Offline Handwritten Chinese Character Recognition Based on New Training Methodology

Weike Luo and Guangtao Zhai(✉)

Department of Electronic Engineering,
Shanghai Jiao Tong University, Shanghai, China
{lwk9419,zhaiguangtao}@sjtu.edu.cn

Abstract. Deep learning based methods have been extensively used in Handwritten Chinese Character Recognition (HCCR) and significantly improved the recognition accuracy in recent years. Famous networks like GoogLeNet and deep residual network have been applied to this field and improved the recognition accuracy. While the structure of the neural network is crucial, the training methodology also plays an important role in deep learning based methods. In this paper, a new data generation method is proposed to increase the size of the training database. Chinese characters could be classified into different kinds of structures according to the radical components. Based on this, the proposed method segments the character images into sub-images and recombines them into new character samples. The generated database, including recombined characters and rotated characters, could improve the performance of current CNN models. We also apply the recently proposed and popular center loss function to further improve the recognition accuracy. Tested on ICDAR 2013 competition database, the proposed methods could achieve new state-of-the-art with a 97.53% recognition accuracy.

Keywords: Offline handwritten Chinese character recognition
Deep learning · Data generation · Center loss

1 Introduction

According to the types of input data, HCCR can be divided into online and offline recognition. For offline recognition, the main target is to extract the feature from the grayscale images and classify them into different categories. It plays an important role in mail sorting, handwritten notes reading and handwritten input on mobile devices. Due to its high practicality, this research has received content attention and plenty methods have been proposed to improve the recognition accuracy.

Nowadays, deep learning based methods have become more and more popular in this field. As shown in a recent report [1], all of the top-ranked methods are using deep neural networks. Among these works, the multi-column deep learning

G. Zhai et al. (Eds.): IFTC 2017, CCIS 815, pp. 235–244, 2018.
https://doi.org/10.1007/978-981-10-8108-8_22

network (MCDNN) is believed to be the first successfully applied CNN model in HCCR [2]. The team of Fujitsu won the first place of the ICDAR 2013 by using an alternately trained relaxation convolutional neural network (ATR-CNN) [10]. After that, Zhong et al. [3] proposed a streamlined version of GoogLeNet [4] which is 19 layers deep and further improved the performance. Recently, Zhong et al. [9] applied the deep residual network [12] which is 34 layers deep and has become the state-of-the-art method.

While various CNN models have been proposed to improve the recognition accuracy, most of these methods only focus on the structure of neural network. The training methodology plays a crucial role in deep learning based methods. Existing works of training methodologies include new loss function, training data generation, the modification of back propagation algorithm, etc. Compared to traditional methods, convolutional neural networks provide a powerful solution to extract feature from raw images directly. Instead of designing complex methods for preprocessing, the current trend is to provide a large database for training which includes the generated samples. The input images are manually distorted and rotated to make the neural network more stable to the input data. Inspired by this, we design a new method to recombine the input images based on the structural information of the character. Most of the complex Chinese characters could be regarded to own specific structure like top-bottom structure and left-right structure. To generate new training samples, the original character image is first segmented into different parts according to the structure and then recombined within the same category to generate new samples. The rotated images are also generated for training. The total number of the generated data accounts for twice as big as the original database.

To further improve the performance, we also add the center loss function. As far as we know, most of the existing works in HCCR only use softmax layer to train the network. As current networks are designed to become deeper and deeper, simply using the softmax layer to train the network could not get an excellent result. Researchers start to design new loss functions to help the neural networks converge faster and get better performance. The center loss function is designed for classification which has a large number of categories. The extracted feature of the input images will be pulled towards the center of the category in the feature space. The whole network is jointly trained by softmax loss and center loss. As a result, while the inter-class distance trained by the softmax loss becomes large, the intra-class distance is reduced by the center loss. With a benefit of the intra-class distances and easy implementation, the center loss function soon becomes popular in face recognition.

The experiments are conducted on 2013 ICDAR competition database. The results show the proposed methods have 97.53% recognition accuracy and achieve the new state-of-the-art result.

