

PAPER • OPEN ACCESS

Handwritten digit recognition based on corner detection and convolutional neural network

To cite this article: Chen Zhong *et al* 2020 *J. Phys.: Conf. Ser.* **1651** 012165

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Handwritten digit recognition based on corner detection and convolutional neural network

Chen Zhong^{1*}, Yuhang Wang¹, Decheng Zhang¹ and Ke Wang²

¹Shenyang Fire Science and Technology Research Institute of MEM, Shenyang, Liaoning, 110034, China

²State Key Laboratory of Robotics and Systems, Harbin Institute of Technology, Harbin, Heilongjiang, 150080, China

*Corresponding author's e-mail: zhongchen@efire.cn

Abstract. In the field of inspection and testing, it is necessary to manually fill a large number of test record forms, and the digital information accounts for a large proportion in the record forms. For the purpose of automatic and accurate recognition of handwritten form data, this paper proposes a handwritten digit recognition method based on Harris corner detection algorithm and convolutional neural network(CNN). The Harris corner detection algorithm is used to identify the positioning marks in the form image, which determine the possible fill-in positions of handwritten digits. Through a 14-layer CNN model, and using the ReLu function as the excitation function, the handwritten digit features of the target area in the image are recognized. The experimental results show that the average recognition accuracy rate of this method on the test set can reach 98.14%. So a technical solution for intelligent recognition of record forms is thus provided.

1. Introduction

Due to the rapid growth of information, the manual filling and entry of various forms has become more and more complicated. With the continuous development of office automation, handwritten digit recognition technology is frequently used in various industries, such as: test records, mailing lists, and checks. In the above-mentioned application fields, the accuracy of handwritten digit recognition is generally required to be higher, so related recognition algorithms are constantly improving.

Lee Y[1] combined the advantages of the KNN classifier, radial basis function networks (RBF) and backpropagation network, and proposed a relatively balanced method in algorithm complexity, classification accuracy and resource consumption. Nibaran Das et al.[2] used genetic algorithm (GA) to search for the best classification features in local area, and used support vector machine (SVM) classifier to recognise numbers. By experimental verification, the maximum recognition accuracy of this method was 97%, which is better than that of simulated annealing algorithm (SA) of 96.7% and mountain climbing algorithm (HC) of 96.7%. Min Feng et al.[3] proposed a handwritten digit recognition method based on principal component analysis (PCA) network and compressed sensing. An improved PCA network was used to extract features in the image. Then a sparse random measurement matrix that meets the conditions of compressed sensing RIP was used for the projection of the extracted feature space. Thus a low-dimensional compressed subspace that can retain high-dimensional image feature space information was obtained. Finally, SVM was used to train and recognize the features after dimensionality reduction.



In recent years, with the rapid development of deep learning technology in the field of image recognition, traditional methods of manually extracting image features are gradually replaced by deep network learning methods. Among them, CNN as an efficient recognition algorithm, has attracted extensive attention [4, 5]. Because the network can directly input the original image, avoids the complicated pre-processing of image and have high accuracy, it is widely used. In this paper, the Harris corner detection algorithm [6, 7] is used to extract the handwritten digit target area in the image. Based on 14-layer convolution neural network structure, handwritten digit is trained and recognized. The experimental results verify the effectiveness and accuracy of the algorithm.

2. Data set extraction based on Harris corner detection

2.1. Harris corner detection

Assuming that a fixed-size window is used to slide on the image, the change degree of pixel grayscale in the window before and after the slide is compared. If the sliding in any direction has a large grayscale value change, it can be considered that there are corner points in the window. The change of the pixel grayscale in the window before and after sliding can be expressed as:

$$E(u, v) = \sum_{u, v} w(x, y) (i(x+u, y+v) - i(x, y))^2 \quad (1)$$

Where $i(x, y)$ and $i(x+u, y+v)$ represent the grayscale value before and after the window is slid, $w(x, y)$ represents a weight value. A first-order Taylor expansion for $i(x+u, y+v)$ is performed as:

$$E(u, v) = \sum_{u, v} w(x, y) (i_x(x, y)u + i_y(x, y)v + o(x^2 + y^2))^2 \quad (2)$$

