Offline Handwritten Text Recognition Using Deep Learning: A Review

Yintong Wang1,2, Wenjie Xiao2 and Shuo Li2*

1 College of Computer Science, Zhejiang University, Hangzhou, 310058, China
2 School of Information Engineering, Nanjing Xiaozhuang University, Nanjing, 211171, China
* ch.nj.ls@njxzc.edu.cn

Abstract. The area of offline handwritten text recognition(OHTR) has been widely researched in the last decades, but it stills an important research problem. The OHTR system has an objective to transform a document image into text data. Compared with online handwriting recognition, the dynamic information about the writing trajectories in OHTR is not available. Many advancements have been proposed in the literature, most notably the application of deep learning methods to OHTR. In this paper, we introduced how this problem has been handled in the past few decades, analyze the latest advancements and the potential directions for future research in this field.

1. Introduction
Offline handwritten text recognition is considered one of the earliest computer vision problems to be handled by many researches[1-3]. Since its inception as a field of research in more than a half century now, researchers never stopped studying on it. The reason can be generally contributed to the following two aspects. First, the application demand of OHTR is growing rapidly[4, 5], including handwritten manuscript recognition, bank form recognition and historical document processing. Second, the long-standing complicated nature of the OHTR itself[6, 7], including the variability of writing styles, a huge amount of character types and the complicated structures of texts.

This paper reviews the existing offline handwritten text recognition methods using deep learning, and hopes to provide some reference for the further development of this problem. We focus on the methods based on deep neural network learning as they have been the state-of-the-art for the last decades. Especially these methods based on Recurrent Neural Networks(RNNs) and Convolutional Neural Networks(CNNs) have dominated all OHTR problems and became the de facto standard for the problem[8].

The rest of this paper is organized as follows. Section 2 reviews the framework of offline handwritten text recognition. In Section 3, we summarize these methods of offline handwritten text recognition. In Section 4, we introduce the common experiment datasets for offline handwritten text recognition. Finally, we draw the conclusions and future works in Section 5.

2. Framework of Offline Handwritten Text Recognition
Usually, OHTR can be defined as the process of a recognition system to transform a text image data into its equivalent character representation, and then be processed and stored in the form of ASCII text. Typically, OHTR system consists of three main parts: preprocessing, features extraction and...
classification[6, 9]. Figure 1 gives the flowchart of the offline handwritten text recognition. In the preprocessing stage, we can improve and enhance the quality of the offline handwritten text image for next appropriate analysis[10]. It should be emphasized that it can be decomposed into many smaller tasks, such as paragraph detection, text-line segmentation, word/character segmentation, image normalization and so on.

In the features extraction stage, we extract the representative features from text image to ensure that these features can be used to acquiring the good performance of the classification system. According to whether the representative features are relevance to the classification tasks without retraining, it includes the hand-crafted feature-based methods[1] and the automatic learning feature-based methods[11]. The formers are very limited as they require prior knowledge about information on the features' position and relevance. Obviously, these approaches are vulnerable with the variation of handwriting styles, background colours and other nonuniform problems of text image. CNNs and RNNs as the two representative techniques of the latter, these techniques fully utilize the deep neural network learning architecture to gain automatically the representative features, and solves the problems of translation, scaling and distortion to become one of the most robust systems. However, there is a generally drawback for automatic learning feature-based methods requiring more computing resource in the training phase to acquire the representative features.

In the classification stage, the representative features are inputted to a trained classifier, which can predict the character/word class. There are many approaches have been applied to recognize the text image, which generally divided into segmentation-free approach[12] and segmentation-based approach[13]. The segmentation-based approach as a traditional method performs explicit segmentation of text-line image into many individual characters before employing a trained classifier to recognize the individual character's class. It is worth noting that the performance of these approaches is a highly correlation of the performance of word/character segmentation, and any segmentation error will be accumulated and directly affect the recognition accuracy of the classifier. By contrast, the segmentation-free approach, the word/line/multi-lines image as input data, allows to recognize the document image without performing explicit segmentation. It seems that these approaches are the most commonly used especially when the separations between lines/characters are hard to determine, such as complex background, touching text-lines, overlapping characters. These methods are commonly used methods are CNNs and RNNs combined with hidden Markov model(HMM) or connectionist temporal classification(CTC). With the continuous deepening of text recognition research, they are considered to perform better on using the contextual information, this information includes the output calculated at each time step based on the past and future context of the connected characters or words.