To summarize, our main contributions are summarized as follows:

(1) We propose a new training data generation method based on the structural information of Chinese character. Compared to the previous character

generation methods, our methods could generate characters with different writing styles within the same category.
(2) We apply the center loss function to deep residual network in HCCR. The center loss significantly improves the training efficiency and recognition accuracy.

The rest of this paper is organized as follows: Sect. 2 introduces a previous data generation method in HCCR. The details of the proposed methods are described in Sect. 3. Section 4 presents the experimental results. The conclusion is shown in Sect. 5.

2 Related Work

Among the recent works, Chen et al. [6] are believed to be the first one to generate the new training data for deep learning based methods in HCCR. In their work, both local distortion and global distortion are added to the input character image. The local distortion is to add a small displacement to the grayscale input image including the X, Y coordinates and gray value. The global distortion is to rotate the image by a small random angle. Since then, plenty of works start to increase the size of training database. Most of the recently proposed methods use the rotated input images to improve the performance of the CNN model.

However, to keep the shape of the character, the displacement and rotation angle must be carefully designed and kept in a small range. Therefore, the generated samples are quite similar to the original ones. The methods only consider the displacements on the input images without making use of the relationship between the characters in the same category. The results show the influence of the generated data on the training is limited. To solve this problem, we design a novel method to generate new training data beyond single input image.

3 Proposed Method

3.1 Radical Region Based Data Generation

The size of the training database plays a key role in deep learning based methods. As shown in face recognition, the bigger database is more likely to get a higher recognition accuracy. Therefore, a great number of methods are proposed to increase the size of training database and make the network more stable to the distorted input. For HCCR, the input data is hard to be aligned without classifying the character. Compared to designing complex traditional methods for normalization, it is more efficient to provide various data to train the neural network. The proposed method is to generate data by recombining different radical regions of the character.

Chinese character, as the logogram, is composed of different radical components, which are the graphical components of the character. These radical components need to be written in certain regions. According to the position, Chinese character could be classified into different structures as shown in Fig. 1.

Structure of Chinese Character	Examples
Single-element Character	一 白 才
Left-right Structure	胡 跟 帆
Left-mid-right Structure	湖 摊 树
Top-bottom Structure	罗 李 昊
Top-mid-bottom Structure	爱 蒙 囊
Semi-enclosed Structure	句 床 司
Enclosed Structure	围 国 回
Pyramid structure	森 淼 晶

Fig. 1. The structural information of Chinese characters

Since different people write Chinese character in different writing styles, the samples would contain various kinds of distortion. To reduce or enlarge the radical components deliberately is one of the ways to make the handwritten character more beautiful. This is also one of the reasons why HCCR is more difficult than printed ones. Therefore, it is essential to generate enough data from samples, which contain different writing styles, to make the convolutional neural network more stable to different writing styles.

According to Fig. 1, the Chinese character could be divided into four categories: the single-element character, center-border character, the two-part character (top-bottom and left-right structure), the three-part character (pyramid, left-mid-right and top-mid-bottom structure). Each of the characters is firstly manually classified into the four categories based on the location of the radical components and then segmented into sub-images except for the single-element character as shown in Figs. 2 and 3. For example, the character with left-right structure is segmented into two parts: left region and right region. Then the segmented images are then recombined with the sub-images of the same character. After normalization, the generated images are resized to the original size for training.

Besides the recombined data, the rotated data is also created for training. For each data, including the original data and recombined data, the image is rotated by a small random angle between -10° and 10° as shown in Fig. 4.

Therefore, the training database is composed of three parts: the original database, the recombined database and the rotated database. The size of the recombined database and rotated database is the same as the original database.

