Pattern Recognition 58 (2016) 190–203

Contents lists available at ScienceDirect

Pattern Recognition

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

DropSample: A new training method to enhance deep convolutional neural networks for large-scale unconstrained handwritten Chinese character recognition

Weixin Yang [a], Lianwen Jin [a,*], Dacheng Tao [b], Zecheng Xie [a], Ziyong Feng [a]

[a] College of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640, China
[b] Centre for Quantum Computation & Intelligent Systems, University of Technology, Sydney, Australia

ARTICLE INFO

Article history:
Received 24 March 2015
Received in revised form
20 March 2016
Accepted 13 April 2016
Available online 23 April 2016

Keywords:
Convolutional neural network
Deep learning
Handwritten character recognition
Domain-specific knowledge

ABSTRACT

Inspired by the theory of Leitner's learning box from the field of psychology, we propose *DropSample*, a new method for training deep convolutional neural networks (DCNNs), and apply it to large-scale online handwritten Chinese character recognition (HCCR). According to the principle of *DropSample*, each training sample is associated with a quota function that is dynamically adjusted on the basis of the classification confidence given by the DCNN softmax output. After a learning iteration, samples with low confidence will have a higher frequency of being selected as training data; in contrast, well-trained and well-recognized samples with very high confidence will have a lower frequency of being involved in the ongoing training and can be gradually eliminated. As a result, the learning process becomes more efficient as it progresses. Furthermore, we investigate the use of domain-specific knowledge to enhance the performance of DCNN by adding a domain knowledge layer before the traditional CNN. By adopting *DropSample* together with different types of domain-specific knowledge, the accuracy of HCCR can be improved efficiently. Experiments on the CASIA-OLHDWB 1.0, CASIA-OLHWDB 1.1, and ICDAR 2013 online HCCR competition datasets yield outstanding recognition rates of 97.33%, 97.06%, and 97.51% respectively, all of which are significantly better than the previous best results reported in the literature.

1. Introduction

A traditional isolated online handwritten Chinese character recognition (HCCR) method typically employs the following framework: (1) pre-processing of an input handwritten character (e.g., linear or non-linear normalization [2] and addition of imaginary strokes [3–7]), (2) feature extraction (e.g., 8-directional feature extraction [6] and discriminative directional feature extraction [8]), and (3) classification via machine learning methods (e.g., modified quadratic discriminant function (MQDF) [9,10], support vector machine [11], and hidden Markov model (HMM) [12]). In contrast, deep learning methods [13–17], which have attracted a considerable amount of research and industry attention in recent years, deviate from the above-mentioned framework by providing an alternative end-to-end solution to HCCR without any dedicated feature extraction or pre-processing technique, enabling potentially high performance. Owing to the availability of large-scale training data, new training technologies (e.g., Dropout [18], DropConnect [19], layer-wise pre-training [20]), and advanced computing hardware platforms (e.g., GPU [21]), the convolutional neural networks (CNNs) originally proposed by LeCun in the 1990s [22,23] have been extensively investigated in recent years. The traditional CNN has been extended with deeper architectures (e.g., [24,25]; we refer to this variant of CNN as Deep CNN (DCNN) in this paper), advanced training technologies, and effective learning algorithms (e.g., [18,40]) to address the various challenges posed by computer vision and pattern recognition problems. Consequently, significant breakthroughs have been achieved, such as image recognition [24–27], facial recognition [28,29], handwriting recognition [13,15,18,30], pose recognition [31], text detection, and natural scene image recognition [32–36]. Furthermore, DCNN with different structures has been successfully applied to the field of HCCR field [13–17], which poses a major challenge because it involves a large vocabulary (e.g., as many as 3755 classes for the GB2312-80 level-1 standard), many similar and confusable characters, and different writing styles with unconstrained cursive techniques [37].

The performance of current DCNNs is highly dependent on the greedy learning of model parameters via many iterations on the basis of a properly designed network architecture with abundant labeled training data. Most DCNN models treat all the training

* Corresponding author. Tel.: +86 13802994591.
E-mail address: lianwen.jin@gmail.com (L. Jin).

http://dx.doi.org/10.1016/j.patcog.2016.04.007
0031-3203/

Table 1
Explanation of several psychological terminologies mentioned in this paper.

Terminology	Interpretation
Forgetting curve	It describes the exponential loss of information that one has learned [64]
Recency effect	It describes the increased recall of the most recent information because it is still in the short-term memory [64]
Spacing effect	It is the phenomenon whereby animals more easily remember of learned items when they are studied a few times spaced over a long time span rather than repeatedly studied in a short span of time [64]
Spacing repetition	It is a learning technique that incorporates increasing intervals of time between subsequent reviews of previously learned material in order to exploit the psychological spacing effect [39]

samples uniformly during the entire learning process. However, we have found that the error reduction rate during the learning process is initially high but decreases after a certain number of training iterations. This may be attributed to most of the samples being well recognized after a certain number of training iterations; thus, the error propagation for adapting the network parameters is low, while confusable samples, which are difficult to learn and account for a relatively low ratio of the training dataset, do not have a high likelihood of contributing to the learning process. Some previous DCNNs consider this phenomenon as a signal to manually reduce the learning rate, or as a criterion for early stopping [38], thereby neglecting the potential of the confusing samples that have thus far been insufficiently well-learnt.