Because $o(x^2 + y^2)$ has little effect on the change of grayscale value, it can be ignored. Expanding the square term of the above formula and passing the matrix transformation, it can be expressed as:

$$E(u, v) = (u, v) \begin{bmatrix} \sum_{(u, v) \in w} i_x^2(u, v) & \sum_{(u, v) \in w} i_x(x, y) i_y(x, y) \\ \sum_{(u, v) \in w} i_y(x, y) i_x(x, y) & \sum_{(u, v) \in w} i_y^2(u, v) \end{bmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \quad (3)$$

Let $A = \sum_{(u, v) \in w} i_x^2(u, v)$, $B = \sum_{(u, v) \in w} i_x(x, y) i_y(x, y)$ and $C = \sum_{(u, v) \in w} i_y^2(u, v)$ there is a real symmetric matrix M and a similar diagonal transformation matrix P , satisfying: $PMP^{-1} = PMP^T$. Therefore, the above formula can be expressed as:

$$\begin{cases} E(u, v) = (u, v) \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = M \\ \begin{bmatrix} A & B \\ B & C \end{bmatrix} = P \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} P^{-1} = P \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} P^T \end{cases} \quad (4)$$

Where λ_1 and λ_2 are the eigenvalues, let be the rotation of (u, v) , and further simplify to:

$$E(u, v) = (u, v) \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \lambda_1 u^2 + \lambda_2 v^2 \quad (5)$$

Therefore, the amount of grayscale change can be measured by λ_1 and λ_2 :

1) When $\lambda_1 \gg \lambda_2$ or $\lambda_2 \gg \lambda_1$, because the changes of u and v have an unbalanced effect on $E(u, v)$, there would be a large change in a certain direction, so an edge in the image is detected;

2) When λ_1 and λ_2 are both small, since the changes of u and v have little effect on $E(u, v)$ in any direction, so a smooth area which has little grayscale change in the image is detected;

3) When λ_1 and λ_2 are both large, the changes of u and v affect $E(u,v)$ drastically in any direction, so corners of the image are detected.

In order to quantify the above three situations, set a small positive number Δk as the threshold, and establish the corner feature parameter R :

$$R = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \quad (6)$$

When $R < -\Delta k$, it is the edge of the image; when $|R| \leq \Delta k$, it is the flat area of the image; when $R > \Delta k$, it is the corner of the image.

2.2. Filling location identification

In order to establish a sample database of handwritten digital images, a digital template is used to collect the handwriting information of each test recorder. The black squares around the template are used to mark the grid positions of each handwritten digit, as shown in figure 1. Harris corner detection algorithm is used to search and mark corner positions in the image, and adjust the value of Δk to ensure that all black squares can be recognized as corners. Then, coordinates of a rectangle with the smallest area that can include all corners can be obtained, which can lock the target area of handwritten digits in each row and column. Finally, the image in the rectangle is divided into rows and columns to obtain sample pictures of each handwritten digit.

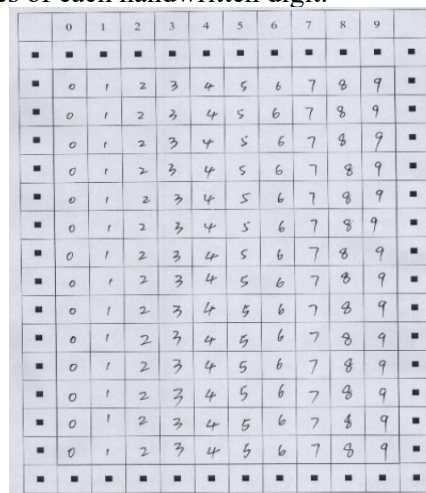


Figure 1. Handwritten data collection template.

Through the above method, a total of more than 14,000 sample pictures containing 0~9 numbers were collected, and then the data set was randomly divided into training set, validation set and test set according to the ratio of 8:1:1, as shown in figure 2. Using a similar method, black squares are set in the record form template to identify the required filling position of handwritten digits, and the corner detection is used to locate the target area to be identified in the record form.