3. Offline Handwritten Text Recognition Using Deep Learning
As we know, most of the state-of-the-art works use CNNs, RNNs or their hybrid architectures to performing text recognition tasks at present. This Section introduces these existing methods in OHTR from character recognition, word/line recognition and multi-lines recognition.

3.1. Character Recognition Methods
Achievements in deep neural network learning enable researchers to successfully use CNNs in the OHTR domain, and then greatly outperforming modified quadratic discriminant functions(MDQF) methods[1]. The multi-column deep neural networks[14, 15] is the first reported method to successfully apply CNNs to OHTR. Wu et al.[16] proposed an alternately trained relaxation convolutional neural networks for OHTR. Zhong et al.[9, 11] subsequently used a shallow version of GoogLeNet to integrating traditional Gabor features with offline handwriting character as input data.
Li et al. [17] introduced an occluded offline handwriting character recognition method using improved GoogLeNet and deep convolutional generative adversarial network to recognize. Wang et al. [18] proposed radical analysis network with densely connected architecture to utilize the offline handwriting Chinese character two-dimensional structures and its radicals. Unlike most of the above methods based on CNNs, Wang et al. [19] introduced an encoder-decoder architecture, namely radical aggregation network, to utilize the radical-level composition of offline handwriting Chinese characters. Although the accuracy of the above method for character recognition is close to the level of human understanding, they reported an accuracy of 96.74% on HWDB1.0 and HWDB1.1 dataset, deep neural networks approaches must to face high calculation cost and a large number of parameters.

To solve the above problem, Zhou et al. [20] proposed an novel framework, which applied a Kronecker fully connected layer to replace the layers after the four inception groups. Zhang et al. [4] introduced an voting module to combine CNNs and traditional normalization-cooperated direction-decomposed feature maps. Xiao et al. [21] proposed global supervised low-rank expansion to accelerate calculations in the network convolutional layers. The experimental results shown that it effective reduce the neural networks calculate cost and compress the network size with only a little drop in classification accuracy. Li et al. [13] introduced weighted average pooling to balance the network parameters' number and the classification accuracy, and designed a cascaded model in a single CNN by adding extra mid output layers, which reduces the average inference time significantly.

As the demand for handwriting recognition changes, the improvement and optimization of single-character recognition accuracy and speed can no longer content the actual application needs. They are facing many new challenges, such as the error-accumulation of caused by error-segmentation, indivisible characters and text lines, and complex context information and so on. Therefore, we need to further explore new methods based on the existing single-character recognition to meet actual needs of offline handwritten text recognition.

3.2. Word/line Recognition Methods

The recognition of offline handwritten document has been commonly approached by the use of sequential pattern recognition techniques. Text lines are processed along a temporal sequence by learning models that leverage their sequence of internal states as memory cells, in order to be able to tackle variable length input signals. The first handwritten word/line recognition method was based on HMM [12], a novel handwritten character sequence recognition method based on the windowed Bernoulli mixture HMMs. Bianne et al. [22] constructed a handwritten text line recognizer using HMM, decision tree and a set of expert knowledges. Bluche et al. [23] introduced a handwritten word recognition method based on HMM and CNN.

