Survey on Handwritten Text Recognition Using Different Deep Learning Techniques

Mr. Rahul Kumar Namdeo
M.Tech. Scholar, PG Dept. of CSE
rahulnamdev75@gmail.com
SIRTS, Bhopal

Prof. Chetan Gupta
Asst. Prof., Dept. of CSE
chetangupta.gupta1@gmail.com
SIRTS, Bhopal

Dr. Ritu Shrivastava
HOD, Dept. of CSE
ritushrivastava08@gmail.com
SIRT, Bhopal

Abstract — recognizing handwritten text is an active research topic among researchers. The introduction of deep learning improves the accuracy of text recognition. In this paper, we survey different deep neural networks for handwritten text recognition that is based on recurrent neural network (RNN) architecture. We also propose a hybrid model, a combination of convolution neural network (CNN) and recurrent neural network (RNN). CNN as feature extractor and RNN to encode the visual features and decode the sequence of letters present in the handwritten image. This paper presents a detailed review in the field of Handwritten Text Recognition.

Keywords - Handwritten Text Recognition, Convolution Neural Network, Recurrent Neural Network, BLSTM

I. INTRODUCTION

With the increasing use of digital technology nowadays, more and more data produced is of digital nature, but there are still many areas where digital technology has not reached completely. In developing nations like India, many documents in the form of handwritten documents are still produced, like post office addresses written in handwritten language, application forms in banks or documents used in government organizations are in handwritten language. To use this data and extract information from it, it is necessary to recognize this data and convert it into digital form to overcome this handwritten recognition techniques used, which has been a top research topic for many decades. Handwriting Recognition or Handwritten Text Recognition (HTR) is a challenging task as there is a high level of change in the writing style of every person in this world. Before the introduction of machine learning and deep learning, HTR was performed using pattern matching techniques such as hidden Markov model (HMM) and support vector machines (SVMs). Though the pattern matching techniques work well, their accuracy is low, so to overcome this, deep learning using different neural network architectures is explored and the result is far better than the pattern matching techniques.

A recurrent neural network (RNN) is a type of artificial neural network whose nodes are connected in a directed graph along a temporal sequence. This enables it to exhibit a temporal dynamic behavior. With their internal state (memory), RNNs can process variable length sequences of inputs. This makes them useful for tasks like recognizing connected handwriting. This work studies the effectiveness of CNN-RNN hybrid architecture by including a variety of data augmentation and normalization schemes that help the network to learn invariance specific to handwritten images. Additionally, we discuss the role of using Pre-training such deep networks with synthetic data to enhance the recognition performance.

Following is the detail of dataset used in training for HTR

Table 1: Dataset of IAM

Dataset	Words	Lines	Writers

IAM	1,15,320	13,353	657

IAM dataset: We use the IAM handwriting dataset as shown in Table 1, which contains offline handwritten content of over 1500 structures, each reflecting the writings of 657 scholars with 13,353 transcribed lines, involving 115,320 words. All related structure name metadata is provided in related XML documents after a section was sectioned and physically verified. In addition to word-level recognition, IAM is prepared to recognize accents and capitalization even when they are not considered during assessment. Our training includes all word pictures, even those with division errors. If a single word on a line has an off-base division, irrespective of whether the more words are divided accurately, the whole line is marked with a mistake. This method is robust to errors of this kind. The extra preparing information can prevent mistakes and be beneficial.

II. LITERATURE SURVEY

Geetha R et.al [1] use deep neural networks and the sequence-to-sequence (Seq2Seq) approach, this paper proposes a hybrid handwritten text recognition (H2TR) model. Convolution neural networks (CNN) and recurrent neural networks (RNN) are combined with a long-short-term memory network (LSTM) in this hybrid model. These networks are used to extract the characteristics of the handwritten image. Later, the extracted features are modeled with a sequence-to-sequence approach and fed to RNN-LSTM for encoding the visual features and decoding the sequence of letters available in the handwritten image.

Luc Mioulet [2] uses several combination methods involving various BLSTM-CTC systems. Researchers use BLSTM-CTC architecture to identify which techniques are the most relevant to improve on an isolated word recognition test (the WR2 challenge of the ICDAR2009 competition). Paper investigates three degrees of combination: early integration (feature fusion), mid level combination, and late fusion (output combinations).

Albert Zeyer et al proposed [3] a modified bidirectional RNNs is used to enable online-recognition by moving a window over the input stream and performing one forwarding through the RNN on each window. In experiments researcher results show a better performance of this online-enabled bidirectional LSTM much better than the unidirectional LSTM.

