A Handwritten Chinese Character Recognition based on Convolutional Neural Network and Median Filtering

Yuan Zhuang\textsuperscript{1,a}, Qiong Liu\textsuperscript{2,b,*}, Chengjun Qiu\textsuperscript{1,c}, Cong Wang\textsuperscript{1,d}, Fudong Ya\textsuperscript{3,e} Ahamed Sabbir\textsuperscript{1,f}, Jiaqi Yan\textsuperscript{1,g}

\textsuperscript{1}College of Mechanical, Naval Architecture & Ocean Engineering, Beibu gulf university, Qinzhou, Gungxi, China
\textsuperscript{2}College of Science, Beibu gulf university, Qinzhou, Gungxi, China
\textsuperscript{3}School of Computer, Electronics and Information, Guangxi University, Nanning, Guangxi, China
\textsuperscript{a}email: johnny160005@foxmail.com
\textsuperscript{*}Corresponding author: \textsuperscript{b}email: zhuangyuan@bbgu.edu.cn

Abstract: with the rapid growth of researches toward computer vision and pattern recognition, methods that based on convolutional neural network (CNN) have shown unique advantages on handwritten characters recognition, also provided impressive results. This paper proposes a model based on CNN to deal with matters of handwritten Chinese character recognition. Different with conventional recognition system, in this model, input images are preprocessed by median filtering to smooth and reduce noise. For testing the stability and performance of the model, two testes are managed respectively. In integral test, experimental results show that the accuracy rate of recognition approach to 90.91% after 5000 times training, mean square error is decreased to 0.0079 at last. Meanwhile, this system also has a good performance at real-time test.

1. Introduction
Handwritten Chinese character recognition is no only one of the research hotspots in pattern recognition, but also one of the most difficult research topics in the field of character recognition. It is being widely used at bank bill identification, mail sorting, office automation and etc. [1], provided a huge benefit to social and economic development. However, Chinese have been considered as one of the hardest languages to write at worldwide, due to its various typefaces and written forms. Therefore different individuals have different writing styles, this causes many difficulties while recognizing Chinese handwritten words. In the past few decades, researchers have proposed many traditional methods to improve the recognition performance of offline handwritten Chinese characters, but the recognition accuracy still lags far behind that of human beings [2].

Years ago, Deep learning and convolutional neural network were successfully applied in computer vision to solve complex problems of recognition, these methods also bring new inspiration and solutions to handwritten Chinese characters recognition (HCCR). In 2012, Ciresan D et al. [3] proposed a Multi-column deep neural networks for handwritten Chinese character classification to recognize 3755 classes of Chinese characters, gained 93.5% overall accuracy, said the accuracy of the model nearly approaches that of human recognizing the same characters. In 2015, Ming-Ke Zhou et al. [4] used both Discriminative quadratic feature learning and deep CNN on HCCR task, gained 94.44% and 94.47%
accuracy rate respectively, although these two results are extremely closed, the deep CNN method needs lesser preprocessing procedures on the input data than that of counterpart. Relative works have also done by HE Z, et al. [5] ZHANG X Y, et al. [6] and ZHONG Z, et al. [7] Recently, researches of CNN tend to focus more at optimizing the running speed and storage capacity [8], therefore many methods to optimize CNN model have come out, in order to improve these. Xiao et al. [9] proposed a 9-layer-CNN with an improved computational speed and a compact size that optimized by Global Supervised Low-rank Expansion (GSLRE) method and Adaptive Drop-weight (ADW) technique, this network has a 90% reduction of computational cost, but with only a 0.21% drop in accuracy.

Despite the advantage of CNN that it does not require strictly about input [10], noise of image (e.g., Gaussian, random, and salt pepper noises) influence the accuracy rate of recognition. Median filter is one of the well-known order-statistic filters to eliminate these noises [11], especially to reduce speckle noise and salt pepper noise. The main principle of a median filter is applying a median of the gray levels from a pixel neighborhood instead of average operation, to replace the gray level of each pixel [12].

In order to eliminate these potential effects as well as further improve the accuracy rate of handwritten Chinese characters recognition, this paper combines the advantages of both median filter and convolutional neural network. In this model, median filter acts as a preprocessor to reduce noise that may affect the accuracy of recognition, a 6 layers CNN was built to recognize and classify target images, the dataset used is from CASIA-HWDB1.1. This method enhances the efficiency of CNN with an increased accuracy rate and less time consumed.

2. Method

2.1. dataset
The features of handwritten Chinese in specific forms are selected as the discriminative features of the model. Since handwritten Chinese contains numerous writing forms that greatly different with each other, each form has distinctive features, accordingly a large number of data are required for training CNN. Here, CASIA-HWDB1.1 was used as the dataset for this model, CASIA-HWDB1.1 was published by Chinese Academy of Sciences, it contains 3755 classes of single Chinese word written by 300 more individuals. A total number of 20 characters are selected from this set with each of the characters contributing 100 images to train, 60 images for testing the accuracy in the final. Each of these images has unique features, as shown in Fig 1. These 20 characters are summarized below at Table 1.

