Nephi: An Open Source Pytorch Library for Handwriting Recognition

Oliver Nina
School of Electrical and Computer Engineering
Ohio State University
Columbus, OH
Email: nina.3@osu.edu

Russell Ault
College of Medicine
Ohio State University
Columbus, OH
Email: ault.95@osu.edu

Roger Pack
FamilySearch
Lehi, UT
Email: roger.pack@familysearch.org

***Abstract*—We present an open source library for handwritten text recognition (HTR) in Pytorch. The problem of offline handwriting recognition has attained greater attention recently due to significant improvements in this area [1], as well as recent relevant competitions such as [2]. Many of the current libraries and tools used in these HTR advancements have not yet been released publicly. To facilitate more rapid progress in this area, we develop an open source libraries to help more novice amateurs and professional researchers contribute to this area.**

Our proposed open source library provides a baseline performance for HTR based on convolutional recurrent neural networks (CRNN) implemented in Pytorch. Our library is a simple CRNN implementation based on a library used for scene text in the wild [3]. Our library uses primarily the READ dataset for training, although it can be easily extended to other datasets. In this paper we give a basic overview of the model used in our library and provide some results we obtained for the READ dataset.

I. Introduction

In recent years the problem of handwritten text recognition has gained increased attention as well as a substantial improvement over the historical HMM models and similar shallow models.

The advent of deep neural networks and their success in other areas such as object classification [4], image description [5], speech recognition [6] and a plethora of other problems has made the application of deep learning to handwritten text recognition rather interesting and compelling.

Moreover, in recent years, several competitions have encouraged worldwide research groups to participate more actively in solving this problem. The success of such competitions has brought about great improvements not seen for several decades before the onset of this dominance of deep learning models for HTR.

Despite the great recent improvements on deep learning models for HTR, there has been little effort to consolidate efforts into a common open source library that can be used for rapid experimentation by novices and expert users who would like to get involved in the problem of handwriting recognition.

II. Method

In this section we give a brief background of the methodology used by our library mainly the CRNN-CTC model.

A. CRNN-CTC

The CRNN or convolutional recurrent neural network model was introduced in [3] for solving typed text recognition. This was an extension to a more well known model for HTR based on column pixel based features and the RNN-CTC.

The main difference between the two models is that in CRNN the features extracted from the image, rather than being based on raw column pixels from the input images to be passed to the recurrent neural network, the image features are instead obtained by a Convolutional Neural Network (CNN) whose weights are trained together with the weights of the RNN. This behavior makes CRNN more dynamic as they learn the appropriate features per input image.

For the recurrent network (RNN) side of it an LSTM unit is used in a similar way as the traditional RNN-CTC. The gating functions for an LSTM are as follows:

$$\begin{aligned}
\mathbf{g}_i &= \sigma(\mathbf{W}^\star \mathbf{x}_i + \mathbf{U}^\star \mathbf{h}_{i-1} + \mathbf{C}^\star \boldsymbol{\nu}_i) \\
\mathbf{z}_i &= g(\mathbf{W}^z \mathbf{x}_i + \mathbf{U}^z \mathbf{h}_{i-1} + \mathbf{C}^z \boldsymbol{\nu}_i) \\
\mathbf{c}_t &= \mathbf{i}_t \odot \mathbf{z}_t + \mathbf{f}_t \odot \mathbf{c}_{t-1} \\
\mathbf{h}_t &= \mathbf{o}_t \odot g(\mathbf{c}_t),
\end{aligned} \tag{1}$$

where σ and g represent a sigmoid and a hyperbolic tangent (TANH) function respectively. The operation $\odot$ represents a Hadamard product. The input, forget and output gating functions of the LSTM unit correspond to $\mathbf{g}_t = \{\mathbf{i}_t, \mathbf{f}_t, \mathbf{o}_t\}$ with its corresponding weights $\mathbf{W}^\star$ and $\mathbf{U}^\star$, where $\star = \{i, f, o\}$.

And the probability of the CTC for each character label is given by:

$$p(\mathbf{l}|\mathbf{x}) = \sum p(\pi|\mathbf{x}) \tag{2}$$

where $\mathbf{l}$ is the character labels or a space character. $\mathbf{x}$ is the output classification and π is that different paths of alignment of the sequence of characters of the outputs.

Thus the final output is obtained by:

$$h(\mathbf{x}) = \arg\max_l p(l|\mathbf{x}) \tag{3}$$

III. Proposed Library

Our library is primarily based on the already open source CRNN library presented in [3]. Although this library was originally used for scene text recognition in the wild, the main model and framework have also been proven to be helpful to perform offline handwriting recognition with state-of-the-art accuracy [1]. Moreover, CRNN have recently been shown to allow better generalization for transfer learning in handwriting recognition than other fully recurrent state-of-the-art architectures [7].

