2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT) | 978-1-7281-9772-2/20/$31.00 ©2020 IEEE | DOI: 10.1109/RTEICT49044.2020.9315679

Handwritten Text Recognition using Deep Learning

Nikitha A
Computer Science
Ramaiah Institute of Technology
Bangalore, India
sahithyaniki@gmail.com

Dr. Geetha J
Computer Science
Ramaiah Institute of Technology
Bangalore, India
geetha@msrit.edu

Dr. JayaLakshmi D.S
Computer Science
Ramaiah Institute of Technology
Bangalore, India
jayalakshmids@msrit.edu

***Abstract*—There are many researchers working on handwritten text recognition (HTR) and also contributing to HTR domain. Even though many research methods are existing for HTR, there is a need for some more improvements in the accuracy of the HTR systems. This paper is a contribution of the application of the Deep Learning algorithm for the HTR system. In this paper first we will collect the data for training the handwritten texts, later features have been extracted from those text datasets and perform training of the model using Deep Learning approach. In this work we are going to use the strategy to recognize in terms of words rather those characters so that accuracy will be improved. The built model using LSTM deep model achieves a very good accuracy. Lastly, this developed approach of the HTR system is integrated into the OCR system and comparison of results are reported in this paper. Two approaches have been compared in this paper on IAM handwritten data set, and found that 2DLSTM based approach outperforms the other approach.**

***Keywords*—LSTM, Deep Learning, Convolutional Neural Network, 2DLSTM.**

I. INTRODUCTION

This system is built to recognize handwritten text and then convert the recognized into digital form.

A. Overview

Although there exists many technologies that can be used in the creation of text based documents, many people tend to use pen and paper and create physical documents which are important. Data storage and physical data retrieval of traditional documents is a most challenging problem. Handwriting recognition is growing rapidly in the present globalization. Handwriting recognition is the ability of the computer to trans- late the human writing to digital form. Handwritten recognition is a technology that can be used to recognize handwritten characters and also can be deployed on other devices like PDA and tablet PCs. Advantages of handwritten recognition is that electronic storage can be adopted which requires fewer employees to sort the documents and to organize the documents. Other than that electronic data retrieval can be performed. Moreover, another advantage is that ancient preservations. Old family records, personnel dairies which might be corrupted due to accidents or any other reasons that is when handwriting recognition software is very helpful. Apart from data storage and data retrieval the most challenging problem associated with handwriting recognition is that most of the people tend to write with their style and dealing with accuracy is also a major problem. This system which we have implemented provides solutions to tackle such problems. We have provided a solution using a stack of bidirectional LSTMs (BLSTMs) as they provide better results. Handwriting text will be in various different types of image formats: handwritten notes, historical documents, whiteboards, medical records etc. There are many applications of Automated HTR system, a need for complete Optical Character Recognition (OCR) solution has to be included in HTR in images. This motivates the need for the research in the domain of HTR and OCR systems.

B. Deep Learning

Deep Learning is a subdivision pertaining to Machine Learning, which is also a subdivision pertaining to Artificial Intelligence. Artificial intelligence is an approach which empowers the system to mirror human behaviors. Machine Learning is an approach to attain Artificial Intelligence using algorithms trained along datasets and lastly Deep Learning is a kind of Machine Learning influenced by the complexion of the human brain. Deep Learning is a Machine Learning approach that grasps features and tasks directly from data. Data can be images, text, or sound. Deep Learning is generally indicated as end to end learning with Deep Learning features picked up with Neural Network without human intervention. Deep Neural Networks requires hours or even months to train. The training time increases with the amount of data and the number of layers in the network.

C. Convolutional Neural Network

The Convolutional Neural Network (CNN) is a deep learning network that is used for image classification. CNN's basic concept is to use predefined convertible filters to distinguish patterns in image edges, parts of objects, and expand on this information to detect complete objects such as animals, humans, cars, etc. Hubel and Wiesel's reporting on CNN has been going on since 1959. The models worked but models were not able to automate learning. CNNs are mainly used to identify images, cluster and classify images, detect objects, etc. CNNs use comparatively less preprocessing than other image processing algorithms. The CNN's communication pattern resembles the visual cortex structure of animals. Compared to other algorithms for the image classification, CNNs use very little preprocessing. This means the network is learning the filters which have been hand-engineered in conventional algorithms. This freedom from previous expertise and human initiative in the design of features is a huge gain.

D. Overall Description

Handwritten Text Recognition HTR, is a method which allows us to convert documents, like scanned image or images browsed by the user into digital form. This method is very helpful as it minimizes time without the necessity of re-typing the documents. It can perform the action in few minutes. It can recognize texts the in pictures and convert it into the digital form using a simplified process as illustrated in figure 1. This process generally consists of three stages: Open the file, recognize the data and then display in a convenient format.