In order to recognize unconstrained offline handwritten texts, a hybrid hidden Markov model/Artificial Neural Network (ANN) model is proposed by Salvador Espan˜a-Boquera, Maria Jose Castro-Bleda, Jorge Gorbe-Moya, and Francisco Zamora-Martinez [4]. The structural part of the optical models has been modeled with Markov chains, and a Multilayer Perceptron is used to estimate the emission probabilities. This paper also presents new techniques to remove slope and with supervised learning methods, it is possible to remove the slant from handwritten text and normalize the size of text images. Slope correction and size normalization are achieved by classifying local extreme of text contours with Multilayer Perceptrons. Slant is also removed in a no uniform way by using Artificial Neural Networks.

Ronaldo Messina et.al. [5] Paper proposed Multidimensional Long-Short Term Memory Recurrent Neural Networks (MDLSTM-RNN) in recognizing lines of handwritten Chinese text without explicit segmentation of the characters. MDLSTM-RNN shows a better performance in Latin and Arabic text recognition.

Based on synthetic data and domain-specific image normalization and augmentation, Kartik Dutta et.al. [6] Presented effective methods for training CNN-RNN hybrid architecture. Each of these modules contributed to improving the recognition rates at both the line and word levels.

C. Wigington et al. [9] Paper introduces two data augmentation and normalization techniques for handwriting recognition that, when used in conjunction with a CNN-LSTM, significantly reduce Word Error Rate (WER) and Character Error Rate (CER). The researcher applies normalization and augmentation to both training and test images in the experiment and achieves a low WER and CER over hundreds of authors, multiple languages, and a variety of collections written centuries apart.

J. Puigcerver et.al. [10] Paper introduces an alternative model that uses only convolution and one-dimensional recurrent layers to generate results that are superior or equivalent to those of the present state-of-the-art architecture while also being substantially faster. Furthermore, it finds that random distortions as synthetic data augmentation during training greatly enhances the accuracy of the model.

Carlos AJ et.al. [11] Paper proposal is based on the transfer learning (TL) from the parameters learned with a bigger database. It first investigates, for a reduced and fixed number of training samples, and shows how the learning from a large database, the IAM, can be transferred to the learning of the CLC of a reduced database. It focuses on which layers of the network could be not re-trained. The Experiment shows that the best solution is to re-train the whole CLC parameters initialized to the values obtained after the training of the CLC from the larger database. As a byproduct, the learning times are quite reduced. Similar good results are obtained from the Parzival database when trained with this reduced number of lines and this new approach.

Voigtlaender P et al. [18] Paper uses an efficient GPU based implementation which greatly reduces training times by processing the input in a diagonal-wise method. It uses this implementation to explore deeper and wider architectures than previously used for handwriting recognition and show that especially the depth plays an important role. it produces better results on two databases with a deep multidimensional network.

Doetsch P et al. [19] Paper researcher uses a modified topology for long short-term memory recurrent neural networks that controls the shape of the squashing functions in gating units. It also proposes an efficient training framework based on a mini-batch training on sequence level combined with a sequence chunking approach. The framework is evaluated on publicly available data sets containing English and French handwriting by utilizing a GPU based implementation.

Word Error Rate (WER) and Character Error Rate (CER) measurement

We measured the recognition performance by comparing the output of the recognizer with the reference transcription and computing the Word Error Rate (WER). Word error rate is defined as the sum of the number of insertions, substitutions, and deletions divided by the total number of words in the reference set by the total number of words in the test set.

$$\mathrm{WER} = 100 \times \frac{\text{insertions} + \text{substitutions} + \text{deletions}}{\text{total number of words}}$$

There can only be a null WER if the output of the recognizer matches the reference transcription precisely.

We also measured the Character Error Rate (CER) for the final test experiments. Character expressions are similar to WER expressions, except that they are characterized by characters rather than words.

$$\mathrm{CER} = 100 \times \frac{\text{insertions} + \text{substitutions} + \text{deletions}}{\text{total number of character}}$$

It is highly recommended to offer not only the value of the WER (or CER) when comparing different systems as shown in Table 2 and also in

Table 3 we also compare the accuracy of different approaches . It also comes with a confidence interval.

Word Recognition Results On the IAM Dataset			
Method	**References**	**WER**	**CER**
CNN-LSTM	Carlos AJ et al.[11]	19.07	6.07
SEQ-TO-SEQ	Sueiras J et al[20]	23.8	8.8
RNN	Doetsch P et al. [19]	8.9	2.8
HMM/ANN	Salvador E et al.[4]	22	9.8
CNN-RNN HN	kartik dutta et al.[6]	17.82	5.7
CRNN-STN	P. Krishnan et al.[8]	12.61	4.8
CNN-LSTM	C. Wigington et al.[9]	19.07	6.07

Table 2: Comparasion of WER and CER

Accuracy comparison with IAM datasets				
References	**Algorithm**	**Word Accuracy %**	**Letter Accuracy %**	**Average %**
Wigington et al. [9]	CNN	94.17	96.85	95.51
Doetsch et al [10]	RNN	87.8	95.3	91.55
R. Geetha et al [1]	CNN–RNN (LSTM)	95.2	97.48	96.34

Table 3: Accuracy Comparison

III. PROBLEM DOMAIN

1. The variability and ambiguity of strokes varies greatly from person to person.
2. The handwriting style of an individual person also varies from time to time and is inconsistent.
3. Poor quality of the source document/image due to degradation over time.
4. Text in printed documents sits in a straight line, whereas humans need not write a line of text in a straight line on white paper.
5. Cursive handwriting makes separation and recognition of characters challenging.
6. Text in handwriting can have variable rotation to the right, which is in contrast to printed text where all the text sits up straight.