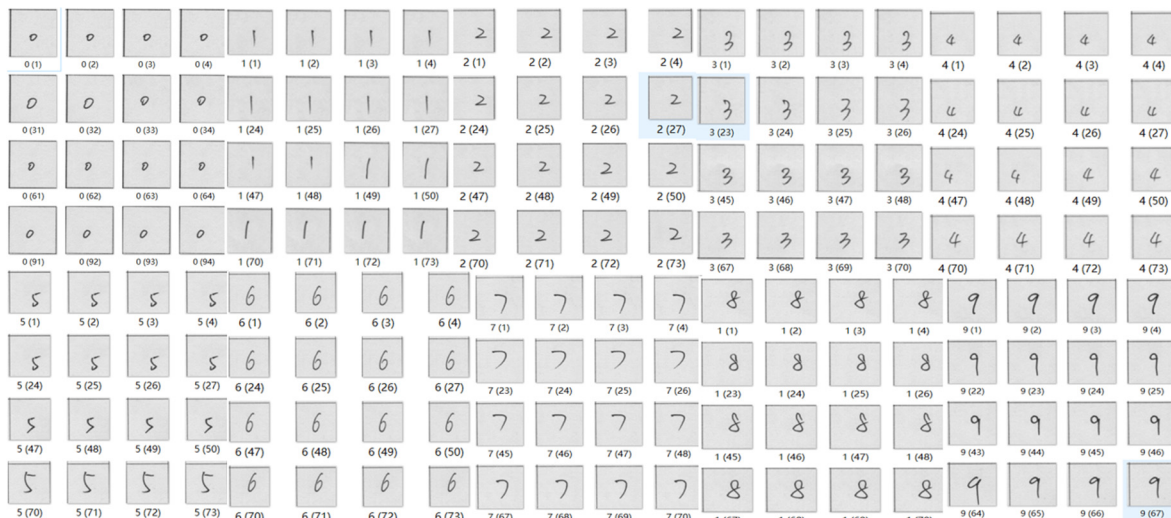


Figure 2. 0~9 handwritten digit samples.

3. CNN structure

3.1. Hierarchical structure design

This paper adopts a hierarchical design of CNN structure to realize the training and recognition of handwritten digital samples. A total of 14 hierarchical structures are constructed, including input layer, 5-level convolution layer, 5-level pooling layer, 2-level full connection layer and softmax output layer, as shown in figure 3.

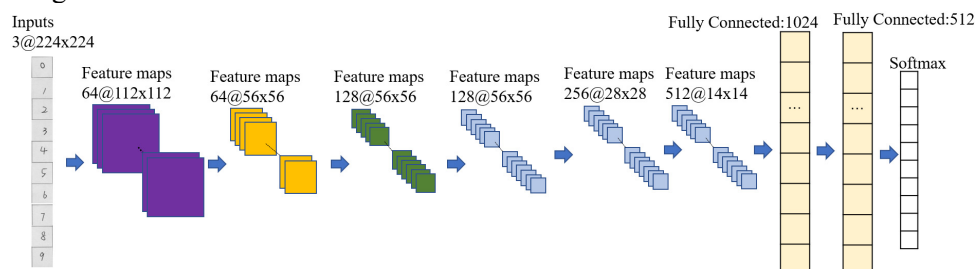


Figure 3. Learning model of handwritten digit recognition.

The input layer can be regarded as a three-dimensional neuron reflected by a $224 \times 224 \times 3$ RGB image. The convolutional layer convolves each small part of the image features extracted by the upper neural network, so as to obtain more abstract and easy-to-process features. The excitation layer uses the excitation function to perform nonlinear transformations on the results of convolution operation in the convolution layer. The pooling layer calculates the energy in a certain area of the feature response graph that can represent the characteristics of the area, so as to reduce the feature map in the convolution structure. The fully connected layer classifies and labels the training samples, and predicts the samples' category through continuous propagation. The Softmax layer is used to predict the output result of the network. According to the probability of the samples' category in the last fully connected layer, the maximum probability of the category is obtained by using Softmax logistic regression.

3.2. Model parameter settings

The commonly used excitation functions of the excitation layer are Sigmoid function, Tanh function and ReLu function [8, 9]. Through experimental comparison, it is found that the ReLu function does not require input normalization to prevent saturation. Therefore, this paper uses the ReLu function as the excitation function.

