# Classification of Handwritten Chinese Numbers with Convolutional Neural Networks


1st Rasoul Ameri
*Faculty of Computer Engineering*
*Shahrood University of technology*
Shahrood, Iran
rasoul.ameri@yahoo.com

2nd Ali Alameer
*School of Biological Sciences*
*Queen's University Belfast*
Belfast, UK
a.alameer@qub.ac.uk

3rd Saideh Ferdowsi
*School of Computer Science & Electronic Engineering*
*University of Essex*
Colchester, UK
s.ferdowsi@essex.ac.uk

4th Vahid Abolghasemi
*School of Computer Science & Electronic Engineering*
*University of Essex*
Colchester, UK
v.abolghasemi@essex.ac.uk

5th Kianoush Nazarpour
*School of Informatics*
*University of Edinburgh*
Edinburgh, UK
kianoush.nazarpour@ed.ac.uk



***Abstract*—Deep learning methods have become the key ingredient in the field of computer vision; in particular, convolutional neural networks (CNNs). Appropriating the network architecture and data pre-processing have significant impact on performance. This paper focuses on the classification of handwritten Chinese numbers. Firstly, we applied various methods of pre-processing to our collected image dataset. Secondly, we customised a CNN-based architecture with minimal number of layers and parameters specifically for the task. Experimental results showed that our proposed methods provides superior classification rate of 99.1%. Our results also show that the proposed method has competitive performance compared to smaller neural networks with fewer parameters, e.g. Squeezenet and deeper networks with a larger size and number of parameters, e.g., pre-trained GoogLeNet and MobileNetV2.**



***Index Terms*—Chinese number classification, Convolutional neural network, Deep learning, Image processing, hand written recognition.**


## I. Introduction

Handwritten recognition, in a broad sense, is a well studied pattern classification problem and has attracted much attention due to wide range of applications [1], [2]. By definition, it refers to ability of machines to recognise and interpret a written text by human. Numerous applications such as reading postal addresses, bank check amounts, car plates, and books can be benefited by this intelligent technique [3], [4]. Nevertheless, it is challenging to design a high performance system due to various handwritten types and languages by individuals around the world. This paper addresses a particular and yet less studied handwritten classification type, i.e. Chinese numbers classification.

Despite the popularity and importance of Chinese characters, there is little work carried out on the classification of handwritten Chinese numbers. One reason could be due to the lack of a comprehensive database. Recently, we have established a wealthy database of this type and proposed a dictionary learning based technique for this purpose [5]. Our open source database paves the path for further research on improvement of Chinese number recognition systems.

The traditional handwritten number classification methods work on three major steps. In the first step, the image enhancement and noise removal is normally applied. These include methods such as RGB-to-grayscale conversion, size normalization, etc. The second step involves extracting features from the image of interest, e.g. histogram of oriented gradients (HOG) [6]. In the final step, the extracted features are classified by using traditional classifiers such as support vector machine (SVM) [7]–[9] and k-nearest neighbor (kNN) [10], [11]. In [12], the authors used three classification methods to recognise Arabic characters handwriting. They used HOG to extract the features; classification was performed using SVM and kNN. In [13], the authors proposed a system for handwriting recognition based on SVM classifier. They used statistical methods for feature extraction and the SVM for classification.

Experimental results suggest that deep learning methods outperform traditional machine learning in a large margin. The CNN structure was first proposed by Fukushima in [14] and then LeCun et al. applied gradient based learning to CNN to classify handwritten numbers [15]. Deep learning achieves promising performance in computer vision [16], [17], natural language processing [18], [19], and speech recognition [20], [21]. Convolutional neural network (CNN) plays a crucial role in solving various computer vision problems. By definition, CNN is a multi-layer neural network that learns formative features at each layer of its hierarchy. Thus, in this paper we propose to employ this novel method for Chinese number recognition.

Various methods based on deep learning have been proposed for recognition of handwritten numbers and characters. In [22], researchers explored the performance of convolutional neural network for addressing this problem, such LeNet, AlexNet, ZfNet and VGGNet. They used ICDAR2013 dataset consisting of 200 classes. In [23], researchers proposed a deep convolutional neural network model by minimizing that intra-class variation and maximizing the inter class variation. They used a DB1.1 database and 2015 ICDAR Chinese handwriting recognition. Researchers in [24] proposed method base on CNN for recognizing Arabic digits dataset. MADBase database was used for their analysis.