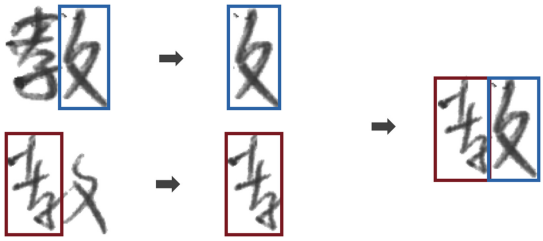

Fig. 2. An example of the left-right recombined data

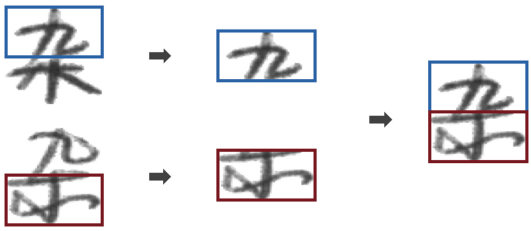

Fig. 3. An example of the top-bottom recombined data

Fig. 4. An example of the rotated data

3.2 Center Loss

Softmax loss is the most widely used loss function in neural network. With current networks become deeper and deeper, simply using softmax loss shows its limitation. In order to get faster convergence and better results, plenty of loss functions have been proposed in the past few years.

The center loss is proposed to reduce the intra-class distances and make it easier to separate different classes [11]. High practicality and easy implementation make center loss soon become popular in face recognition. As far as we know, Yang et al. [13] was the first to apply center loss to HCCR and proposed a light CNN model. The effect of the center loss can be represented as Fig. 5. The star represents the center of the feature of the same category. It is believed that if the feature is tightly clustered around the center point, it will be easier for classification.

In training, the output feature $\mathbf{x_i} \in R^d$ of each input image i is pulled towards the center of each category $\mathbf{c_{y_i}}$. Parameter d is the feature dimension while y_i represents the label of input image i. The formula of the center loss is shown as below:

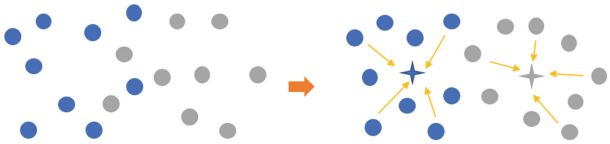

Fig. 5. The effect of center loss

$$\mathcal{L}_c = \frac{1}{2}\sum_{i=1}^{m} \|\mathbf{x_i} - \mathbf{c_{y_i}}\|_2^2 \tag{1}$$

It is impossible to calculate the centers of categories in the whole database due to the limitation of memory. Thus, the center $\mathbf{c_{y_i}}$ is updated within the mini-batch in the training.

However, all of the feature will be pulled to the zeros if only center loss is used. In this way, the intra-class distance will be the smallest while the network could not classify the characters. Therefore, the center loss and softmax loss are jointly used in practice. The formula of the softmax loss could be represented as below:

$$\mathcal{L}_s = -\sum_{i=1}^{m} log\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}} \tag{2}$$

where n represents the number of class. $\mathbf{W}_j \in R^d$ is the jth column of the weights $\mathbf{W} \in R^{d \times n}$ in the last fully connected layer and $\mathbf{b} \in R^n$ is the bias term.

The whole loss function could be represented by the following equation. The parameter λ is used to balance the loss functions.

$$\mathcal{L} = \mathcal{L}_s + \lambda\mathcal{L}_c = -\sum_{i=1}^{m} log\frac{e^{W_{y_i}^T \mathbf{x_i} + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T \mathbf{x_i} + b_j}} + \frac{\lambda}{2}\sum_{i=1}^{m} \|\mathbf{x_i} - \mathbf{c_{y_i}}\|_2^2 \tag{3}$$

3.3 Details of Implementation

In experiments, the deep residual network (DRN) is used for feature extracting. Compared to traditional CNN models, the DRN adopts residual learning to every few stacked layers which is realized by a residual building block defined as:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x} \tag{4}$$

The $\mathbf{x}$ and $\mathbf{y}$ are the input and output of the layers with the same dimension. Function $\mathcal{F}(\mathbf{x}, \{W_i\})$ represents the residual mapping to be learned and $\{W_i\}$ is the parameters to be optimized.