The learning rate has a great influence on the handwritten digit recognition system. When learning rate is too fast, the weight swing will be large, which can only fluctuate around the minimum point of the cost function, so that the global minimum point cannot be obtained. But when the learning rate is too slow, it will greatly increase resource consumption. In order to select the learning rate reasonably, this article uses exponential decay to set the learning rate, which can ensure that the training speed is faster in the early stage and slowed down in the later stage, which is conducive to obtaining the optimal solution. In this paper, the initial value of learning rate is set to 0.1.

For evaluating the handwritten digit recognition model, the average accuracy rate A is introduced as the evaluation index:

$$A = \frac{1}{p} \sum_{i=1}^p \frac{n_{ai}}{n_i} \quad (7)$$

Where, n_{ai} is the number of accurate samples for each prediction, n_i is the number of samples passed into the network at one time, and p is the number of training rounds in each iteration.

4. Experimental results

4.1. Data set test analysis

In the training process of convolution neural network model, a training batch is set to 128, so that every 200 iterations output a log and the accuracy is calculated. After about 16,000 training sessions, the accuracy rate remained at a stable level. At the same time, the robustness of the model is enhanced by randomly adjusting the brightness, contrast or rotating the input picture. According to the test results, the curves of accuracy and loss with training times are drawn, as shown in figure 4 and 5.



Figure 4. Accuracy curve on training set.

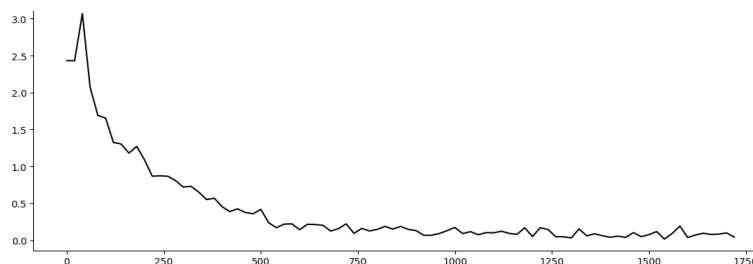


Figure 5. Loss curve on training set.

The confusion matrix of the CNN model on the training set and the confusion matrix on the validation set are shown in figure 6 and 7.

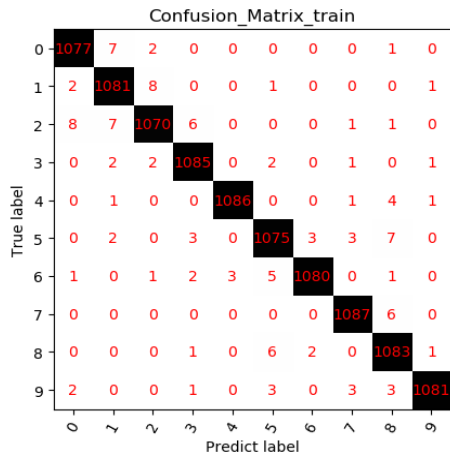


Figure 6. Training set confusion matrix.

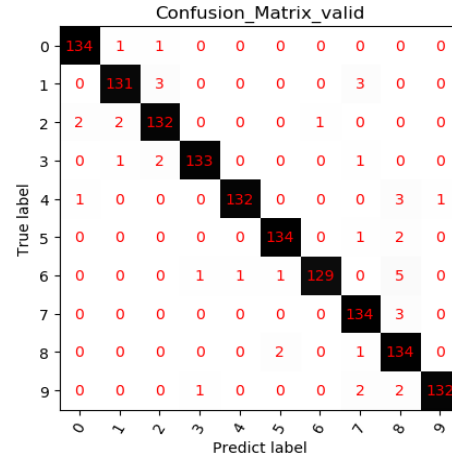


Figure 7. Validation set confusion matrix.

On the test set, the classification accuracy of handwritten digits 0~9 by the CNN model is shown in Table 1. The confusion matrix of the test set is shown in figure 8.