IV. CONCLUSION

In handwritten character recognition, much research and work has been done. So far, however, 100 percent accuracy has not been achieved, which indicates that more work needs to be done in this direction. While separate characters are more accurate, a different writing style affects word recognition. Although a holistic approach eliminates complicated segmentation, they use a limited vocabulary. Segmentation methods are less accurate because they are complex. The classifier displays good accuracy when a limited number of words are used since it is dealing with a smaller number of variations.

V. References

[1]. Geetha, R., Thilagam, T. & Padmavathy, T. “Effective Offline Handwritten Text Recognition Model based on a Sequence-to-Sequence Approach with CNN–RNN Networks”, Neural Computing and Applications, volume 33, pages 10923–10934 2021.

[2]. Luc Mioulet, G. Bideault, C. Chatelain, T. Paquet, S. Brunessaux “BLSTM-CTC

Combination Strategies for Off-line Handwriting Recognition", International Conference on Pattern Recognition Applications and Methods, 2015.

[3]. Albert Zeyer, Ralf Schlueter, Hermann Ney "Towards Online-Recognition with Deep Bidirectional LSTM Acoustic Models", Published at INTERSPEECH, 2016.

[4]. Salvador E-B, Maria JC-B, Jorge G-M, Francisco Z-M "Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models", IEEE Transactions on Pattern Analysis and Machine Intelligence ,33 (4):767–779 2011.

[5]. Ronaldo Messina, Jerome Louradour "Segmentation-free Handwritten Chinese Text Recognition with LSTM-RNN", 13th International Conference on Document Analysis and Recognition ICDAR ,2015.

[6]. Kartik Dutta, Praveen Krishnan, Minesh Mathew; C.V. Jawahar "Improving CNN-RNN Hybrid Networks for Handwriting Recognition", 16th International Conference on Frontiers in Handwriting Recognition ICFHR, 2018.

[7]. U.-V. Marti and H. Bunke, "The IAM-database: An English Sentence Database for Offline Handwriting Recognition", IJDAR, 2002.

[8]. P. Krishnan, K. Dutta, and C. V. Jawahar, "Word Spotting and Recognition using Deep Embedding", DAS, 2018.

[9]. 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 ICDAR, 2017.

[10]. J. Puigcerver, "Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition?", ICDAR, 2017.

[11]. José Carlos Aradillas, Juan José Murillo-Fuentes, Pablo M. Olmos "Boosting Handwriting Text Recognition in Small Databases with Transfer Learning", International conference on frontiers in Handwriting Recognition ICFHR Conference, 2018.

[12]. H Haşim Sak, Andrew Senior, Françoise Beaufays "Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition",arXiv preprint: 1402.1128, 2014.

[13]. Albert Zeyer, Patrick Doetsch, Paul Voigtlaender, Ralf Schlüter, Hermann Ney "A Comprehensive Study of Deep Bidirectional LSTM RNNs for Acoustic Modeling in Speech Recognition" ,arXiv:1606.06871 2016.

[14]. Graves, A., Fernandez, S., Gomez, F., Schmidhuber, J. "Connectionist Temporal Classification: Labelling Unsegmented Sequential Data with Recurrent Neural Networks". In 23rd Int'l Conf. on Machine Learning ICML, pp. 369–376 2006.

[15]. Anita P, Dayashankar S "Handwritten English Character Recognition using Neural Network", International Journal of Computer Science & Communication, Vol.1(2):141–144 2010.

[16]. Bengio Y "Learning Deep Architectures for AI",Foundations and Trends in Machine Learning ,Vol. 2(1):1–127, 2009.

[17]. Alex Krizhevsky,Ilya Sutskever, Geoffrey E. Hinton "Imagenet Classification with Deep Convolution Neural Networks",NIPS, 2012.

[18]. Paul Voigtlaender; Patrick Doetsch; Hermann Ney "Handwriting Recognition with Large Multidimensional Long Short-Term Memory Recurrent Neural Networks", ICFHR 2016.

[19]. Doetsch P, Kozielski M, Ney H “Fast and Robust Training of Recurrent Neural Networks for Offline Handwriting Recognition”, 14th international conference on frontiers in Handwriting Recognition, Heraklion, 2014, pp 279–284. ICFHR, 2014.

[20]. Sueiras J, Ruiz V, Sanchez A, Velez JF “Offline Continuous Handwriting Recognition using Sequence to Sequence Neural Networks” ,Neuro computing, 2018.