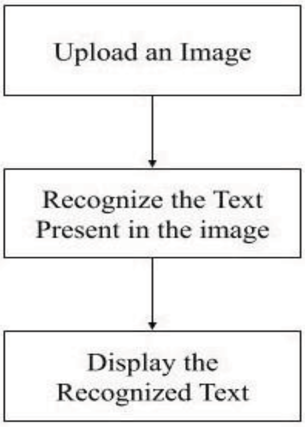

Fig. 1. HTR Process

II. LITERATURE SURVEY

Humans have been writing their thoughts for quite a long time in the form of letter, transcripts etc.; for dissemination to others. But since computer technology has evolved the format of handwritten text quickly transferred to digital generated machine Text and so people feel the need for such a way of transforming of handwritten text to digital text, as processing those data are very simple and convenient. There have been more researchers who are working on handwriting text recognition (HTR) techniques. Some of research is mentioned below.

A. . Hand-written Text Recognition

They have followed the usage of line level strategies by combining the Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) [4] recurrent neural networks. These techniques have been used to extract features, and train the model with Connectionist Temporal Classification (CTC [3]). There are many data-driven deep-learning based approaches which will extract and select prominent features to be employed from the training sample set, whereas in traditional methods which employ hand-engineered features. However above said techniques have resulted very good improvements in recognizing handwritten text on public data set [9], [8], scaling up these approaches for the purpose of supporting the trending disciplines or to a new coding languages will be extremely challenging due to higher costs and difficulties present in obtaining and labeling a dataset of handwriting training data set. By literature survey it is understood that the challenging issue in developing an HTR system is collecting and labeling the training and testing data set of handwritten text.

This paper is a contribution to the field of HTR and OCR and also tries to address the common issues by using large training data of handwriting dataset: The data set from online HTR system [10- 11] has been used to training the model to recognize line text in the existing HTR system [12]. The data is a series of strokes (x, y) coordinates along the timestamp of the users writing using their fingers or a stylus on the screens.

The HTR is crucial with the necessity of invading handwritten texts into the devices in a world where a big number of humans can acquire an access to HTR services using mobility devices. Based on the literature survey, this is the primary work which explains the interrelationship between online HTR systems with offline handwriting recognition with reusing an existing training dataset to build a fully functional HTR system. HTR presented in this paper [14] is trained on images which are used from the trajectory data of the online HTR systems. In order to obtain acceptable accuracy in the real time scenarios for handwritten text, adoptive image degradation techniques have been used for generating training datasets for HTR to OCR systems [14]. After enhanced handwritten images are rendered, the handwritten images are preprocessed by using an image degradation algorithm which has been applied to realistic images. In addition to the collecting online handwritten dataset, also handwriting synthetic image data set has been used to increase the variability in the training dataset to achieve better output. This paper demonstrated the feasibility of this aforementioned technique for Latin script, but there is strong proof that the same approach with pre- processing operation will work for other languages also.

For the purpose of line recognition for HTR systems, many researchers have experimented with different architectures: one among is LSTM technique, similar to many of the existing well proved approaches. However, one primary issue with respect to recurrent models is that recurrent models will not perform training operations and cannot run easily on any specialized hardware as the Feed Forward Networks. Therefore many researchers proposed a fully Feed Forward Network model which can achieve a good comparative accuracy as the LSTM based architectures.

However the handwritten text line recognition approach is the dominant step in many HTR systems also it's merely one of the important components of a complete handwritten text recognition system. We will outline the new steps that have been used to combine Handwritten text recognition reinforce into a text recognition system consisting of text recognition, directions detection, scripts identification, and text line recognition models [15].

III. PROPOSED METHODOLOGY

A. Problem Statement

The problem with respect to handwritten text recognition is very complex, and even now there is no single approach that solves the problem of handwritten text recognition efficiently. Recent works accomplished until now use the strategy of recognizing in terms of character and the accuracy obtained by the recent works is not up to the mark or not with good accuracy.

B. Methodology

The project follows an approach of developing the Handwritten Text Recognition (HTR) such that the handwritten text can be recognized accurately. In this project first we will collect the data for training the handwritten texts, later features have been extracted from those text datasets and perform training of the model using Deep Learning approach. In this work we are going to use the strategy to recognize in terms of words rather those characters so that accuracy will be improved. The built model using LSTM deep model achieves a very good accuracy. The general methodology of HTR is demonstrated in figure 1. This figure presents the workflow of HTR systems. It involves pre-processing, training and classification steps as shown in figure 2. Each of the pre- processing steps are explained below.

