A Survey of Optical Character Recognition Technologies And Its Advancement

Yiru Xiong

***Abstract*—Optical Character Recognition (OCR) is the process to convert different types of images of text into machine readable, editable and searchable text formats. As one of the first and most mature research topics in computer vision field, OCR technologies have been rapidly evolved due to their wide applications in various business domains and the overall advancement of artificial intelligence. Researchers in computer vision, natural language processing (NLP), document understanding, and various artificial intelligence fields have contributed to the advancement of OCR from different perspectives. The evolution of OCR in the past decade was driven significantly by breakthroughs in deep learning. Besides the evident improvement in its accuracy, the state-of-the-art OCR technology has also demonstrated advancement in terms of its adaptability, accessibility, and robustness. This survey paper focuses on reviewing some of the major contributions to the evolution of OCR technologies over the past decade. The goal is to provide insights for future researchers on existing technologies and potential technology gaps.

***Keywords*—Optical Character Recognition, Scene Text Recognition, OCR Survey, Document Image Analysis, Feature Extraction, Image Classification, Postprocessing, Computer Vision, Object Detection, Transformer, Encode-Decoder, Mask, Pretrain models, Transfer Learning, Postprocessing**

I. INTRODUCTION

In the artificial intelligence world, computer vision serves as the “eye” of most AI-powered applications. As one of the first and most mature topics in computer vision, OCR technologies have been rapidly evolved due to their wide applications in various business domains and the advancement of deep learning. OCR technologies can be applied to handwritten documents, scanned PDFs, webpage screenshots, images captured by cameras, and many more sources where text information needs to be extracted from. Nowadays, OCR applications and services can be accessed from many resources including cloud-based ones and can be further improved with the integration of machine translation and NLP for direct multilingual support. Together with these mentioned technologies in AI, OCR technologies transform the entire information extraction process from images, videos, and numerous digital sources in this digital age. How does OCR technology evolve over the past decade? What are some popular and latest OCR technologies, and how does the advancement in deep learning expediate the performance of OCR services and applications? These topics will be elaborated and discussed in greater details in the following chapters.

II. METHFOLOGIES

As the first step, an exploratory study was carried out to understand how OCR technology has evolved over the past decade. The goal is to discover changes in the major architectures and adopted models, and how these updates contribute to the improvement of OCR performance. Through this exploration phases, major contributors among relevant researcher papers were selected and reviewed further. A preliminary literature review was conducted first and a small subset of papers were then selected for a more comprehensive systematic review and comparison.

For this study, more than eight research papers dated as early as 2013 were selected for preliminary review, and among them, 5 research papers were reviewed in details and included in this survey paper.

The selected papers include a survey paper which serves as a good summary of general OCR technologies before 2017, and one focus on applying deep learning technologies to OCR for handwritten texts. The other two represent the most cutting-edge technologies and the latest trends in the OCR field – transformer-based architectures. This survey will include a brief summary of each mentioned paper, followed by an analysis of their major contributions, performance review, and potential insights for future studies.

III. LITERATURE REVIEW

A. A Survey on Optical Character Recognition System [1]

This survey paper serves as a roadmap of existing research done in the filed of OCR before the paper was published in 2017. It reviewed papers in five major phases in the process of OCR, namely preprocessing, character segmentation, feature extraction, character classification, and the post processing phase.

- Image acquisition: this process involves acquiring images, and approaches include digitization, binarization, and compression. More approaches have been explained in this paper [5].
- Pre-processing: this phase include noise removal, skew removal, thinning, and morphological operations.
- Segmentation: this phase includes implicit and explicit segmentation and separate image into constituent characters
- Feature Extraction: various features are extracted from image including geometrical feature and statistical features
- Classification: this phase is where the magic could play a role in. By utilizing Bayesian, Nearest Neighborhood, and Neural Network models, image can be categorized into specific classes based on selected features
- Post-processing: in this phase the OCR results are further improved through contextual and dictionary-based methods

The major contribution of this paper is to separate OCR into these five major tasks, and summarized relevant studies on each phase into one readable paper. The paper outlined potential subtasks future researchers can improve on to contribute the improvement of OCR from different perspectives.

B. Optical Character Recognition Using Deep Learning Techniques for Printed and Handwritten Documents [2]

This research paper pioneers in the sense of applying deep learning methods into OCR on handwritten texts specifically. It has always been a challenge for OCR applications to achieve a close to perfect accuracy on images tasks contain handwritten texts. Researchers introduce a new architecture named CL-9, which was built upon a Convolutional Recurrent Neural Network (CRNN) model with 7 CNN layers and 2 LSTM layers. This proposed model was tested on handwritten text images and achieved accuracy an 94.79% accuracy for printed text source, and 75.2% for handwritten discrete source.

Through this study, researchers demonstrate the power of leveraging deep learning in OCR tasks, which set up a solid bedrock for further deep learning-based OCR research. As a drawback of this paper, the accuracy on the handwritten cursive texts is only 65.7%, which is far from a satisfaction level, especially in practical usages. This study did not go further by fine tuning their proposed model or suggesting potential improvement methods for future studies.