With the rise of neural networks learning, RNNs [24] used to deal with character sequence data has started to become more popular. In these approaches, BLSTM [25] or MDLSTM [26] were the two most commonly used models. Through the comparison of two methods based on 1D-LSTM and 2D-LSTM layers, Puigcerver [27] proved that multi-dimensional recurrent layers may not be necessary to acquire good classification accuracy for offline handwritten character sequence recognition. Hassan and Abdelkarim [28] build a text recognition architecture combine of Deep CNNs and RNNs. Chowdhury et al. [29] introduced a novel approach that combines a CNNs with a recurrent Encoder-Decoder network to map an image to a sequence of characters corresponding to the text-line image. Kundu et al. [30] used improved RNNs architecture for the Generator and patch GAN architecture for the discriminator with different combinations of loss functions. Although all those approaches use recurrent architectures to properly conceal and learn serial information, they may suffer of the lack of parallelization during the training stage and require a huge amount of labelled training data.

Currently, most of the offline handwritten word/line recognition methods are combining the recurrent neural networks and connectionist temporal classification layers [31]. Carbonell et al. [7] introduced an end-to-end OHTR model that integrates a one-stage target detection neural network and branches for recognizing document data and named entities, so it can obtained the training errors of each task while learning shared features. Puigcerver [27] presented a composite architecture that
combines a CNNs with a deep one-dimensional RNN-CTC model. Liu et al.[32] propose a novel efficiency and effective OHTR algorithm with CNNs-only network for the challenging offline handwritten Chinese text recognition task. Like other approaches, the CTC-based methods must to face up its limitations, such as the high complexity and the slowness training speed, for gaining a reasonable performance in offline handwritten word/line recognition processes[3].

3.3. Multi-lines Recognition Methods

For OHTR problems, the coupled nature of segmentation and recognition is one of the most important challenges. Text line/word segmentation is still an error-prone process, which will greatly affect the performance of text recognition systems[5, 33, 34]. Fortunately, this problem has been progressively solved by the segmentation-free OHTR module, which also proved much cheaper to labelling data for and more classification accuracy.

For multi-lines recognition of OHTR, many of the approaches train individual text paragraph detection, text line segmentation, text line recognition, and then combine the above steps into a text recognition system. Moysset et al.[35] introduced a multi-lines recognition method which integrates fully convolutional neural network based text line localization network and MDLSTM based text recognition. Bluche et al.[36, 37] proposed a modification of the MDLSTM recognizing full paragraphs or multi-line of document without an explicit segmentation. The method converting the two-dimensional multi-lines representation into a single sequence of predictions by the collapse convolutional neural network layers, which can recognition one text line at a time. Wiginton et al.[6] introduced a multi-lines handwritten text recognition method combining region proposal network to gain the start position of each text line, normalize text lines by line follower network, and fulfil text recognition using CNN-LSTM network. Obviously, the method is through pertained individual networks separately, and then jointly trained them together to achieve the multi-line text recognition. Although the above methods achieve multi-lines recognition to a certain extent, they still face to some problems. They require to pre-train their encoder subnetwork on single-line document before training on multi-line versions, and very slow compared with most current text recognition methods based on text line segmentation.

In response to the previous problem, Tensmeyer and Wiginton[38] proposed an adapting text recognition method in a weakly supervised manner without requiring text line segmentation, it solving the text alignment between the predicted text line transcriptions and its ground truth. Peng et al.[2] presented an end-to-end offline handwritten Chinese text recognition using fully convolutional network, which can acquire text line segmentation and text recognition results at the same time. Yousef et al.[8, 39] proposed a multi-line recognition method based on existing neural network recognition model, it provided enough spatial capacity to collapse a 2D extraction feature into 1D without losing essential feature information, and trained using exactly their original procedure.

4. Offline Handwritten Text Datasets

This Section focuses on six common OHTR datasets, they summary information shown in Table 1, and the details are as follows:

(1) IAM dataset[40], English documents extracted from the LOB corpus, copied by different authors. The database includes 1,199 document images, in which 459,308 characters distributed in 10,373 text lines, and 79 different symbols.

(2) Bentham dataset[41], as annotating handwritten historical documents, constructed during the transScriptorium project. The database includes 433 historical document images, in which 524,065 characters distributed in 11,473 text lines, and 93 different symbols.

(3) BH2M dataset[42], as handwritten marriages documents during Barcelona historical, which written by the one author. The dataset contains 174 handwritten documents. The database consists of annotated images in an XML hierarchical structure.