Despite the complexity in the characters of the Chinese numbers, the literature shows reduced effort to solve problems associated with its recognition and classification. This is mainly due to the lack of a standard dataset. We recently collated a large dataset from a large number of native Chinese people specifically for this task. As a result, we pioneered a method to classify handwritten Chinese number classification using dictionary learning [5]. Following that work, here we have designed a CNN method for handwritten Chinese number recognition. To optimize performance, we propose the method consists of two phases; a) image processing: we first proposed Otsu's method to smooth the images, we then adapted the images in accordance with our CNN input channels. b) CNN: we customized a CNN architecture with appropriate number of layers, including convlutional/max pooling layers and classification based on CNN method.

In next section, we describe the proposed method including the dataset description and preparation, image pre-processing, and neural network structure. Section 3 is devoted to experimental settings and the corresponding results. Finally, the paper is concluded in Section 4.

## II. Material and Method

### A. Dataset

Our handwritten Chinese number dataset consisting of 15,000 samples was collected at Newcastle University [25]. It includes 15 Chinese numbers, where the numbers written by 100 people in different handwriting. The participants wrote numbers on a standart A4 sheet which was scanned at the resolution of 300×300 pixels. The size of all images is 64-by-64. One sample of handwritten Chinese number can be seen in Fig. 1.

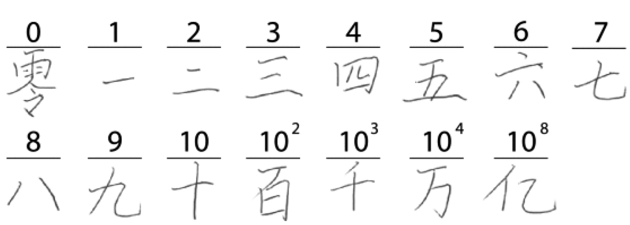



Fig. 1: Simplified example of handwritten Chinese numbers. The equivalent English numbers are given on the top of each handwriting.

### B. Image pre-processing

The preprocessing phase includes two parts; image enhancement and noise removal. Initially, each scanned image

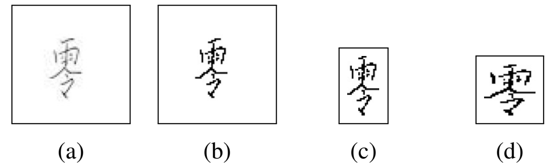



Fig. 2: Example result of image processing step for Chinese number '0'. (a) the original grayscale image of size 64×64; (b) binarized image using Otsu's method; (c) cropped image; (d) resized image to 32×32. The images are made negative for ease of presentation.

is converted to gray-scale, then, a global image threshold is determined by using Otsu's method to convert the image from gray-scale to binary image (shown in Fig. 2b). The cropping operation is performed on the image so that each number in an image can be fitted in a rectangular box that is equal in size to the width and height of the number. The result of cropping operation is shown in Fig. 2c. In the last step, images are resized to 32×32 pixels as shown in Fig. 2d.

### C. Classification using CNN

In this paper, a customized CNN architecture is proposed. It includes a set of empirically selected combination of layers consisting of convolutional, max-pooling, and fully connected layers. The network architecture design is illustrated in Fig. 3. The architecture of CNN consists of two parts: feature extraction and a soft-max layer that outputs a probability distribution. For feature extraction, we propose a set of convolution and max-pooling layers. In the convolution layer, features from previous layers are convolved with learnable kernel. The max-pooling layer performs the downsampled operation on the input layers. The fully connected layer computes the score of each class from extracted feature in the preceding steps. In the final softmax layer, a score vector that outputs a probability distribution is generated; the class with the highest score is used for classification.

