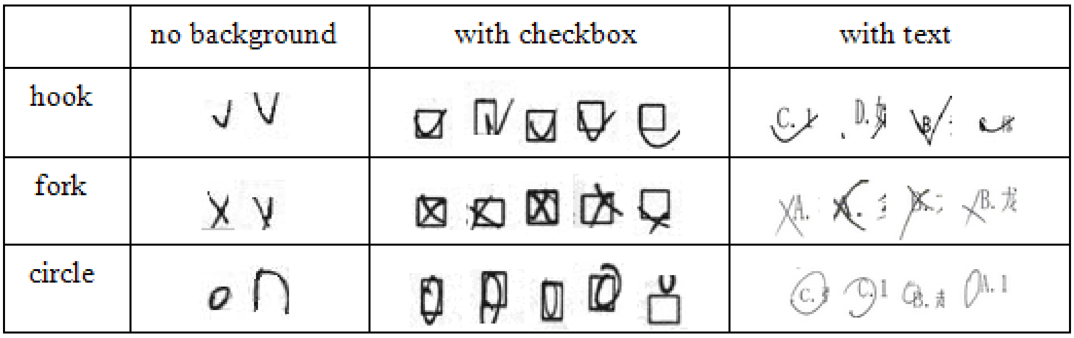

FIGURE 4. Handwritten symbol samples in image library

TABLE 1. Recognition rate with different backgrounds

No.	Train types	Train amounts	Test types	Test amounts	Average recognition rate
1	no background	1200	no background	120	100%
2	with checkbox	3000	with checkbox	300	100%
3	no, with checkbox	4200	no, with checkbox	420	100%
4	with text	2400	with text	240	99.16%
5	with checkbox, with text	5400	with checkbox, with text	540	99.07%
6	no, with checkbox, with text	6600	no, with checkbox, with text	660	98.79%
7	no, with checkbox	4200	with checkbox, with text	540	95.19%
8	with checkbox	3000	with text	240	92.9%

4.2. **Samples selection.** Convolution neural network needs a large number of samples to train, so it should contain types as much as possible. In the above image library, each symbol includes 11 classifications, and every classification includes 200 samples. We took all the samples of the library as training samples and selected 660 images from other documents as testing samples.

4.3. **Experiment results.** We designed the program using VS2010 and CUDA6.0 and tested the proposed method. GPU is from NVIDIA GeForce GTX750. The experiment is divided into eight groups, and each group is to test recognition rate of handwritten symbols with different backgrounds; the results are summarized in Table 1.

From Table 1, we can find that convolutional neural network is a reliable method of image recognition and a good recognition rate can be achieved when the train types are same as the test ones. When the structures of half of the symbols are incomplete or there are noise segments in their images, the recognition rate of three symbols without background or with checkbox background has achieved 100%. Template matching method, by contrast, can only recognize '$\surd$', '$\times$' and '$\circ$' in ideal circumstances, for the non-standard symbols shown in Figure 1, its recognition rate is only 93% and the error rate is 6.4%. [4] does not consider background factors either and the recognition rate of '$\surd$', '$\times$' achieves 100%, and the one of '$\circ$' is 99%. For the non-standard symbols, although [4] improves the recognition rate by some additional conditions, the method is limited by the relevant threshold, and false reject rate still exists.

In the 4th, 5th and 6th groups of tests, the images with text background are put into the training and test samples. Due to the great difference between text characters and '$\surd$', '$\times$' and '$\circ$' symbols, the recognition rate is a little down, but it has achieved more than 98.7%. In the 7th and 8th, the types of the training and test samples are different, but we can find that convolution neutral network has good generalizing ability. In the follow-up work, aiming at handwritten symbol images with text background, we will adopt appropriate pretreatment and increase the number of training samples to further improve the recognition rate.

5. **Conclusion.** Aiming at the recognition of '$\surd$', '$\times$' and '$\circ$' in data form, we establish a recognition method based on CNN, and illustrate the design of network structure and parameter selection. The proposed method can recognize handwritten symbol precisely on the unpretreated and uncomplicated image without text background, and it can overcome the interference of background elements on recognition and take off the writing restrictions of existing methods, which is of high resistance to noise and robustness. By using small-size convolution kernels to convolute several times, the partial features of handwritten symbols can be extracted effectively. At the same time, it combines with the max pooling technology to further highlight the symbol features, thus improves the ability of correctly recognizing for symbols with partial or fractured structure.

The experimental results show that the proposed method has obvious advantages on recognition ability and generalization ability, and it can be used to the various recognitions of handwritten symbols.

Acknowledgement. This work is partially supported by Scientific and Technological Plan Project of Shenyang City (F15-194-5-00) and Scientific and Technological Plan Project of Liaoning Province (2015410010).

REFERENCES

[1] Y. Zhang, S. Y. Yu, C. Y. Zhang et al., Extraction and removal of frame line in form bill, *Journal of Computer Research and Development*, vol.45, no.5, pp.909-914, 2008.
[2] C. S. Liu, S. Y. Pan, Y. F. Zheng et al., A frame line detection and removal algorithm for form document recognition, *Journal of Electronic and Information Technology*, vol.24, no.9, pp.1190-1196, 2002.
[3] Y. X. Li, Y. F. Sun and Y. Z. Zhang, Adaptive distance-weighted character and form line separating algorithm, *Computer Engineering*, vol.33, no.4, pp.206-208, 2007.
[4] J. J. Zhang, G. Xiao and Y. M. Zhang, Recognition approach for ballot symbol based on run features, *Journal of Computer Applications*, vol.32, no.7, pp.1906-1909, 2012.
[5] X. N. Song, Z. Liu, D. J. Yu et al., Hybrid detection approach for handwritten specific symbol in form bill image, *Journal of Nanjing University of Science and Technology*, vol.36, no.6, pp.909-914, 2012.
[6] H. Y. Wang, *Research on High Accuracy Recognition and Preprocessing for Handwritten Character*, Master Thesis, Anhui University of Technology, 2015.
[7] L. R. Hu, J. G. Wu and X. Guo, Speedy method of ballot image recognition, *Computer Engineering and Design*, vol.33, no.12, pp.4629-4633, 2012.
[8] L. R. Hu, *Method for Non-negative Matrix Factorization and Its Application in Ballot Image Recognition*, Ph.D. Thesis, Anhui University of Technology, 2013.
[9] J. Q. Shen, *Research on Layout Understanding and Fast Recognition of Ballot Image*, Master Thesis, Zhejiang University of Technology, 2009.
[10] Y. Zheng, Q. Liu, E. Chen, Y. Ge and J. L. Zhao, Time series classification using multi-channels deep convolutional neural networks, *Lecture Notes in Computer Science*, vol.8485, pp.298-310, 2014.