(4) HIT-MW dataset[43], as the first collection of Chinese offline handwritten documents, collected by Artificial Intelligence Lab, Harbin Institute of Technology. The dataset includes 853
handwritten documents written by more than 780 authors, in which 8,664 text lines, 186,444 characters.

(5) HWDB1.0-1.1, as part of CASIA-HWDB dataset[44], constructed by the Institute of Automation of Chinese Academy of Sciences. The dataset written by 720 authors, contains 2,853,165 characters, 4,052 categories, in which, 3,881 Chinese characters and 171 alphanumeric and symbols.

(6) HWDB 2.0-2.2, as part of CASIA-HWDB dataset[44], constructed by the Institute of Automation of Chinese Academy of Sciences. The dataset contains 5,091 pages, in which 1,349,414 characters distributed in 52,230 text lines, and 2,703 different symbols.

Table 1. Information for the OHTR common datasets

| Database       | Pages | Lines | Words | Characters | Class | Language |
|----------------|-------|-------|-------|------------|-------|----------|
| IAM            | 1199  | 10373 | 89896 | 459308     | 79    | English  |
| Bentham        | 433   | 11473 | 96155 | 524065     | 93    | English  |
| BH2M           | 174   | 5498  | 56645 | —          | 3360  | Barcelona|
| HIT-MW         | 853   | 8664  | —     | 186444     | 3041  | Chinese  |
| HWDB1.0-1.1    | —     | —     | —     | 2853165    | 4052  | Chinese  |
| HWDB2.0-2.2    | 5091  | 52230 | —     | 1349414    | 2703  | Chinese  |

5. Conclusion
Over the last decade, many researchers have presented a large variety of methods for Offline handwritten text recognition using deep learning. While unstrained handwriting text recognition remains a challenging recognition problem, classification accuracy has jumped significantly in the last decade, mainly due to advancements in deep learning technology including new ideas, algorithms and model structure. Analyzing the recent contributions to the offline handwritten text recognition domain, we can generally summarize their research focus into the following two aspects: (1) Weakly supervised handwritten text recognition. The fundamental purpose of deep learning is to obtain equivalent or better learning effects based on less domain knowledge or expert experience. This is the goal pursued by most researchers engaged in OHTR. (2) Faster and compact module for handwritten text recognition. Deep learning networks intuitively appear to require the storage of a large number of parameters, need longer training time, and then incur high computational cost. From the authors' point of view, this trend will continue for future work, better handwritten text feature extract and recognition will explore by researchers, and improve classification with limited marking information of offline handwritten text.

Acknowledgments
This work was financially supported by the Natural Science Foundation of Jiangsu Province (BK20180142), the and Jiangsu Government Scholarship for Overseas Studies (JS-2019-104).

References
[1] Kimura, F., Takashina, K., Tsuruoka, S., Miyake, Y. (1987) Modified quadratic discriminant functions and the application to Chinese character recognition. IEEE transactions on pattern analysis and machine intelligence, 9(1): 149-153.
[2] Peng, D., Jin, L., Wu, Y., Wang, Z., Cai, M. (2019) A Fast and Accurate Fully Convolutional Network for End-to-End Handwritten Chinese Text Segmentation and Recognition. International Conference on Document Analysis and Recognition, pp. 25-30.
[3] Kang, L., Riba, P., Villegas, M., Fornés, A., Rusiñol, M. (2020) Candidate fusion: Integrating language modelling into a sequence-to-sequence handwritten word recognition architecture. Pattern Recognition, 112: 107790-1-12.
[4] Zhang, X., Bengio, Y., Liu, C. (2017) Online and offline handwritten Chinese character recognition: A comprehensive study and new benchmark. Pattern Recognition, 61: 348-360.
[5] Moysset, B., Kermorvant, C., Wolf, C. (2018) Learning to detect, localize and recognize many text objects in document images from few examples. International Journal on Document Analysis and Recognition, 21(3): 161-175.