Table 1. Test set classification results.

digits \ test results	0	1	2	3	4	5	6	7	8	9
recall rate (%)	99.26	98.52	97.04	98.52	96.30	97.78	97.04	99.26	95.56	94.93
Accuracy (%)	96.43	97.08	98.50	95.68	97.74	99.25	97.04	98.53	95.56	98.50

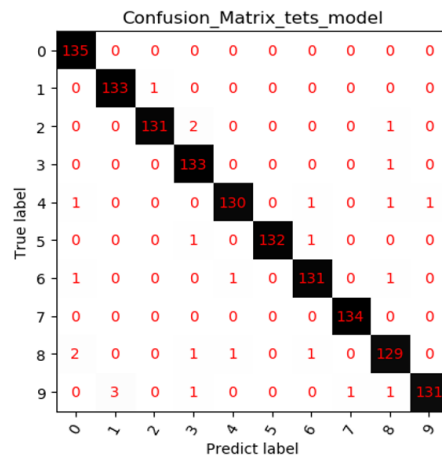


Figure 8. Test set confusion matrix.

4.2. Recognition effect and confidence

By inputting a single picture, the prediction results and confidence levels is given by the model. Two pictures with different writing standards of the same number were tested respectively (see figure 9). The CNN model accurately predicts the number in the picture, and gets the confidence value by Softmax. The probability of being 0 is 99.91% (see figure 9(a)) and 97.82% (see figure 9(b)), respectively.

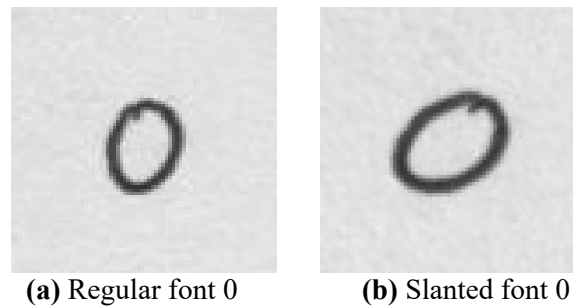


Figure 9. Digital samples of different writing standards.

From the recognition results, it can be seen that when the numbers in the picture have irregular writing such as slant, the CNN model can also accurately give the recognition results, which has strong robustness.

5. Conclusion

This paper proposes a handwritten digit recognition method based on Harris corner detection algorithm and CNN. The black squares are set on the handwritten number template to calibrate the possible position of handwritten numeral, and the corner detection algorithm is used to locate and extract the target area of handwritten numbers. Based on the platform of PyTorch deep learning framework, a 14-layer CNN is built. By training more than 14000 handwritten digital images and testing on a test set composed of more than 1000 images, the accuracy curve and loss curve are drawn and the confusion matrix is given. Experimental results show that the average recognition accuracy on the test set can reach 98.14%.

Acknowledgments

In this paper, the research was sponsored by Science and Technology Program of Liaoning Province(2019-ZD-0865), and National Key R&D Plan Project (Project No. 2017YFF0207004).

References

- [1] Lee, Y. "Handwritten Digit Recognition Using K Nearest-Neighbor, Radial-Basis Function, and Backpropagation Neural Networks." *Neural Computation* 3.3(2014):440-449.
- [2] Das, Nibaran, et al. "A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application." *Applied Soft Computing* 12.5(2012):1592-1606.
- [3] Feng, Min, Y. Xianyi, and Z. Yanduo. "Method of handwritten digit recognition based on improved PCANet." *Journal of Huazhong University of Science and Technology(Natural Science Edition)* (2018).
- [4] Karpathy, Andrej, et al. "Large-Scale Video Classification with Convolutional Neural Networks." *Computer Vision & Pattern Recognition IEEE*, 2014.
- [5] Ciresan, Dan Claudiu, et al. "Flexible, High Performance Convolutional Neural Networks for Image Classification." *International Joint Conference on Ijcai DBLP*, 2011.
- [6] Shengli, Cai, and Z. Huiqing. "Measurement of Image Rotation Based on Harris Corner Detection." *Computer Measurement & Control* 19.1(2011):30-32.
- [7] Liping, Chang, J. I. Xiaoping, and Z. Liang. "Block Image Registration Method Based On Harris Corner Detection." *video engineering* (2013).
- [8] Gayatri Khanvilkar, Deepali Vora. "Activation Functions and Training Algorithms for Deep Neural Network" *International Journal of Management, Technology And Engineering*, Volume 8 Issue VI (2018):226-233.
- [9] Zhi-Lin, Q. U., and H. U. Xiao-Fei. "Research on Convolutional Neural Network Based on Improved Activation Function." *Computer Technology and Development* (2017).