Inspired by the theory of Leitner's learning box from the field of psychology [1], we propose a new training method, namely *DropSample*, to enhance the efficiency of the learning process of DCNN, and we employ it to solve the challenge of cursive online HCCR. The explanation of several relevant psychological terminologies are presented in Table 1. The principle of spacing repetition is useful in many contexts, but it requires a learner to acquire a large number of items and retain them in memory indefinitely. Leitner's learning box is designed as a simple implementation of the principle of spacing repetition for learning [39]. A direct application of Leitner's learning box theory is that material that is difficult to learn will appear more frequently and material that is easy to learn will appear less frequently, with difficulty defined according to the ease with which the user is able to produce a correct learning response. The *DropSample* training method proposed in this paper adopts a similar concept to design a learning algorithm for DCNN. To this end, each training sample is assigned to a box with a quota function that is dynamically adjusted according to the classification confidence given by the DCNN softmax output. After a learning mini-batch iteration, samples with high confidence in this mini-batch will be placed in a box with low appearance frequency, whereas those with low confidence will be placed in another box with high appearance frequency, thus they are more likely to appear for selection as the training data after a short learning interval. A certain amount of noisy data (e.g., mislabeled samples and outliers) always exists in the training dataset, and such data may be useful in the initial stages of training to avoid overfitting; however, they gradually prevent the network from achieving high prediction accuracy. They should therefore be placed in a box with low appearance frequency, and thus can gradually be eliminated. A schematic of *DropSample* is shown in Fig. 1.

To address the specific challenge of online HCCR, we propose the incorporation of domain-specific technologies in DCNN models to account for domain-specific information that may be useful but cannot be learnt by the DCNN. By employing the new training technology of *DropSample* together with various types of domain-specific knowledge, we find that recognition accuracy can be improved significantly. In this way, we obtain a single network that achieves test error rates of 2.99% and 3.43% on CASIA-OLHWDB 1.0 and CASIA-OLHWDB 1.1 [41], respectively, both of which are lower than the state-of-the-art results reported in previous studies [13,15,37]. Furthermore, the different types of domain-specific knowledge contribute to corresponding DCNN classifiers, which may be complementary and can therefore be integrated to achieve better accuracy. The ensemble ultimately reduces the test error rates to 2.67% on CASIA-OLHWDB 1.0, 2.94% on CASIA-OLHWDB 1.1, and 2.49% in the case of the ICDAR 2013 online HCCR competition dataset [16], which are significantly better than the best results reported in previous studies [13,15,16,37]; this confirms the effectiveness of the proposed *DropSample* training method.

The remainder of this paper is organized as follows. Section 2 introduces related studies reported in the literature. Section 3 presents the proposed DCNN architecture and the configurations employed. Section 4 provides a detailed description of the *DropSample* training method. Section 5 describes the domain-specific knowledge. Section 6 presents the experimental results and analysis. Lastly, Section 7 summarizes our findings and concludes the paper.

2. Related work

Over the past four decades, numerous studies have investigated HCCR [42], resulting in the development of many techniques such as non-linear normalization, data augmentation, directional feature extraction, and quadratic classifier modification. Non-linear normalization methods such as modified centroid-boundary alignment (MCBA) [43], line density projection interpolation (LDPI) [2], and line density projection fitting (LDPF) [43], as well as their pseudo-two-dimensional extensions [2,44], are based on line density equalization that can reduce within-class variation of character shape. Data augmentation techniques enhance insufficient training data by generating various handwriting styles via affine transformation [45], cosine functions [46], distorted generation [47], deformation transformation [48], and style consistent perturbation [57]. Two of the most popular feature extraction methods for representation are 8-directional feature extraction [6,10] and discriminative directional feature extraction [8,49]. By employing classifiers that use the modified quadratic discriminant function (MQDF) [9,10], or its variant such as discriminative learning quadratic discriminant function (DLQDF) [37], a traditional HCCR system can yield fairly good recognition performance.

In recent years, CNNs have produced outstanding results in the fields of machine learning and pattern recognition. The idea of CNN was first proposed by Fukushima [50] in 1980. It was formally developed by LeCun et al. [22,23] and improved by Simard et al. [51], Cireşan et al. [13,14], and others. GPU acceleration hardware [21] has facilitated the development of deep CNN (DCNN), which includes a deeper architecture with additional convolutional layers. DCNN offers several advantages. For example, it enables integrated training of feature extractors and classifiers to provide systematic optimization. In addition, it does not require pre-training and provides useful properties such as availability of raw data, effective feature extraction, and excellent generalization capability [52].

CNN has acquired a reputation for solving many computer vision problems in recent years, and its application to the field of HCCR has been shown to provide significantly better results than traditional methods [13–17]. The multi-column deep neural network (MCDNN) method proposed by Cireşan et al. [13] shows remarkable ability in many applications and attains near-human performance on handwritten datasets such as MNIST. In addition, it has provided promising results for HCCR. Graham proposed a variation of CNN called DeepCNet [15], which won first place at the ICDAR 2013 online HCCR competition [16]. By combining the path signature feature and employing a spatially sparse architecture, DeepCNet produced a best test error rate of 3.58% [15] on CASIA-OLHWDB 1.1, which is lower than that achieved by MCDNN (5.61%) [13] and DLQDF (5.15%) [37].