![Fig.1. different features of same word in different writing forms](image-url)
2.2. Preprocessing
Here a preprocessing stage is needed before training the model, as to reduce noise that probably cause recognition accuracy to be lower, one more reason is the CNN required 28x28 input images. Hence, in the first state, we perform a median filtering method to these input image to mitigate effects. In image processing, a Two-dimensional median filtering method is based on a two-dimensional sliding window, it moves from left to right and from top to bottom on the pixel points of the image in turn. When each pixel is passed, the gray values in the window are sorted to generate monotonous gray sequence. The median value is chosen as the output for this pixel point [13]:

$$g(x,y) = med\{f(x - i, y - j); (i,j) \in S\}$$ (1)

Where, \(f(x, y)\) is the original image, \(g(x, y)\) is the image after the process, \(S\) is the two-domain pattern.

By applying median filtering method, input images are smoothed as shown below in fig.2, whereafter these images are downsized to 28x28 to fit the input size of CNN.

Fig.2. the median filtering of image: a. original image for “且”, b. filtered image for “且”, c. downsized image for “且”

2.3. Design of handwritten Chinese characters recognition
We implement the program in MATLAB, it consists of 3 main parts, the first part is mainly to read data, where all images are transferred to matrices, after that these images are randomly divided into two sets which are for training and testing respectively, i.e., 100 images for training and 60 images for test the accuracy. Finally, after all data are loaded, a 6-layers convolutional neural network is ready to construct.

Fig.3 shows the architecture of this handwritten Chinese character recognition system. In this system, there are two convolutional layers that involve six and twelve filters severally. In the first convolutional layer, matrices transferred from input are convoluted with different kernels according to different filters, each one matric output six feature maps. After subsampling in the first pooling layer, these feature maps are then preformed convolution with twelve kernels in the second convolutional layer, where outputs twelve feature maps to the second pooling layer. After second pooling, the network processes all images to 12 feature maps with 4x4 size each, whereafter, these maps are sent to full connected layer to determine the correct identification.
Fig. 3. Architecture of handwritten Chinese character recognition system based on CNN and median filtering.

Training and testing are proceeded after the network is constructed. By setting the epochs, which is according to the times that training set passing through the neural network for weights update, the network can be trained as needed. The weights of the neural network are randomly initialized to a small number, to mitigate the saturation of the model. The train algorithms and rules are generalized detailedly in the next part. When training is completed, images for testing are used to verify the recognition capability of the system, where output the main parameters such as accuracy rate and training mean square error.

2.4. Learning algorithm

When training the model, the error or cost function needs to be first calculated by present output of the network for every mini-batch. Gradient descent is then used to update the weights of the model. The algorithm sets to reduce the gradient of the error function with respect to the weights and biases by updating the parameters at a magnitude directly proportional to the current gradient. Compared to typical gradient descent, in mini-batch gradient descent, the update is done several times with smaller steps during one pass of the training set, where the updating terms are computed for every mini-batch [14]. The method has been proven to lead to faster convergence and mitigate over-shooting of the model. Details of the training algorithm is summarized in Algorithm 1 and the mini-batch size is set to 50 [15]. Mathematically, the updating rules can be summarized as:

\[ E_N = \frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{C} (t^n_k - y^n_k)^2 \]  
\[ w_l = w_l - \eta \frac{\delta E(w,b)}{\delta w_l} \]  
\[ b_l = b_l - \eta \frac{\delta E(w,b)}{\delta b_l} \]

Where, \( E(w,b) \) is the cost function, \( w_l \) is the weight, \( b_l \) is the bias, and \( \eta \) is the learning rate.

3. Result and discussion

3.1. Testing

As mentioned above, dataset has been separated into two sets, one for training with the specific number of epochs, the else is used to test the performance of the model. After the training procedure, the matrices of the rest 1200 images (i.e., 60 images for 20 characters each) are loaded to the network to identify the characters shown in these images. For understanding the effect of epochs on recognition accuracy rate, 10 epochs of training is set as the baseline, to the end of 5000 epochs. By 10 epoch, it can be seen that the model fails to recognize the most of characters, with only 4.58% of 1200 samples have successfully been identified. Whereas the recognition accuracy rate shows a shift when the number of epochs increases to 100, with 46.92% of samples can be identified. The sharp increase has lasted to 200 times of epoch, to 80.75%. However, it becomes slow to grow, with 86.17% accuracy rate of 500 epochs,
87.08% of 1000 epochs, 89.01% of 2000 epochs, finally reached to 92.28% of 5000 epochs. As seen from Table 2, the model precedes to recognize better when the number of training epoch increase, showing an early sign of convergence at 100 epochs (Fig. 4). The curve graph of identification accuracy rate is shown in Fig. 5. Meanwhile, the mean square error decreases from 2.9718 of 10 epochs to 0.0083 at the end of 5000 epochs, yet the minimum occurs at nearly 2000 epochs with 0.0074, shown in Fig. 6.

