

A Survey on Chinese OCR Techniques based on Deep Learning

Xiang Li, *Yeonjoon Lee

Major in Bio Artificial Intelligence, Department of Computer Science and Engineering,
Hanyang University, *Department of Computer Science and Engineering, Hanyang University
lx913247225@hanyang.ac.kr, yeonjoonlee@hanyang.ac.kr

딥러닝 기반 중문 OCR 기술 동향 연구

이향, * 이연준

한양대학교 컴퓨터공학과 바이오인공지능융합전공,
* 한양대학교 컴퓨터공학과

Abstract

With the rapid development of Deep Learning, OCR techniques based on Deep Learning are also prospering. The OCR techniques can effectively detect text in natural images, it is a process of scanning manuscripts or books, and analyzing scanned images to obtain information about characters and pages. Therefore, OCR technology is a relatively convenient and simple character input method. In this paper, we offer an overview of the most recent OCR methods based on Deep Learning.

I Introduction

Optical Character Recognition (OCR) is the process of recognizing various forms of characters from handwriting to print in an image and converting them into character codes that can be entered by the machine. There are two phases in the OCR process: text detection and text recognition. In the first stage, only characters in the text area will be located. and then they are going to be extracted for text detection. A typical OCR system is divided into two main branches by process: detection then recognition and end-to-end text spotting. The OCR was originally developed to recognize English, but there are very big differences between Chinese and English characters, recognition of English character requires only 26 letters whereas each Chinese characters, Hanzi, is made up of different graphemes and is different from each other. So the OCR system on English letters cannot be directly applied to Chinese Hanzi. In the past, there were few OCR techniques for Chinese Hanzi, but many of them have emerged in recent years. In this paper, we will offer an overview of these OCR techniques.

II Method

In this section, we will briefly explain some DB-based methods used for OCR towards Chinese characters.

II.1 Traditional OCR Methods

In text detection and recognition, the most classical algorithm is CRNN (Convolutional Recurrent Neural Network) proposed by Shi et al.^[1] in 2017, which combines CNN and RNN and is based on end-to-end image sequence recognition, instead of cutting a single text first. The whole CRNN structure consists of three parts, from bottom to top: in the CNN Layer features are extracted from input images; RNN Layer, using BLSTM (Bidirectional Long Short-Term Memory) neural network to predict the feature sequence, learn each feature vector in the sequence, and output the predicted label; CTC Layer, using CTC (Connectionist Temporal Classification) loss function, converts a series of labels obtained from the CNN layer into the final label sequence.

Xie et al.^[2] used CRNN based on Pytorch deep learning framework to study Chinese and open source. This method combines CTPN^[3] (Connectionist Text Proposal Network) to realize the continuous process of text detection and text recognition. Chinese scene text detection has obvious boundary and sequence characters. Xie et al. trained with 3.6 million Chinese Text data sets and were able to finetune with 97.7% accuracy. Their study introduced deep learning framework and inspired subsequent research towards OCR of Chinese Hanzi.

II.2 PP series OCR Methods

PaddlePaddle is a deep learning framework developed by Baidu Inc., China. Based on PaddlePaddle framework, scientists have developed the PP series OCR techniques in recent years, which are remarkable creation in this field. Du et al.^[4]'s research focuses on DB^[5] (Differentiable Binarization) in text detection and CRNN in text recognition, proposing PP-OCR, and use a correction module that can effectively detect cursive, rotated, curved, clutter and multilingual. Based on PP-OCR, Du et al.^[6] optimized the model and proposed PP-OCRv2: utilizing CML (Collaborative Mutual Learning) Knowledge Distillation and CopyPaste Data Augmentation in text detection; Enhanced CTC loss functions, lightweight backbone network and U-CML (Unified-Collaborative Mutual Learning) in text recognition. Li et al.^[7] continued to attach improvement strategies to this foundation and proposed PP-OCRv3: there are 9 improvements, 3 for text detection and others for text recognition, Experiments on real data demonstrate that the accuracy of PP-OCRv3 is higher than that of PP-OCRv2 at the same inference cost. The advantage of PP-OCR series is that technology is well integrated. Figure.1 shows the comparison of PP-OCR in 3 versions with the same Chinese scene text picture.

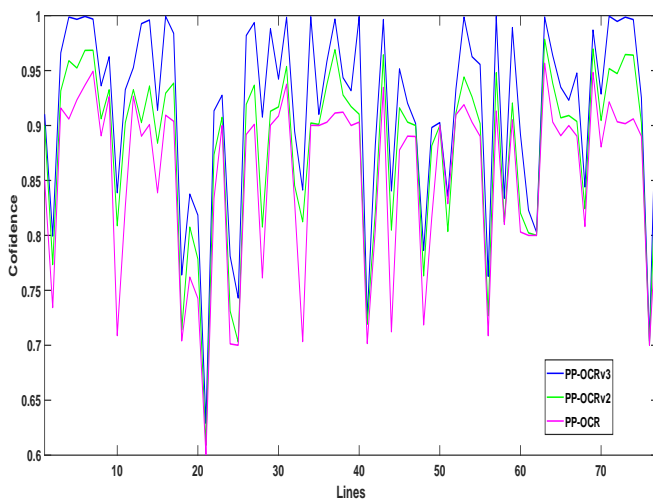


Figure 1: **Confidence between PP-OCR series**

II.3 MMOCR Method

Kuang et al.^[8] proposed MMOCR, although its end-to-end recognition isn't done very well, it offer a comprehensive toolbox for text detection, recognition and understanding. MMOCR has implemented 14 state-of-the-art algorithms, offered a wide range of trained models. MMOCR implements not only the classic CRNN algorithm, but also the algorithm based on Transformer (encoder-decoder or attention).

III Conclusion

The paper briefly introduced recent research on Chinese OCR. Different techniques are applied in many scenarios, the improvement strategies are also being optimized. With the marketization and standardization of this field, it will make tremendous progress.

ACKNOWLEDGMENT

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2022-00155885, Artificial Intelligence Convergence Innovation Human Resources Development (Hanyang University ERICA)) and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2020-0-01343, Artificial Intelligence Convergence Research Center(Hanyang University ERICA)).

References

- [1] Shi, Baoguang, Xiang Bai, and Cong Yao. "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition." *IEEE transactions on pattern analysis and machine intelligence* 39.11 (2016): 2298-2304.
- [2] Xie Jinheng. CRNN_Chinese_Characters_Rec documentation. <https://github.com/Sierkinhane>, 2020, 11
- [3] Tian, Zhi, et al. "Detecting text in natural image with connectionist text proposal network." *European conference on computer vision*. Springer, Cham, 2016.
- [4] Du, Yuning, et al. "Pp-ocr: A practical ultra lightweight ocr system." *arXiv preprint arXiv:2009.09941* (2020).
- [5] Liao, Minghui, et al. "Real-time scene text detection with differentiable binarization." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 34. No. 07. 2020.
- [6] Du, Yuning, et al. "PP-OCRv2: bag of tricks for ultra lightweight OCR system." *arXiv preprint arXiv:2109.03144* (2021).
- [7] Li, Chenxia, et al. "PP-OCRv3: More Attempts for the Improvement of Ultra Lightweight OCR System." *arXiv preprint arXiv:2206.03001* (2022).
- [8] Kuang, Zhanghui, et al. "MMOCR: a comprehensive toolbox for text detection, recognition and understanding." *Proceedings of the 29th ACM International Conference on Multimedia*. 2021.