Pattern Recognition 61 (2017) 348–360

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Online and offline handwritten Chinese character recognition: A comprehensive study and new benchmark

Xu-Yao Zhang [a,*], Yoshua Bengio [b], Cheng-Lin Liu [a,c]

[a] NLPR, Institute of Automation, Chinese Academy of Sciences, China
[b] MILA, University of Montreal, Canada
[c] CAS Center for Excellence in Brain Science and Intelligence Technology, China

ARTICLE INFO

Article history:
Received 30 April 2016
Received in revised form
5 August 2016
Accepted 6 August 2016
Available online 9 August 2016

Keywords:
Handwriting recognition
Chinese characters
Online
Offline
Directional feature map
Convolutional neural network
Adaptation

ABSTRACT

Recent deep learning based methods have achieved the state-of-the-art performance for handwritten Chinese character recognition (HCCR) by learning discriminative representations directly from raw data. Nevertheless, we believe that the long-and-well investigated domain-specific knowledge should still help to boost the performance of HCCR. By integrating the traditional normalization-cooperated direction-decomposed feature map (directMap) with the deep convolutional neural network (convNet), we are able to obtain new highest accuracies for both online and offline HCCR on the ICDAR-2013 competition database. With this new framework, we can eliminate the needs for data augmentation and model ensemble, which are widely used in other systems to achieve their best results. This makes our framework to be efficient and effective for both training and testing. Furthermore, although directMap+convNet can achieve the best results and surpass human-level performance, we show that writer adaptation in this case is still effective. A new adaptation layer is proposed to reduce the mismatch between training and test data on a particular source layer. The adaptation process can be efficiently and effectively implemented in an unsupervised manner. By adding the adaptation layer into the pre-trained convNet, it can adapt to the new handwriting styles of particular writers, and the recognition accuracy can be further improved consistently and significantly. This paper gives an overview and comparison of recent deep learning based approaches for HCCR, and also sets new benchmarks for both online and offline HCCR.

1. Introduction

Handwritten Chinese character recognition (HCCR) has been studied for more than fifty years [1,2] to deal with the challenges of large number of character classes, confusion between similar characters, and distinct handwriting styles across individuals. According to the type of input data, handwriting recognition can be divided into online and offline. In online HCCR, the trajectories of pen tip movements are recorded and analyzed to identify the linguistic information expressed [3], while in offline HCCR, character (gray-scaled or binary) images are analyzed and classified into different classes. Offline HCCR finds many applications, such as mail sorting [4], bank check reading, book and handwritten notes transcription, while online HCCR has been widely used for pen input devices, personal digital assistants, smart phones, computer-aided education, and so on. Moreover, HCCR is also an important integral part for handwritten text recognition (both online [5] and offline [6]) which considers segmentation and recognition simultaneously. High character recognition accuracy is essential for the success of handwritten text/string recognition [7].

To promote academic research and benchmark on HCCR, the National Laboratory of Pattern Recognition (NLPR), Institute of Automation of Chinese Academy of Science (CASIA), has organized three competitions at CCPR-2010 [8], ICDAR-2011 [9], and ICDAR-2013 [10]. The results of competition show improvements over time and involve many different recognition methods. An overwhelming trend is that deep learning based methods gradually dominate the competition. From the very beginning, all submitted systems at CCPR-2010 were traditional methods. In ICDAR-2011, the team of IDSIA from Switzerland submitted their system [11] based on convolutional neural network (convNet) and won the first place on offline HCCR. This is the first work on using convNet for HCCR. Later for ICDAR-2013, both the winners of online and offline HCCR were using convNets. The team from Fujitsu R&D Center used a 4-convNet voting method to win the competition of offline recognition, while the team from University of Warwick

* Corresponding author.
E-mail addresses: xyz@nlpr.ia.ac.cn (X.-Y. Zhang), yoshua.bengio@umontreal.ca (Y. Bengio), liucl@nlpr.ia.ac.cn (C.-L. Liu).

http://dx.doi.org/10.1016/j.patcog.2016.08.005
0031-3203/

used a sparse convNet [12] to win the competition of online recognition.

Deep learning methods can directly learn discriminative representations [13] from raw data, and therefore can provide end-to-end solutions for many pattern recognition problems. However, the well-studied domain-specific knowledge is shown to be still helpful for further improving the performance [14,15] of HCCR. The most important domain knowledge of HCCR includes the character shape normalization and direction-decomposed feature maps. The character recognition community has proposed many useful shape normalization methods such as nonlinear normalization [16], bi-moment normalization [17], pseudo 2D normalization and line density projection interpolation [18]. Shape normalization can reduce the within-class variations and hence increase the recognition accuracy [19]. Another important domain knowledge is the direction-decomposed feature map. By decomposing the gradient (for offline image) or the local stroke (for online stroke trajectory) into different directions (from 0° to 360°), we can obtain multiple feature maps, each representing a direction of original gradient/stroke. This is a strong prior knowledge of Chinese character which is produced by basic directional strokes during writing process. Representing Chinese character as directional features had been the state-of-the-art method [19–21] for a long time before the arrival of convNet.