In our library we have added several modules to help transition from scene type text recognition to HTR and provide tools to parse and incorporate well-known datasets for HTR such as the READ dataset[1]. We have added additional validation parsing tools, image resizing, and loss calculation metrics to the outputs.

In contrast to other methods that participated in the HTR competition described in [2], our framework is simple and stream-lined. Whereas some other methods perform several steps of pre-processing of the images including and not limited to binarization and data augmentation, our model rather takes any input image and performs the training and validation of a CRNN-CTC model which is the state-of-the-art deep learning model for performing HTR.

We have released our library as a Github repository :
https://github.com/olivernina/nephi

Our library is currently in beta version. We are currently adding new features as we keep improving the performance or our current baseline.

IV. Results

The model used for the results in all the images was trained for 850 epochs. Figure 1 shows best results obtained in the validation set of the READ dataset. Notice that the results in this figure are nearly perfect compared to the ground truth. This is very impressive considering the model was trained with a relative small number of epochs.

Figure 2 shows the worst results on the validation set of the READ dataset. Notice that the resizing of some of the images, deforms the text and that negatively affects the performance of the model.

Table I shows the character error rate (CER) and word error rate (WER) of our trained model. The rates in this table are a little lower compared to state of the art methods such as [1]. The reason for this is because our method is much simpler. In our method we do not employ any pre-processing steps such as binarization, warping or data augmentation as done in [1].

V. Conclusion

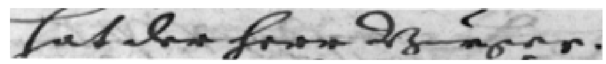
In this paper we have presented an open source library for handwriting text recognition based on the READ dataset.

The lack of open source libraries and publicly available resources in this topic, motivated us to develop a framework that we can use to create a baseline to build upon.

Our library is based on Pytorch and allows anyone interested in this topic to experiment with a method similar to the one used by state of the art methods. Thus, it should foster innovation and research contributions in this field.

[1]https://read.transkribus.eu/

(a) **-Prediction**: wegen der: 50 fl. so sein **-Ground Truth**: wegen der: 50 fl. so sein



(b) **-Prediction**: hat der herr brger **-Ground Truth**: Hat der herr Brger

(c) **-Prediction**: khaffng. vnd dardrch **-Ground Truth**: khaffng. vnd dardrch

Fig. 1. Best Results on READ Validation Set

(a) **-Prediction**: fp **-Ground Truth**: die herrn gehaimen Rth vmb Passb

(b) **-Prediction**: . **-Ground Truth**: :der:

(c) **-Prediction**: n **-Ground Truth**: :mit:

Fig. 2. Worst Results on READ Validation Set

Our preliminary results are very encouraging and show that our model can be trained on raw input images without any pre-processing steps and achieve competitive results.

TABLE I
Performance on READ Dataset

SET	CER	WER
Training	0.0580	0.0279
Validation	0.1665	0.5338

VI. Acknowledgements

We would like to thank the insightful comments and valuable feedback received by Dr. William Barrett, Dr. Douglas Kennard and Seth Stewart. We would also like the reviewers and organizers of the Family History Technology Workshop 2018 and the many other people that indirectly or directly have contributed to our project.

REFERENCES

[1] C. Wigington, S. Stewart, B. Davis, B. Barrett, B. Price, and S. Cohen, "Data augmentation for recognition of handwritten words and lines using a cnn-lstm network," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 01, Nov 2017, pp. 639–645.

[2] J. A. Sanchez, V. Romero, A. H. Toselli, M. Villegas, and E. Vidal, "Icdar2017 competition on handwritten text recognition on the read dataset," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 01, Nov 2017, pp. 1383–1388.

[3] B. Shi, X. Bai, and C. 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*, vol. 39, no. 11, pp. 2298–2304, 2017.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'12. USA: Curran Associates Inc., 2012, pp. 1097–1105. [Online]. Available: http://dl.acm.org/citation.cfm?id=2999134.2999257

[5] A. Karpathy and L. Fei-Fei, "Deep Visual-Semantic Alignments for Generating Image Descriptions," *arXiv:1412.2306 [cs]*, Dec. 2014, arXiv: 1412.2306. [Online]. Available: http://arxiv.org/abs/1412.2306

[6] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech Recognition with Deep Recurrent Neural Networks," *arXiv:1303.5778 [cs]*, Mar. 2013, arXiv: 1303.5778. [Online]. Available: http://arxiv.org/abs/1303.5778

[7] T. Bluche and R. Messina, "Gated convolutional recurrent neural networks for multilingual handwriting recognition."