Convolution Neural Networks for Offline Chinese Handwriting Recognition

Okke van der Wal | u956749
Tilburg University
o.a.vdrwal@tilburguniversity.edu

Abstract

In this research, we explored the performance of convolutional neural network on recognizing offline handwritten Chinese characters. We compared two different convolutional networks and our purpose was to demonstrate the effect of adding an extra convolutional layer. Comparison of the experimental results demonstrate that a deeper network with an extra layer do not necessarily result in a better validation accuracy.

1 Introduction

In general, humans can easily recognize media such as characters, images and sounds. For computers, this is often much more difficult, due to various factors, including the variations in shape, which makes it complicated for the machine to distinguish characters from each other. However, recent developments have shown that machines have become better able to classify certain images or characters. Moreover, deep convolutional neural network (CNN) has become the architecture of choice for these types of problems.

There has been a lot of research on using deep CNN to recognize handwritten digits, English alphabet, or the Latin alphabet. Compared to recognising handwritten digits and handwritten English alphabet, the recognition of handwritten Chinese characters is a more challenging task due to various reasons. First, there are many more classes for Chinese characters than for digits and English characters. In comparison, there are 10 digits for digit recognition tasks and there are 26 characters for the English alphabet, while there are in total over 50,000 Chinese characters. Approximately 3,000 are for everyday use. Second, Chinese characters have more complicated structures compared to digits or English characters. Finally, the handwritten character can vary significantly, depending on the writer. Figure 1 portrays an example of the Chinese character, 且, and it shows how the example in the dataset can deviate from the original character and illustrates the difficulty of classifying Chinese characters.

Figure 1. Handwriting samples from the dataset

In 2010, the National Laboratory of Pattern recognition (NLPR), Institute of Automation of Chinese Academy of Sciences (CASIA) released a database of online and offline handwritten characters [1]. Studies on Handwritten Chinese character recognition therefore focus on either online or offline recognition. Offline recognition is often considered to be more difficult, since online recognition is stroke trajectory-based and offline recognition is image-based. As well as in English alphabet as in Chinese alphabet, stroke trajectory can help better predict a character, especially when characters are similar [2].

The introduction of the deep convolutional neural network has improved the accuracy of offline recognition significantly, as demonstrated by various research projects [2]. Despite advances, there remain various challenges, particularly in the recognition of individual Chinese characters. Zhang [3] proposed several convolutional neural network architectures to solve this. Due to time and computational constraints, the purpose of this research is to explore the effect of using two different simplified versions of the networks used by Zhang. In general, adding an extra layer to a CNN often improves the performance of a network [5]. However, previous research has demonstrated that adding an extra convolutional layer might not improve the performance of the networks for recognizing offline Chinese characters [4]. The research question addressed in this paper therefore reads as follows: To which extent could offline Chinese characters be classified correctly using a simplified Deep Convolutional Neural Network and what is the effect of adding an extra convolutional layer?

The paper is organized as follows: we will first introduce the dataset and our network configurations. Then we will explain how we implement and train our network. Afterwards we present our experimental results and discuss these.

2 Architecture

To explore the performance of deep convolutional neural networks on classifying handwritten Chinese characters, and answer the research question, two different simplified versions of the architecture described by Yuhao Zhang [3] were used. The input image fed into the network is a fixed-size 64x64 grayscale image. The network is a multi-layer architecture consisting of fully connected layers with a softmax classifier. The image is passed through several convolutional layers, each of which has 3x3 filters. Table 1 portrays the various layers used in both architectures of the network.