Below we provide a detailed description of the layers/parameters that constitute our proposed CNN. The first layer is a convolutional layer with kernel size "5×5" pixels

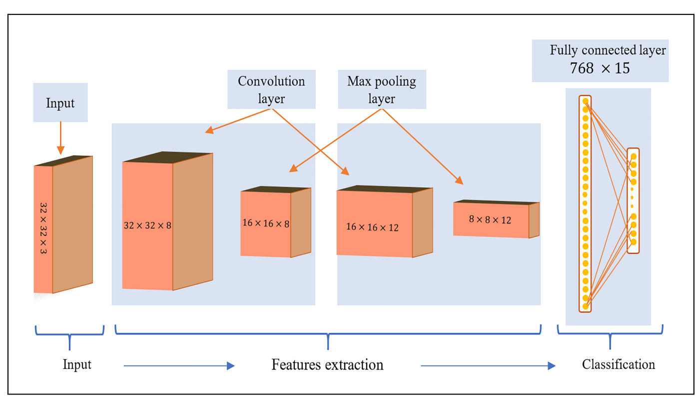
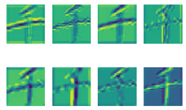
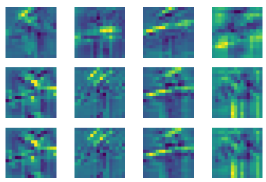



Fig. 3: Visualization of the proposed CNN architecture.

(a) Input image.



(b) Conv.1 outputs.



(c) Conv.2 outputs.

Fig. 4: Output of CNN layers for $10^3$. The input image size is $32\times32$. The output of the first and second layers consist of 8 and 12 convolutional images of size $32\times32$ and $16\times16$, respectively.

and 8 output channels. The second layer is a max-pooling layer with “2x2” kernel size. The third layer is a convolutional layer with kernel size “5x5” pixels and 12 output channels. The fourth layer is a max-pooling layer with “2x2” kernel size. The following layers are fully connected neural layer with 768 – 15 neurons in each layer. As an illustrative example the output of each CNN layers for Chinese number $10^3$ is given in Fig. 4. It is noted that all programs are implemented using Python language and Pytorch library [31]. The results are generated on a machine equipped with Intel core i7 with 2.20 GHz, NVIDIA GeForce 1050 graphics card and 8GB of memory.

## III. Experimental results

Chinese numbers were classified via a 5-fold cross validation. Eighty percent of all samples were used for training and twenty percent for testing. There are 1000 samples per classes in dataset. At each trial, 800 samples were used for training and 200 samples for test.

The Stochastic gradient decent (SGD) optimizer is used for training [26]. The SGD approach needed a short training time compared to traditional gradient decent. In this paper, SGD is used for training with momentum 0.95 and learning rate 0.002. The selected momentum value accelerated to accelerate the training process with the SGD approach.

Rectified linear unit (ReLU) was used for activation function [27]. The training procedure was stopped after 12 epochs as shown in Fig. 5.

In Fig. 6 confusion matrix for all 15 classes is illustrated. It is seen that the number of correct classification by the trained network for all classes is more than 196 (out of 200), meaning

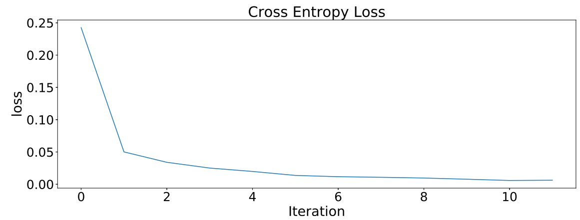



Fig. 5: Cross entropy loss in training stage.

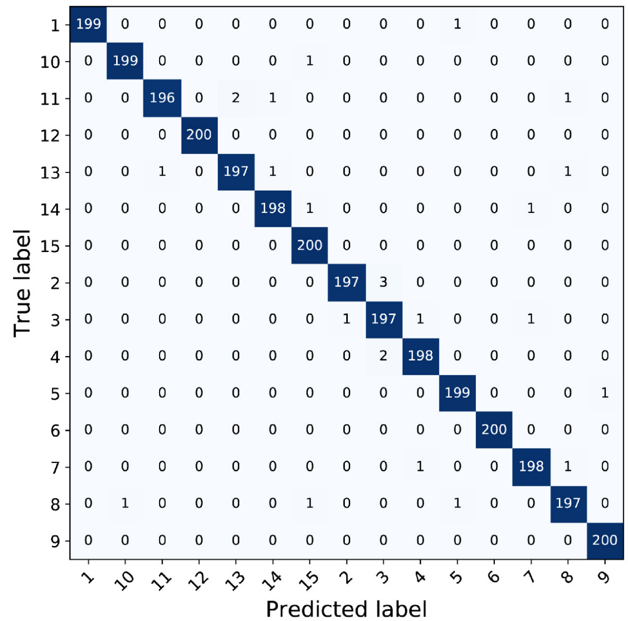



Fig. 6: Confusion matrix; vertical and horizontal axes show respectively the true and predicted labels. The diagonals and off-diagonals represent the number of correctly-classified and misclassified samples, respectively.