Table 2. The prediction accuracy corresponding to different number of training epoch

| epoch | 10   | 100  | 300  | 500  | 1000 | 2000 | 5000 |
|-------|------|------|------|------|------|------|------|
| accuracy | 45.8% | 45.92% | 84.25% | 86.17% | 87.08% | 89.01% | 92.28% |

Fig. 4. The prediction results for different number of training epoch: a. 100, b. 5000 (if the red start matches the blue circle, that means the sample are identified properly)

Fig. 5. curve of accuracy rate with the increase of training epochs
Besides, the bar chart illustrated recognition accuracies for each single character is shown below at fig. 7, as seen here, the top specific accuracy is for character “丁”, with 96.67%. The lowest which is for character “丝” also shows a recognition accuracy over 80%.

3.2. Real-time recognition
To further test the performance of this model, a real-time recognition is arranged. This test is conducted based on a model that trained 5000 epochs, 5 characters have randomly been written in real-time to be used as the test samples, shown in Fig. 8. Result shows that 4 samples have been correctly identified, shown in fig. 9, although many of these words are written in complex background, except the image for character “丑”. The reason is that this character is composed of many small images which are also in different colors with each else, it may cause an enormous difficulty for the model only trained by uncomplex images of this character.
4. Conclusion
The proposed CNN and median filtering based handwritten Chinese recognition system can identify Chinese words with a good accuracy rate, proving its potential for future usage. Median filtering method has also enhanced the accuracy rate by 1.8%, compared with the 90.48% accuracy rate of a model without median filtering. However, there are few issues regarded the long training time and the small number of samples, the time cost for 5000 epochs is 10948 seconds. As a result, the future work will be focused on increasing the number of characters samples of this model, optimizing the architecture as well as algorithm of the network to reduce computational time, increase the computational efficiency with methods like incorporating hybridization into the architecture of the network to reduce the processing complexity.

Acknowledgment
We would like to express our sincere thanks to Institute of Automation, Chinese Academy of Sciences for which they provide the database for training the model. This work was supported by Guangxi Natural Science Foundation Project: Research of Mangrove Pest Management based Semi-Continuous Dynamic System Theory (project no: 2016GXNSFAA380102)

Reference
[1] Ahlawat, S., & Choudhary, A. (2020). Hybrid CNN-SVM Classifier for Handwritten Digit Recognition. Procedia Computer Science, 167, 2554-2560.
[2] Ptucha, R., Such, F. P., Pillai, S., Brockler, F., Singh, V., & Hutkowski, P. (2019). Intelligent character recognition using fully convolutional neural networks. Pattern recognition, 88, 604-613.
[3] Cireşan, D., & Meier, U. (2015, July). Multi-column deep neural networks for offline handwritten Chinese character classification. In 2015 international joint conference on neural networks (IJCNN) (pp. 1-6). IEEE.

[4] Zhou, M. K., Zhang, X. Y., Yin, F., & Liu, C. L. (2016). Discriminative quadratic feature learning for handwritten Chinese character recognition. Pattern Recognition, 49, 7-18.

[5] He, Z., Zhong, Y., & Cao, Y. (2013). High Accuracy Handwritten Chinese Character Recognition Based on Support Vector Machine and Independent Component Analysis. In Informatics and Management Science V (pp. 725-733). Springer, London.

[6] Zhang, X. Y., Bengio, Y., & Liu, C. L. (2017). Online and offline handwritten chinese character recognition: A comprehensive study and new benchmark. Pattern Recognition, 61, 348-360.

[7] Zhong, Z., Jin, L., & Xie, Z. (2015, August). High performance offline handwritten Chinese character recognition using googlenet and directional feature maps. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR) (pp. 846-850). IEEE.

[8] Liu, X., Zhou, Y., & Wang, Z. Deep neural network-based recognition of entities in Chinese online medical inquiry texts. Future Generation Computer Systems, 114, 581-604.

[9] Xiao, X., Jin, L., Yang, Y., Yang, W., Sun, J., & Chang, T. (2017). Building fast and compact convolutional neural networks for offline handwritten Chinese character recognition. Pattern Recognition, 72, 72-81.

[10] Kavitha, B. R., & Srimathi, C. (2019). Benchmarking on offline Handwritten Tamil Character Recognition using convolutional neural networks. Journal of King Saud University-Computer and Information Sciences.

[11] Chen, J., Kang, X., Liu, Y., & Wang, Z. J. (2015). Median filtering forensics based on convolutional neural networks. IEEE Signal Processing Letters, 22(11), 1849-1853.

[12] Niu, Y., Zhao, Y., & Ni, R. (2017). Robust median filtering detection based on local difference descriptor. Signal Processing: Image Communication, 53, 65-72.

[13] Rhee, K. H. (2019). Forensic detection using bit-planes slicing of median filtering image. IEEE Access, 7, 92586-92597.

[14] Wu, Y. C., Yin, F., & Liu, C. L. (2017). Improving handwritten Chinese text recognition using neural network language models and convolutional neural network shape models. Pattern Recognition, 65, 251-264.

[15] Zhuang, Y., Chuah, J. H., Chow, C. O., & Lim, M. G. (2020, November). Iris Recognition using Convolutional Neural Network. In 2020 IEEE 10th International Conference on System Engineering and Technology (ICSET) (pp. 134-138). IEEE.