C. MaskOCR: Text Recognition with Masked Encoder-Decoder Pretraining.[3]

This paper introduces a novel approach to unify both vision and language pre-training in the classification phase of the framework, which was normally done separately in previous approaches. As text images contain both visual and linguistic information, by focusing on one of the two may lead to the information loss and scarification in classification performance. Masked image modeling approach, however, enable encoder and decoder to learn the data modalities of both vision and language. Researchers tested this proposed method on both Chinese and English text images, and concluded with consistent result that this approach can improve the performance on benchmark datasets.

Encoder-decoder transformer is used for text recognition
A large set of unlabeled real text images are added to pretrain the feature encoder for a better visual representation learning
Directly pre-train the sequence decoder and transform text data into synthesized text images to unify the visual and linguistic representations

The model was trained on both English and Chinese datasets, where two languages are very different in their visual and linguistic presentations

This paper demonstrates the performance of OCR can be improved through adding extra steps in the pre-training phases. The conclusion is backed with strong experiments and results. It is one of the state-of-the-art approaches adopted in the current OCR technologies.

D. Exploring Machine Learning Solutions in the Context of OCR Post-Processing of Invoices [4]

As discussed in earlier section, the final result of the OCR does not solely depend on the computer vision related tasks, appropriate post-processing upon output context from OCR engine is very crucial to boost the accuracy of the extracted information further and tailor the output to specific business requirements. Unlike other research papers focus on the core of OCR system, this paper proposes effective post-processing methods on OCR outputs utilizing machine learning approaches. It is an inspiring piece for future studies on the post-processing phase of OCR at the intersection field of computer vision and natural language processing.

In this study, researchers conducted an experiment with scanned images of 70 real invoices from companies and aimed to discover common errors in their outputs from Azure’s Form Recognizer. The purpose is to explore the possibility of leveraging machine learning based classification method as an error detection, which could be a very useful first step for further correction and processing to enhance the overall accuracy rate.

A Bidirectional Encoder Representations from Transformers (BERT) model was adopted and fined-tuned for the invoice classification. According to the result, the two most common OCR errors in extracting information from invoices are presence of extra words and absence of words. After fine tuning, the BERT model classifier is able to achieve an accuracy close to 83.2% in error detections. It is promising that this can be leveraged as a bedrock for a robust error corrector of OCR errors.

Having an error detector/corrector can definitely make further post-processing more efficient, but there are more untouched fields and perspectives in the post-processing phase that this paper has not brought into discussion. This can be a potential direction for future research.

E. TR-OCR: Transformer-based optical character recognition with pre-trained models [5]

Instead of creating from classic CNN models, this paper introduces a Transformer-based OCR architecture build upon pre-trained model. This proposed end-to-end architecture leverages the pre-trained image transformer and text transformer models. Simple but effective, it is able to outperform most existing models on the printed, handwritten and scene text recognition tasks. As this study is still an ongoing one, researchers provide open-source code in GitHub for further implementations and improvements.

IV. FUTURE WORK

A few observations can be generated through reviewing these research papers in OCR topics:

- To further improve OCR performance, computer vision is not the only niche to be focused on. As the overall OCR tasks involve pre-processing, post-processing, and other necessary steps in different phases, interdisciplinary approaches can be leveraged to enhance the accuracy or efficiency of the overall text information extraction.
- The state-of-the-art machine learning techniques to be utilized in OCR are transformer based. Further studies can be focused more on utilizing pre-trained transformer-based models and incorporating attention into the architecture.
- Machine learning techniques and natural language processing methods can be leveraged in different phases of OCR, such as the post-processing phase, to further enhance the effectiveness, accuracy. and the overall performance of the OCR system.

V. REFERENCES

[1] Islam, Noman, et al. “A Survey on Optical Character Recognition System.” ArXiv:1710.05703 [Cs], 3 Oct. 2017, arxiv.org/abs/1710.05703.

[2] Bagwe, Sanika and Shah, Vruddhi and Chauhan, Jugal and Harniya, Purvi and Tiwari, Amanshu and Gupta, Vartika and Raikar, Durva and Gada, Vrushabh and Bheda, Urvi and Mehta, Vishant and Warang, Mahesh and Mehendale, Ninad, Optical Character Recognition Using Deep Learning Techniques for Printed and Handwritten Documents (July 31, 2020). Available at SSRN: https://ssrn.com/abstract=3664620 or http:// dx.doi.org/10.2139/ssrn.3664620

[3] Lyu, Pengyuan, et al. “MaskOCR: Text Recognition with Masked Encoder-Decoder Pretraining.” ArXiv.org, 9 Oct. 2023, arxiv.org/abs/2206.00311. Accessed 6 Dec. 2023.

[4] Dwyer, J., & Bertse, S. (2022). Exploring Machine Learning Solutions in the Context of OCR Post-Processing of Invoices (Dissertation, KTH Royal Institute of Technology). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-321737

[5] Li, Minghao, et al. “TROCR: Transformer-based optical character recognition with pre-trained models.” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 11, 2023, pp. 13094–13102, https://doi.org/10.1609/aaai.v37i11.26538.

[6] Lund, W.B., Kennard, D.J., & Ringger, E.K. (2013). Combining Multiple Thresholding Binarization Values to Improve OCR Output presented in Document Recognition and Retrieval XX Conference 2013, California, USA, 2013. USA: SPIE