the 98% classification accuracy. For example 197 samples of class 13 are correctly classified as class 13 and one sample of class 13 are incorrectly of class 8, classes 11 and 14. Misclassification happened between class 2 and 3; this is due to the observation that upper section of the number of 2 is similar to that of 3.

Table I demonstrates the obtained results with CNN on Chinese number classification. The results are better than those reported in [5] which was based on dictionary learning. The proposed method obtained 0.1% misclassification error on training data and 0.9% misclassification error on test data.

TABLE I: Classification accuracy of different method for Chinese number classification.

| Method | Accuracy (%) |
|---|---|
| InDPL [5] | 93 |
| The proposed method | 99.1 |

TABLE II: Classification accuracy of different deep models for handwritten Chinese numbers.

| | GoogleNet [28] | MobileNetV2 [29] | SqueezeNet [30] | Proposed method |
|---|---|---|---|---|
| Accuracy | 99.83% | 99.77% | 98.55% | 99.1% |

For completeness, in table II we compare the proposed models with two powerful deep learning platforms, GoogleNet [28] and MobileNetV2 [29]. To replicate results under the same experimental environment, we applied our proposed image processing method prior to feeding to these networks. The accuracy difference between our proposed method and GoogleNet and MobileNetV2 is 0.73% and 0.67%, respectively. Despite the large size and number of parameters of these networks (i.e., 1.27, 7, and 3.5 million parameters in SqueezeNet, GoogleNet, and MobileNetV2, respectively), our proposed network provided competitive performance. On the other hand, the number of calculated parameters of the proposed model is much lower than GoogleNet and MobileNetV2. Also, we compare our proposed method with a lesser deep network, SqueezeNet that well-known CNN architecture compatible with small computers [30]. SqueezeNet produced an average classification accuracy of 98.55% over 5-folds cross validation, which makes our customized CNN fits nicely in the literature.

## IV. Conclusion

Deep learning so far has proven to be a powerful tool for pattern classification. Our proposed image processing method and CNN architecture showed dramatic increase in classification accuracy. Furthermore, combining the proposed image processing method with large sized CNN architectures shows slight improvement in comparison with our proposed minimal architecture. Finally, our proposed architecture is optimised for real-time scenarios, e.g., mobile application, without relying on GPUs or cloud computing.

## References


[1] J. Sueiras, V. Ruiz, A. Sanchez, and J. F. Velez, "Offline continuous handwriting recognition using sequence to sequence neural networks," Neurocomputing, vol. 289, pp. 119–128, 2018.

[2] X. I. A. Mei-Qun et al., "Managing real-time handwriting recognition." Google Patents, 09-Jul-2019.

[3] D. Keysers, T. Deselaers, H. A. Rowley, L.-L. Wang, and V. Carbune, "Multi-language online handwriting recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1180–1194, 2016.

[4] D. Hu, "Research and application of handwritten numeral recognition method," SM thesis, Univ. Nanchang, Nanchang, China, 2012.

[5] V. Abolghasemi, M. Chen, A. Alameer, S. Ferdowsi, J. Chambers, and K. Nazarpour, "Incoherent dictionary pair learning: Application to a novel open-source database of Chinese numbers," IEEE Signal Process. Lett., vol. 25, no. 4, pp. 472–476, 2018.

[6] R. Ebrahimzadeh and M. Jampour, "Efficient handwritten digit recognition based on histogram of oriented gradients and SVM," Int. J. Comput. Appl., vol. 104, no. 9, pp. 10–13, 2014.

[7] A. Boukharouba and A. Bennia, "Novel feature extraction technique for the recognition of handwritten digits," Appl. Comput. Informatics, vol. 13, no. 1, pp. 19–26, 2017.