To improve the accuracies of HCCR, instead of training convNet from raw data, we represent both the online and offline handwritten characters by the normalization-cooperated [22] direction-decomposed feature maps (directMap), which can be viewed as a $d \times n \times n$ sparse tensor (d is the number of quantized directions and n is the size of the map). DirectMap contains the domain-specific knowledge of shape normalization and direction decomposition, and hence is a powerful representation for HCCR. Furthermore, inspired by the recent success of using very deep convNet for image classification [23–25], we developed an 11-layer convNet for HCCR. By combining directMap with convNet, we are able to obtain new benchmarks for both online and offline HCCR on the ICDAR-2013 competition database [10]. Previous works usually adopt different methods to obtain best performance for online and offline HCCR separately. However, with directMap+convNet, we are able to achieve state-of-the-art performance for both online and offline HCCR under the same framework. Due to the embedded domain-specific knowledge, we can also eliminate the needs of data augmentation and model ensemble, which are crucial for other systems to achieve their best performance. This makes our model to be efficient and effective for both the training and testing processes.

The large variability of handwriting styles across individuals is another challenge for HCCR. Writer adaptation [26,27] is widely used to handle this challenge by gradually reducing the mismatch between writer-independent system and particular individuals. Although deep learning based methods have set a high record for HCCR which already surpass human-level performance, we show that writer adaptation in this case is still effective. Inspired from our early work on style transfer mapping [28], we add a special adaptation layer in the convNet to match and eliminate the distribution shift between training and test data in an unsupervised manner. The adaptation can guarantee performance improvements even when only a small number of samples are available, due to the regularization involved in the learning process. During our experiments on 60 writers for both online and offline HCCR, we observed consistent and significant increase of accuracies by the adaptation of the convNet.

The handwriting recognition community has reported many useful and important achievements (from the year of 1980 to 2008) by previous overview papers of [3,29–33]. Nowadays, the deep learning based approaches become the new cutting-edge technology for solving handwriting related problems. This paper can be viewed as an overview of recent progresses (especially through the three competitions [8–10]) in using deep learning methods for the task of handwritten Chinese character recognition (HCCR). The results and comparisons reported here can be used as new benchmarks for future researches in the field of both online and offline HCCR.

The rest of this paper is organized as follows. Section 2 reviews related works. Section 3 describes the procedures for generating online and offline directMaps. Section 4 shows the evolution from traditional methods to convNet. Section 5 introduces the details of the convNet used in our system. Section 6 explains how to add an adaptation layer in convNet for writer adaptation. Section 7 reports the experimental results, and Section 8 draws concluding remarks.

2. Related works

With the impact from the success of deep learning [34,35] in different domains, the solution for HCCR has been changed from traditional methods to convolutional neural networks (convNet) [36]. The first reported successful use of convNet for HCCR (offline) was the multi-column deep neural network (MCDNN) [37,38]. After that, the sparse convNet [39] was used to achieve the best performance for online HCCR in ICDAR-2013 competition. Alternately trained relaxation convolutional neural network was proposed by [40] for offline HCCR. Recently, the highest accuracy for offline HCCR was achieved by [41] through integrating multiple strategies such as local and global distortions, multi-supervised training, and multi-model voting. ConvNet has also been successfully used for handwritten Hangul recognition [42] which is similar to HCCR. Although these methods have outperformed traditional methods by large margins, they are based on end-to-end learning which ignores the long-and-well studied domain-specific knowledge in HCCR.

Recently, [15] combined the traditional feature extraction methods such as Gabor and gradient feature maps with the GoogLeNet [24] to obtain very high accuracy for offline HCCR. Moreover, for online HCCR, [14] and [43] achieved the best performance by using convNet with various domain knowledge including deformation, imaginary stroke map, path signature map, and directional map. These results clearly identify the advantages of using domain knowledge for further improving performance. It should be noted that in the application of deep learning to most image classification tasks, the generation of distorted images for augmenting the training data is also a kind of utilization of domain knowledge. However, in our mind, the most important domain-specific knowledge should be shape normalization and direction decomposition. With our proposed directMap+convNet, we can achieve new benchmarks for both online and offline HCCR, without the help from data augmentation or model ensemble, which are crucial for [15,43] to obtain their best results.

Deep learning based methods have also found applications in other handwriting related problems, such as writer identification [44], hybrid model [45], confidence analysis [46], handwritten legal amounts recognition [47], and text spotting [48]. The convNet can also be combined with the hidden Markov model (HMM) for online handwriting recognition [49]. Recently, the recurrent neural network (RNN) with long-short term memory (LSTM) [50] has been successfully used for handwritten Chinese text recognition without explicit segmentation of characters [51]. The combination of RNN and convNet has also been used for scene text reading by [52,53]. It is evident that more and more character recognition related problems will turn their attention to deep learning methods for high performance solutions.