Handwritten Recognition using Transformer for Medical Prescription Application

Zobeir Raisi[a], Mohamed A. Naiel[a], Steven Wardell[b], Paul Fieguth[a] and John Zelek[a]

[a] Vision and Image Processing Lab, Department of Systems Design Engineering, University of Waterloo, ON, N2L 3G1 Canada

{zraisi, mohamed.naiel, pfieguth, jzelek}@uwaterloo.ca, (519) 888-4567

[b] ATS Automation Tooling Systems Inc., Cambridge, ON, N3H 4R7, Canada

swardell@atsautomation.com, (519) 653-6500

Abstract

Introduction
Optical Character Recognition (OCR) is a solved problem for controlled text typed recognition. Handwriting is still used in the medical field for prescriptions or in the lab for identifying contents in sample containers. The digitization of handwritten records, whether on paper or on containers is a difficult problem that current OCR technology cannot solve. Unlike optical-character recognition (OCR), Handwritten Text Recognition (HTR) involves more challenging scenarios such as variant writing styles and complex data collection process. Recently introduced AI based deep-learning architectures appear to provide superior performance to HTR when compared to that of the classical machine-learning-based methods. The current state-of-the-art HTR methods [10, 1, 12, 11, 8] are mostly inherited from Recurrent Neural Networks (RNNs) architectures [3]. However, these methods [1, 12, 11, 8] require features in one-dimensional (1D) representation and generally fail to recognize characters in 2D images due to losing the spatial information. Transformers [13] have recently showed superior performance for the problem of text understanding and we exploit that architecture for solving HTR.

Methodology
We use a Transformer-based architecture [13] for handwritten text recognition since it better preserves the spatial information within 2D images than the prior approaches. More specifically, Transformers are deep-learning-based methods that scan through each character of a written word by using a self-attention module and update it by aggregating information from the whole word. We propose some modifications to the original Transformer architecture to embed spatial information of 2D image features generated by Convolutional Neural Networks (CNN). Therefore, the proposed model is capable to tackle the various handwriting styles. Due to the lack of training data, we first train the proposed model on the synthetic text recognition datasets in [4, 2] and then fine-tune it on 79 classes of the IAM handwritten recognition training dataset [9] to generate our final model, where the 79 classes include alphanumerical and special characters.

Experimental Results
To the best of our knowledge, there is no publicly available dataset for medical prescription handwritten recognition; Hence, for evaluation, we use the images of the IAM HTR testing dataset [9] that is close to our primary objective of medical prescription HTR. The IAM HTR dataset also includes several challenges such as different writers, special characters and low-resolution images. Experimental results on IAM HTR testing dataset show that the proposed scheme achieves a word recognition accuracy of 91.85% and it outperforms by a large margin of 6% on average the state-of-the-art HTR methods [7, 6, 5]. Moreover, as shown in Figure 1, the qualitative results demonstrate its effectiveness in recognizing handwritten text under the mentioned challenges.

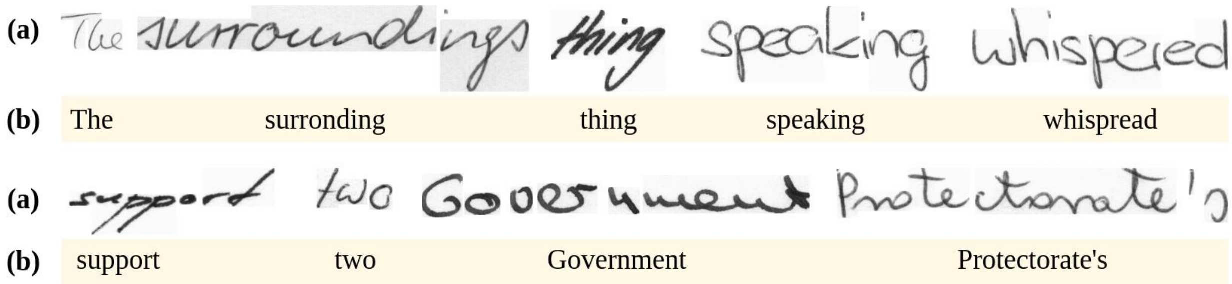

Figure 1: Qualitative results of using the proposed model on IAM HTR dataset [9], which in (a) and (b) show the inputs and the recognized text, respectively.

Future Objectives

In order to evaluate the generalization capability of our proposed model, we aim to evaluate it on different handwritten benchmark datasets as well as compare its performance to the current state-of-the-art methods.

Acknowledgments

This work was supported by the National Science and Research Council (NSERC) of Canada and the Ontario Centers of Excellence (OCE-VIP2) in conjunction with ATS Automation Tooling Systems Inc.

References

[1] Theodore Bluche and Ronaldo Messina. Gated convolutional recurrent neuralnetworks for multilingual handwriting recognition. InProc. Int. Conf. on Doc.Anal. and Recognit. (ICDAR), volume 1, pages 646–651, 2017.

[2] Ankush Gupta, Andrea Vedaldi, and Andrew Zisserman. Synthetic data for textlocalisation in natural images. InProc. IEEE Conf. on Comp. Vision and PatternRecognit., pages 2315–2324, 2016.

[3] Sepp Hochreiter and J¨urgen Schmidhuber. Long short-term memory.NeuralComput., 9(8):1735–1780, 1997.

[4] Max Jaderberg, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Syn-thetic data and artificial neural networks for natural scene text recognition.arXivpreprint arXiv:1406.2227, 2014.

[5] Lei Kang, Pau Riba, Mar¸cal Rusi˜nol, Alicia Forn´es, and Mauricio Villegas. Payattention to what you read: Non-recurrent handwritten text-line recognition.arXiv preprint arXiv:2005.13044, 2020.

[6] Lei Kang, Pau Riba, Mauricio Villegas, Alicia Forn´es, and Mar¸cal Rusi˜nol. Can-didate fusion: Integrating language modelling into a sequence-to-sequence hand-written word recognition architecture.arXiv preprint arXiv:1912.10308, 2019.

[7] Lei Kang, J Ignacio Toledo, Pau Riba, Mauricio Villegas, Alicia Forn´es, andMar¸cal Rusinol. Convolve, attend and spell: An attention-based sequence-to-sequence model for handwritten word recognition. InGerman Conf. on PatternRecognit., pages 459–472. Springer, 2018.

[8] Sri Yugandhar Manchala, Jayaram Kinthali, Kowshik Kotha, Kanithi SantoshKumar, and Jagilinki Jayalaxmi. Handwritten text recognition using deep learn-ing with tensorflow.

[9] U-V Marti and Horst Bunke. The iam-database: an english sentence database foroffline handwriting recognition.Int. J. on Doc. Anal. and Recognit., 5(1):39–46,2002.

[10] Joan Puigcerver. Are multidimensional recurrent layers really necessary for hand-written text recognition? InProc. Int. Conf. on Doc. Anal. and Recognit. (IC-DAR), volume 1, pages 67–72, 2017.

[11] Joan Andreu S´anchez, Ver´onica Romero, Alejandro H Toselli, Mauricio Villegas,and Enrique Vidal. A set of benchmarks for handwritten text recognition onhistorical documents.Pattern Recognit., 94:122–134, 2019.

[12] Guangcun Shan, Hongyu Wang, and Wei Liang. Robust encoder-decoder learn-ing framework towards offline handwritten mathematical expression recognitionbased on multi-scale deep neural network.arXiv preprint arXiv:1902.05376, 2019.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones,Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need.InAdv. in Neural Inform. Process. Sys., pages 5998–6008, 2017.