<table>
<tr><th>Model A</th><th>Model B</th></tr>
<tr><td>5 weight layers</td><td>weight layers</td></tr>
<tr><td>Conv3-64</td><td>Conv3-64</td></tr>
<tr><td colspan="2">Maxpool</td></tr>
<tr><td>Conv3-128</td><td>Conv3-128</td></tr>
<tr><td colspan="2">Maxpool</td></tr>
<tr><td>Conv3-256</td><td>Conv3-256</td></tr>
<tr><td rowspan="2"></td><td>Maxpool</td></tr>
<tr><td>Conv3-512</td></tr>
<tr><td colspan="2">Maxpool</td></tr>
<tr><td colspan="2">FC-1024</td></tr>
<tr><td colspan="2">FC-200</td></tr>
<tr><td colspan="2">Softmax</td></tr>
</table>

Table 1: Network configurations

As portrayed in Table 1, both architectures of the network consist of at least five weight layers. This choice was made since a convolutional network with less than five weight layers is often not conclusive enough to achieve a high accuracy on the handwritten Chinese character offline recognition task, as concluded by previous experiments [4]. In order to keep the size of the output, we consequently fixed the stride size to 1 pixel and the zero-padding size to 1 pixel. Implementing small filter sizes reduces the number of parameters in the network and makes the networks more efficient to train, while not comprising for performance of the network. This is the same approach used successfully in other papers [5].

During training, the inputs to our network are 64×64 grayscale images, as discussed before. Spatial pooling is carried out by max-pooling layers denoted by "maxpool" in Table 1. The max-pooling is carried out over a 2×2 pixels window, with stride of 2 pixel. Since the network includes at least three convolutional layers, we therefore use at least three max-pooling layers. The difference between the first version and the second version is that the second version of the network has an extra convolutional layer. This to explore the effect of adding an extra layer on the performance of the network.

3 Experimental Method

3.1 Data

For this project, we used the CASIA offline database, released in 2011 by the Chinese Academy of Science [1], which consists of online and offline isolated handwritten Chinese character images. More specifically, we used the HWDB1.1 dataset, which totally includes 3,755 Chinese characters and 171 alphanumeric and symbols. Each class contains handwritten images from approximately 240 different writers and each writer contributed one image to each Character class.

3.2 Experiment

The network configuration is a simplified version of the one described in Deep Convolutional Network for Handwritten Chinese Character Recognition by Yuhao Zhang [3]. Training on the full dataset, as in the research by Zhang, would not be within the scope of this experiment. Therefore, due to constraints by time and computation resources, the experiment was run on a subset of the full dataset. Table 2 portrays the information for the subset and the full dataset.

Dataset	#Classes	#Training Examples (per class)	#Validation Examples (per class)	#Total Examples
Subset	200	200	40	48,000
Full	3,755	200	40	759,000

Table 2: Training set and validation set information for the subset and the full dataset

3.3 Hyper-parameters

In running the experiment, various parameters were used to train the network. After several experiments, the optimal number of epoch was determined to be 15 epochs. In training a network one cannot pass the entire dataset into the network at once. The training set for this experiment consisted of 4000 samples and the validation set consisted of 8000 samples. In the cross-validation split, we ensured that the number of examples per character was equally divided, so there would be no differences for individual characters. The optimal batch size was determined to be 128. All convolutional layers in the model used a rectified linear unit as error threshold. Finally, a learning rate of 0.01, Adam as optimizer and categorical cross entropy as its loss function were used.

3.4 Evaluation

The dataset was split in a training set and a validation set and the network is therefore evaluated by the validation accuracy. Moreover, the network is evaluated by the ability to correctly assign the right Chinese character class to a certain input image. Also, the network will be evaluated by their loss, to ensure that network does not overfit.

4 Results

Experiments were run following the network architectures and classification settings introduced in the previous section. The classification results for the 200-class subset is reported in Table 3. The table demonstrates that Model A has achieved a better performance, compared to Model B. It achieves a higher validation accuracy, meaning that on unseen data, it is better able to correctly classify to which Chinese character a certain input image belongs. In addition, the training loss and validation loss is lower for Model A. In general, the lower the loss, the better.

Model	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
A	0.9688	0.1046	0.9141	0.3735
B	0.9457	0.1703	0.9044	0.3946

Table 3: Results for both versions of the network

Figure 2 and Figure 3 portray the development of the validation accuracy for both networks over time, as more epochs are being used. What is interesting to see is that Model A takes a relatively big leap in accuracy at the start of the training phase. One explanation may be that the samples included in the batch size are not evenly distributed. However, the network was trained several iterations, which did not change the performance of the network over time.