[8] M. N. Ayyaz, I. Javed, and W. Mahmood, "Handwritten character recognition using multiclass svm classification with hybrid feature extraction," Pakistan J. Eng. Appl. Sci., 2016.

[9] E. Tuba and N. Bacanin, "An algorithm for handwritten digit recognition using projection histograms and SVM classifier," in 2015 23rd Telecommunications Forum Telfor (TELFOR), 2015, pp. 464–467.

[10] Y. Wang, R. Wang, D. Li, D. Adu-Gyamfi, K. Tian, and Y. Zhu, "Improved Handwritten Digit Recognition using Quantum K-Nearest Neighbor Algorithm," Int. J. Theor. Phys., pp. 1–10, 2019.

[11] N. Ilmi, W. T. A. Budi, and R. K. Nur, "Handwriting digit recognition using local binary pattern variance and K-Nearest Neighbor classification," in 2016 4th International Conference on Information and Communication Technology (ICoICT), 2016, pp. 1–5.

[12] M. Salam and A. A. Hassan, "Offline isolated arabic handwriting character recognition system based on SVM.," Int. Arab J. Inf. Technol., vol. 16, no. 3, pp. 467–472, 2019.

[13] N. A. Hamid and N. N. A. Sjarif, "Handwritten recognition using SVM, kNN and neural network," arXiv Prepr. arXiv1702.00723, 2017.

[14] K. Fukushima, "Neocognitron: A hierarchical neural network capable of visual pattern recognition," Neural networks, vol. 1, no. 2, pp. 119–130, 1988.

[15] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[16] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?," IEEE Trans. image Process., vol. 24, no. 12, pp. 5017–5032, 2015.

[17] M. Boroumand and J. Fridrich, "Deep learning for detecting processing history of images," Electron. Imaging, vol. 2018, no. 7, pp. 1–9, 2018.

[18] L. Deng and Y. Liu, Deep Learning in Natural Language Processing. Springer, 2018.

[19] M. Gardner et al., "Allennlp: A deep semantic natural language processing platform," arXiv Prepr. arXiv1803.07640, 2018.

[20] T. Afouras, J. S. Chung, A. Senior, O. Vinyals, and A. Zisserman, "Deep audio-visual speech recognition," IEEE Trans. Pattern Anal. Mach. Intell., 2018.

[21] D. Amodei et al., "Deep speech 2: End-to-end speech recognition in english and mandarin," in International conference on machine learning, 2016, pp. 173–182.

[22] R. G. Maidana, J. M. Dos Santos, R. L. Granada, A. D. M. Amory, and R. C. Barros, "Deep neural networks for handwritten Chinese character recognition," Proc. - 2017 Brazilian Conf. Intell. Syst. BRACIS 2017, vol. 2018-Janua, pp. 192–197, 2017.

[23] C. Cheng, X. Y. Zhang, X. H. Shao, and X. D. Zhou, "Handwritten Chinese character recognition by joint classification and similarity ranking," Proc. Int. Conf. Front. Handwrit. Recognition, ICFHR, vol. 0, pp. 507–511, 2016.

[24] A. El-Sawy, E.-B. Hazem, and M. Loey, "CNN for handwritten arabic digits recognition based on LeNet-5," in International Conference on Advanced Intelligent Systems and Informatics, 2016, pp. 566–575.

[25] "Handwritten Chinese Numbers." [Online]. Available: $https://data.ncl.ac.uk/articles/Handwritten_Chinese_Numbers/1028083.$ [Accessed: 29-Nov-2019].

[26] L. Bottou, "Stochastic gradient descent tricks," in Neural networks: Tricks of the trade, Springer, 2012, pp. 421–436.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105. 0.5 MB model size," arXiv Prepr. arXiv1602.07360, 2016.

[28] D. Amodei et al., "Deep speech 2: End-to-end speech recognition in english and mandarin," in International conference on machine learning, 2016, pp. 173–182.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510–4520.

[30] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and 0.5MB model size," arXiv:1602.07360.

[31] "PyTorch." [Online]. Available: https://pytorch.org/. [Accessed: 28-Nov-2019].