[6] Wigington, C., Tensmeyer, C., Davis, B., Barrett, W., Price, B., Cohen, S. (2017) Start, Follow, Read: End-to-End Full-Page Handwriting Recognition. Proceedings of the European Conference on Computer Vision, pp. 367-383.

[7] Carbonell, M., Fornés, A., Villegas, M., Lladós, J. (2020) A Neural Model for Text Localization, Transcription and Named Entity Recognition in Full Pages. Pattern Recognition Letters, 136: 219-227.

[8] Yousef M., Bishop, T. E. (2020) OrigamiNet: Weakly-Supervised, Segmentation-Free, One-Step, Full Page Text Recognition by learning to unfold. Conference on Computer Vision and Pattern Recognition, pp. 14710-14719.

[9] Zhong, Z., Jin, L., Xie, Z. (2015) High performance offline handwritten chinese character recognition using googlenet and directional feature maps. International Conference on Document Analysis and Recognition, pp. 846-850.

[10] Wang Y., Xiao, W. (2019) Handwritten Text Line Segmentation Method by Writing Pheromone Diffusion and Convergence. Cognitive Cities Conference, pp. 105-113: Springer.

[11] Min, F., Zhu, S., Wang, Y. (2020) Offline Handwritten Chinese Character Recognition Based on Improved Googlenet. International Conference on Artificial Intelligence and Pattern Recognition, pp. 42-46.

[12] Giménez, A., Khoury, I., Andrés-Ferrer, J., Juan, A. (2014) Handwriting word recognition using windowed Bernoulli HMMs. Pattern Recognition Letters, 35: 149-156.

[13] Li, Z., Teng, N., Jin, M., Lu, H. (2018) Building efficient CNN architecture for offline handwritten Chinese character recognition. International Journal on Document Analysis and Recognition, 21(4):233-240.

[14] Cireşan, D., Meier, U., Schmidhuber, J. (2012) Multi-column deep neural networks for image classification. Conference on computer vision and pattern recognition, pp. 3642-3649.

[15] Cireşan, D., Meier, U. (2015) Multi-Column Deep Neural Networks for Offline Handwritten Chinese Character Classification. Joint Conference on Neural Networks, pp. 1-6.

[16] Wu, C., Fan, W., He, Y., Sun, J., Naoi, S. (2014) Handwritten character recognition by alternately trained relaxation convolutional neural network. International Conference on Frontiers in Handwriting Recognition, pp. 291-296.

[17] Li, J., Song, G., Zhang, M. (2020) Occluded offline handwritten Chinese character recognition using deep convolutional generative adversarial network and improved GoogLeNet. Neural Computing and Applications, 32(9): 4805-4819.

[18] Wang, W., Zhang, J., Du, J., Wang, Z., Zhu, Y. (2018) DenseRAN for Offline Handwritten Chinese Character Recognition. International Conference on Frontiers in Handwriting Recognition, pp. 104-109.

[19] Wang, T., Xie, Z., Li, Z., Jin, L., Chen, X. (2019) Radical aggregation network for few-shot offline handwritten character recognition. Pattern Recognition Letters, 125: 821-827.

[20] Zhou, S., Wu, J., Wu, Y., Zhou, X. (2015) Exploiting local structures with the kronecker layer in convolutional networks. arXiv preprint arXiv:1512.09194, pp.1-17.

[21] Xiao, X., Jin, L., Yang, Y., Yang, W., Sun, J., Chang, T. (2017) Building fast and compact convolutional neural networks for offline handwritten Chinese character recognition. Pattern Recognition, 72: 72-81.

[22] Bianne, A., Menasri, F., Mohamad, R., Mokbel, C., Kermorvant, C., Likforman, L. (2011) Dynamic and contextual information in HMM modeling for handwritten word recognition. IEEE transactions on pattern analysis and machine intelligence, 33(10): 2066-2080.