The whole network is the original 34 layers deep model [12] which is the most widely used one. Softmax loss and center loss are jointly used to train the network. To balance the softmax loss and center loss, the parameter λ is set as 0.008. The general architecture of the CNN is shown in Fig. 6

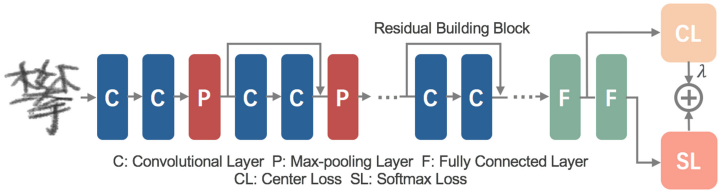

Fig. 6. A general architecture of the proposed network

Considering the influence of the image size on the recognition accuracy and the storage cost, the images are firstly shuffled and then resized into 120×120. We start with a learning rate of 0.02, divide it by 2 per 40k iterations. The maximum iteration is set to 550k. The training takes 2 days to get the best result. Input images are the grayscale images without any preprocessing and the whole network is an end-to-end method.

4 Experiment

4.1 Experimental Data

The databases come from the Institute of Automation of Chinese Academy of Sciences (CASIA) [5]. CASIA-HWDB 1.0 and CASIA-HWDB 1.1 are used for training with a total number of 2,678,424. All of the samples are in 3,755 categories of GB2312-80 level-1 set which covers most of the contemporary used characters. The test database is from the competition test database which contains 224,419 samples comes from 60 writers [7].

4.2 Experimental Results

We test the deep residual network (DRN), center loss (CL) and generated database (GD) separately. The experimental results show the proposed methods could further improve the performance of the current model. The best model improves the recognition accuracy by 0.28% compared to the original model as shown in Table 1.

Table 1. Results on Deep Residual Model

Method	Accuracy
DRN	97.29%
DRN + GD	97.43%
DRN + CL	97.41%
DRN + CL + GD	**97.53%**

Table 2. Results on ICDAR-2013 Offline HCCR Competition Database

Method	Accuracy	Ensemble	Memory
Human Performance [7]	96.13%	n/a	n/a
Traditional Method: DFE+DLQDF [1]	92.72%	no	120.0 MB
MCDNN [2]	95.79%	yes(8)	349.0 MB
ATR-CNN Voting [10]	96.06%	yes(4)	206.5 MB
HCCR-Gabor-GoogLeNet [3]	96.35%	no	27.77 MB
HCCR-Gabor-GoogLeNet-Ensemble [3]	96.74%	yes(10)	270.0 MB
CNN-Single [6]	96.58%	no	190.0 MB
CNN-Voting [6]	96.79%	yes(5)	950.0 MB
DirectMap-ConvNet [1]	96.95%	no	23.50 MB
STN-Residual-34 [9]	97.37%	no	92.30 MB
Proposed Method	**97.53%**	**no**	**124.8 MB**

Table 2 shows the comparison with current methods [9]. Both the recognition accuracy and the memory cost are shown in this table. There is only one end-to-end model in our proposed methods without any ensemble. The results show that the proposed methods could achieve new state-of-the-art performance on the test database.

We also compare the influence of the size of the generated data. The recombined data and rotated are tested separately in five groups. As shown in Fig. 7, the recombined data and the rotated data could further improve the performance of the current model. Both of them have a positive relationship with the recognition accuracy. However, too many generated data would influence the performance on the original database.

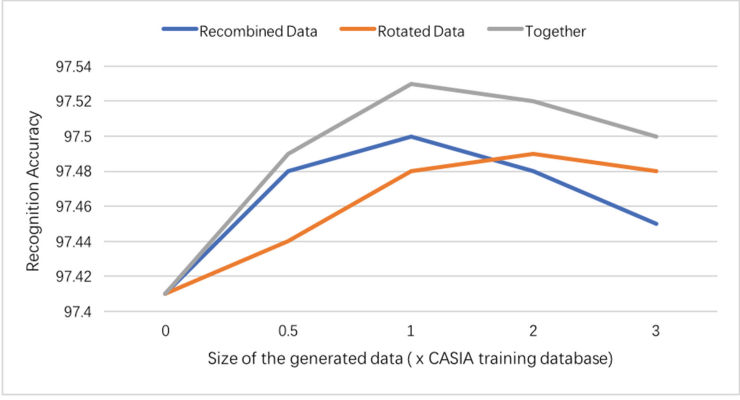

Fig. 7. The recognition accuracy with different size of generate database

4.3 The Recombined Database

The recombined database plays a key role in our methods. Compared to the rotated database, the recombined database is more effective as shown in Fig. 7. Figure 8 shows examples of the generated recombined database. The proposed method could get a good result even for some complex characters.

