УДК 004.932.75'1

DEEP LEARNING FOR HANDWRITING TEXT RECOGNITION: EXISTING APPROACHES AND CHALLENGES

Mohammed.S (ITMO university)
Supervisor –Associate Professor Ph. D, Teslya N.N
(SPC RAS)

Annotation (Handwritten Text Recognition HTR is the process of extracting handwritten text from an image and converting it into a digital form for machine operation. The existing approaches to solve the HTR are based on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which utilize the Connectionist Temporal Classification (CTC) objective function, and approaches based on attention models (Seq2Seq). In this article, we provide extensive comparison of the current Deep Learning approaches for the task of HTR. Also, we outline the current problems that limits the effectiveness of these approaches.)

Introduction. HTR has been of interest to the Pattern Recognition community for many years. Converting images from handwritten text to machine readable format has a large number of uses such as making HTR an open research problem that is still challenging. With the advent of neural networks and Deep Learning architectures, HTR, like many other applications, has dramatically improved performance and many surveys on the current approaches have been published. Although these surveys usually cover the basics, they are general and not focused on the HTR architectures. In addition, new research approaches have emerged. Therefore, it will be realistic and necessary to investigate the current state-of-the-art HTR approaches.

Main part. We focus on conducting a comprehensive and extensive comparison between the current Deep learning approaches for the task of HTR, for such comparison, the IAM dataset and the most usual evaluation metrics for HTR systems were adopted: Character Error Rate (CER).
By the result of comparison, we could figure out some main points:
1-A model with millions of trainable parameters makes it a challenge to be implemented in many real-world applications. Meanwhile, a model with fewer parameters (in thousands), exchange high performance for simplicity of the model.
2-Most of CTC_based approaches have difficulties in remembering long terms due to vanishing/ exploding gradient descent.
3- CTC only allows monotonic alignments, which may be a valid assumption for word-level or line-level HTR tasks, but it lacks the possibility for further research on paragraph or even more complex article styles. Also, it is challenging for detectors to separate words for scripts that do not separate words by spaces such as Chinese, Japanese, and Korean and they are more likely to miss punctuation and diacritic marks.
4- Considering issues related to long images. There are at least two new aspects that need to be considered, which are efficiency and performance: long images affect the efficiency of models with the Self-Attention encoder due to quadratic scaling with image length. This problem can be solved for CTC models without performance loss by chunking the images. Training on images of fixed maximum width affects the performance on longer images of models that make use of the Transformer decoder. This issue can be solved by resizing the images to the train width.

5- Although encoder/decoder method achieved great success in the field of HTR, there is still a lot of room to improve with pre-trained CV and NLP models.

6-TrOCR is the first end-to-end Transformer-based model with pre-trained CV and NLP models with the less CER 2.89%. It doesn’t use CNN as backbone, doesn’t rely on any pre/post-processing

steps and can be extended to multilingual model with minimum efforts. However, the total number of parameters of the base model 334 million, comparing to 558 million parameters for the large model

Conclusion. Although, the researches could enhance the accuracy rate and control the time complex in the field of HTR, two important factors that might slow down the widespread adoption of these techniques in the area are the need for large training data and subsequently advanced hardware and more complex architectures to deal with such high amounts of data and achieve the good improvements. The main contribution of this article is providing a critical state-of-art review of the recent proposed approaches in the HTR field. Also, we identified and discussed the main challenges that still exist.

Mohammed.S (author)	Signature
Teslya N.N. (supervisor)	Signature