[23] Bluche, T., Ney, H., Kermorvant, C. (2013) Tandem HMM with convolutional neural network for handwritten word recognition. IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 2390-2394.
[24] Lipton, Z., Berkowitz, J., Elkan, C. (2015) A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019, pp. 1-38.
[25] Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., Schmidhuber, J. (2008) A novel connectionist system for unconstrained handwriting recognition. IEEE transactions on pattern analysis and machine intelligence, 31(5): 855-868.
[26] Graves A., Schmidhuber, J. (2009) Offline handwriting recognition with multidimensional recurrent neural networks. Advances in neural information processing systems, pp. 545-552.
[27] Puigcerver, J. (2017) Are multidimensional recurrent layers really necessary for handwritten text recognition?. Conference on Document Analysis and Recognition, pp. 67-72.
[28] Bahi, H., Zatni, A. (2019) Text recognition in document images obtained by a smartphone based on deep convolutional and recurrent neural network. Multimedia tools and applications, 78(18): 26453-26481.
[29] Chowdhury, A., Vig, L. (2018) An efficient end-to-end neural model for handwritten text recognition. arXiv preprint arXiv:1807.07965, pp. 1-11.
[30] Kundu, S., Paul, S., Bera, S., Abraham, A., Sarkar, R. (2020) Text-line extraction from handwritten document images using GAN. Expert Systems with Applications, 140: 112916.
[31] Graves, A., Fernández, S., Gomez, F., Schmidhuber, J. (2006) Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. Proceedings of International Conference on Machine Learning, pp. 369-376.
[32] Liu, B., Xu, X., Zhang, Y. (2020) Offline Handwritten Chinese Text Recognition with Convolutional Neural Networks. arXiv preprint arXiv:15619, pp. 1-6.
[33] Carbonell, M., Mas, J., M., Villegas, Fornés, A., Lladós, J. (2019) End-to-end handwritten text detection and transcription in full pages. International Conference on Document Analysis and Recognition Workshops, 5: 29-34.
[34] Chung, J., Delteil, T. (2019) A Computationally Efficient Pipeline Approach to Full Page Offline Handwritten Text Recognition. International Conference on Document Analysis and Recognition Workshops, pp. 35-40.
[35] Moysset, B., Kermorvant, C., Wolf C. (2017) Full-page text recognition: Learning where to start and when to stop. Conference on Document Analysis and Recognition, pp. 871-876.
[36] Bluche, T., Louradour, J., Messina, R. (2017) Scan, attend and read: End-to-end handwritten paragraph recognition with md LSTM attention. International Conference on Document Analysis and Recognition, pp. 1050-1055.
[37] Bluche, T. (2019) Joint line segmentation and transcription for end-to-end handwritten paragraph recognition. Advances in Neural Information Processing Systems, pp. 838-846.
[38] Tensmeyer, C., Wigington, C. (2019) Training Full-Page Handwritten Text Recognition Models without Annotated Line Breaks. International Conference on Document Analysis and Recognition, pp. 1-8.
[39] Yousef, M., Hussain, K., Mohammed, U. (2020) Accurate, data-efficient, unconstrained text recognition with convolutional neural networks. Pattern Recognition, 108: 107482.
[40] Marti, U., Bunke, H. (2002) The IAM-database: an English sentence database for offline handwriting recognition. Journal on Document Analysis and Recognition, 5(1): 39-46.
[41] Sánchez, J. (2013) TransScriptorium: a european project on handwritten text recognition. ACM symposium on Document engineering, pp. 227-228.
[42] Fernández, D., Almazán, J., Cirera, N., Fornés, A., Lladós, J. (2014) Bh2m: The barcelona historical, handwritten marriages database. Conference on Pattern Recognition, pp. 256-261.
[43] Su, T., Zhang, T., Guan, D. (2007) Corpus-based HIT-MW database for offline recognition of general-purpose Chinese handwritten text. International Journal of Document Analysis and Recognition, 10(27): 0037-1-6.
[44] Liu, C., Yin, F., Wang, D., Wang, Q. (2011) CASIA online and offline Chinese handwriting databases. International Conference on Document Analysis and Recognition, pp. 37-41.