Model A:

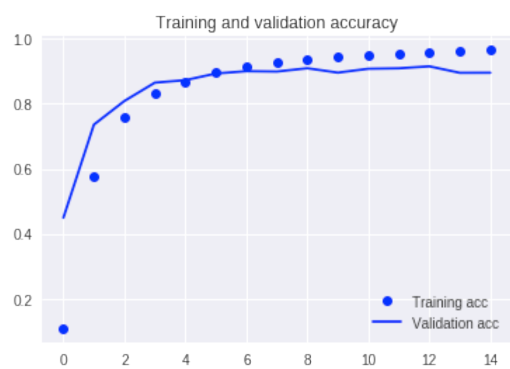

Figure 2: Training and validation accuracy for Model A, as the number of epochs increases. The y-axis shows the accuracy of the network and the x-axis shows the number of epochs.

Model B:

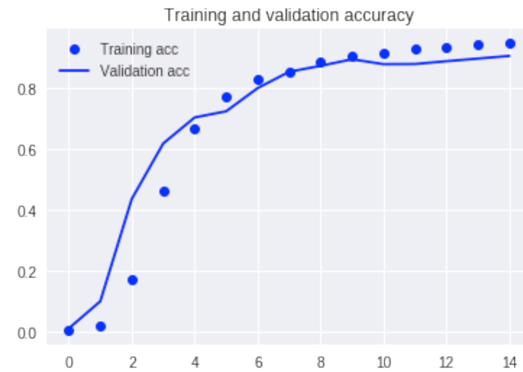

Figure 3: Training and validation accuracy for Model B, as the number of epochs increases. The y-axis shows the accuracy of the network and the x-axis shows the number of epochs.

5 General Discussion

As portrayed in Table 3, the best validation accuracy is achieved by the first version of the network. If we further examine the results achieved by previous research, we can see that there is no accuracy gain, when adding an extra convolutional layer. This is an interesting observation and it aligns with the conclusion from the paper by Zhang [3], which also concluded that adding an extra convolutional layer does not necessarily improve the performance of a convolutional neural network on the CASIA offline database.

The performance of previous CNN-based handwritten characters depended mainly on the budget for computational resources and available time. However, as our results show, adding an extra layer does not necessarily result in a better performance. One should however note that, previous experiments would train on the full CASIA offline database, instead of on a subset as in this experiment [4]. Thus, even though the conclusion is interesting it is less generalizable since we ran the experiment on a subset of the full dataset.

6 Conclusion

In this project, we explored the problem of recognizing handwritten Chinese characters. Specifically, we use a deep convolutional neural network, based on the paper by Zhang. Based on the results, we conclude that a simplified version of the experiment by Zhang can achieve impressive results with a 91% accuracy on the validation set. This result should be interpreted with caution, since the network was trained on a subset of the complete dataset. This subset consisted of 200 characters instead of the 3,755 characters the full dataset is comprised of. Nonetheless, for recognizing offline Chinese characters, using a convolutional neural network yields impressive results. This aligns with previous research that implemented convolutional neural networks.

The research question addressed in this paper read as follows: To which extent could offline Chinese characters be classified correctly using a simplified Deep Convolutional Neural Network and what is the effect of adding an extra convolutional layer? Based on our experiments, we conclude convolutional neural networks can achieve impressive results and that adding an extra convolutional layer does not necessarily improve the results for recognizing offline Chinese characters.

References

[1] C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang. CASIA online and offline Chinese handwriting databases. In Document Analysis and Recognition (ICDAR), 2011 International Conference on, pages 37–41. IEEE, 2011.

[2] Sanjiv Kumar Singh, D. (2015). Review of Online & Offline Character Recognition. *International Journal Of Engineering And Computer Science*, *4*(05).

[3] Zhang, Y. Deep convolutional network for handwritten chinese character recognition. *Computer Science Department, Stanford University*.

[4] Liu, Cheng-Lin & Yin, Fei & Wang, Da-Han & Wang, Qiu-Feng. (2013). Online and offline handwritten Chinese character recognition: Benchmarking on new databases. Pattern Recognition. 46. 155–162. 10.1016/j.patcog.2012.06.021.

[5] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.