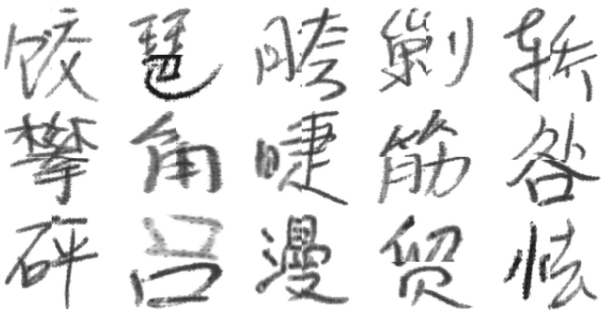

Fig. 8. Recombined database

However, due to the various writing styles, the recombined database also contains some invalid samples as shown in Fig. 9. Most of these samples are removed from the recombined database manually while a small percentage of these invalid samples are kept. We believe these invalid samples could still contribute to improving the performance for it could reduce the over-fitting problem.

Fig. 9. Invalid samples

5 Conclusion

In this paper, two training methods for offline HCCR are proposed to further improve the performance of the current model. A radical region based training data generation method is proposed to increase the size of the training data. The generated data could recombine samples with different writing styles. The center loss function is used in the residual model in HCCR. Both of the methods could improve the performance of current models on the ICDAR 2013 offline HCCR competition database. The best result comes to 97.53% which is the new state-of-the-art as we know.

The CASIA database also gives the human performance (96.13%) on it, which shows there is a limitation of this database. The wrong and similar characters are even impossible for human beings to recognize. By far, the performance of

the neural network is far beyond the human performance. We will continue to improve the recognition accuracy until we find the real limitation of it. While deep CNN model is the key in HCCR nowadays, we believe the semantic meaning behind the character will become crucial in the future.

References

1. Zhang, X.Y., Bengio, Y., Liu, C.L.: Online and offline handwritten chinese character recognition: a comprehensive study and new benchmark. Pattern Recogn. **61**, 348–360 (2017)
2. Cireşan, D., Meier, U.: Multi-column deep neural networks for offline handwritten Chinese character classification. In: 2015 International Joint Conference on Neural Networks (IJCNN), pp. 1–6. IEEE (2015)
3. Zhong, Z., Jin, L., Xie, Z.: 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 (2015)
4. Szegedy, C., Liu, W., Jia, Y., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
5. Liu, C.L., Yin, F., Wang, D.H., et al.: CASIA online and offline Chinese handwriting databases. In: 2011 International Conference on Document Analysis and Recognition (ICDAR), pp. 37–41. IEEE (2011)
6. Chen, L., Wang, S., Fan, W., et al.: Beyond human recognition: a CNN-based framework for handwritten character recognition. In: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), pp. 695–699. IEEE (2015)
7. Yin, F., Wang, Q.F., Zhang, X.Y., et al.: ICDAR 2013 Chinese handwriting recognition competition. In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR), pp. 1464–1470. IEEE (2013)
8. Jia, Y., Shelhamer, E., Donahue, J., et al.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 675–678. ACM (2014)
9. Zhong, Z., Zhang, X.Y., Yin, F., et al.: Handwritten Chinese character recognition with spatial transformer and deep residual networks. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 3440–3445. IEEE (2016)
10. Wu, C., Fan, W., He, Y., et al.: Handwritten character recognition by alternately trained relaxation convolutional neural network. In: 2014 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 291–296. IEEE (2014)
11. Wen, Y., Zhang, K., Li, Z., et al.: A discriminative feature learning approach for deep face recognition. In: Leibe, B., et al. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 11–26. Springer International Publishing, Heidelberg (2016). https://doi.org/10.1007/978-3-319-46478-7_31
12. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: Computer Vision and Pattern Recognition, pp. 770–778. IEEE (2016)
13. Yang, S., Nian, F., Li, T.: A light and discriminative deep networks for off-line handwritten Chinese character recognition. In: Youth Academic Conference of Chinese Association of Automation, pp. 785–790 (2017)