C. Pre-Processing

Whenever a document gets scanned or the original data is specified, this could require some preliminary processing. The pre-processing helps to create the final version of the document that will eventually be processed by a handwritten method of text recognition. The major pre-processing goals are: Noise Removal, Segmentation, Thinning, Binarization, Normalization

1) Noise Removal: During scanning, the input can contain different types of noise that can be unacceptable during pro- cessing. The unnecessary pixels in the image can be noise that is a black pixel where white is always required, and conversely. The input image as shown in fig 2 contains background noise. So before continued processing, some of this noise should indeed be eliminated. A median filter with a filter size of approximately 3 x 3 was chosen from the various noise reduction algorithm. As shown in figure 4 the background from the input image has been eliminated using a noise reduction algorithm that is median filtering.

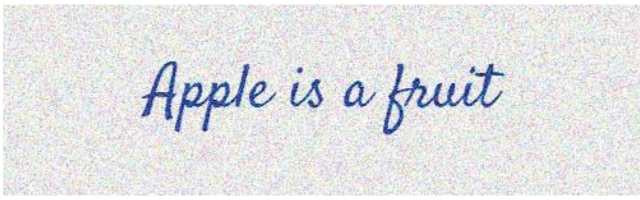

Fig. 2. Noise in the Image

Fig. 3. Noise being eliminated from the Input Image

2) Segmentation: The method of separating the words is segmentation. In HTR, it is very essential because only one word at a point can be recognized by the framework. The basic connectivity principle was used. In this principle, at some point, pixels of a words are considered to be related to each other. Segmentation also consists of the separation process between the image and the background.

In this context in order for the image to be transformed to black and white, we would need to transform a color image into a grey scale. To isolate a word, the picture must be clipped. We may use a fixed- size window, which is equal to the size of a single word, for clipping. As per the requirement, the window therefore breaks a single word. The fig 4 shows segmented words.

Fig. 4. Segmented Words

3) Binarization: Binarization is also an essential step in the processing of images wherein the image pixels are classified into two parts: background that is white and foreground that is black. A binary image can only be found in two shades that is white and black. Therefore a global gray scale intensity threshold is used in the proposed binarization strategy. The binarized image is shown in fig 5.

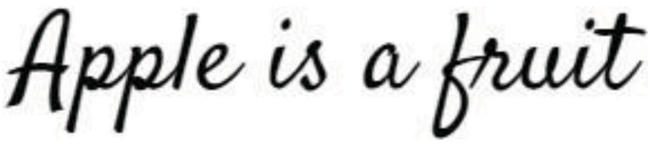

Fig. 5. Binarized Image

4) Normalization: The framework should have the ability to identify various font sizes. So, for that before sending it to the classifier, all incoming characters must be translated into a regular size. The method of changing the picture size to such a size that is accepted by the classifier is called normalization. The neural network comprises of input layers that receives image pixels as the input, for which the number of pixels is always set.

5) Classification: Just one the preprocessing is been completed, output from the pre-processing steps that is an image with a standard size is sent to a classifier. The pixel location of the word is used as the input for the classifier. In a neural network, there is indeed a unique algorithm that could be utilized for classification, but here the only algorithm for back propagation has been used.

D. LSTM Based Model

For modern HTR systems mostly LSTM and other Recurrent Neural Networks (RNN) have been prominently used for text line recognition steps. This paper presented a model, motivated by the CLDNN (Convolutions, LSTMs, and Deep Networks). For all the CNN layers, the inception architectures [20] has been used. For the LSTM layers, four stacked bidirectional LSTMs (BLSTMs) have been used. This is the way, how the model has only a feed- forward network. This paper reported the method with combination achieves sufficient accuracy in LSTM-complaint systems.

E. Training Dataset

To demonstrate the performance of the HTR system, IAM online handwritten dataset, online handwritten dataset and offline handwritten datasets samples are being used. The IAM offline handwriting database [30] (offline) which consist of scanned documents of handwritten text images, which is written by around 500 different persons using different prompt from Lancaster Oslo/Bergen (LOB) texts corpus. The dataset consists of pictures with text lines, which will be grouped based on different training samplings, acceptance, and test datasets respectively. These experiments make use of a combination of various sources of datasets to acquire better accuracy in Handwritten Text Recognition systems: Researchers have access to a considerable bulk of ink datasets which is used to develop an HTR model [10], [11] in different languages. For the purpose of experimentation and reporting the results in this paper, only Latin script is being considered but also planning to extend this system to other languages also. The dataset is feeded into images, pre-processing steps using an image pre-processing pipeline is described in the paper. Later those pre-processed images have been processed using the same degrading operation pipeline, which can be used to train an HTR system on an artificial synthetic data set. In order to increase the accuracy with respect to text recognition on printed image text, the degraded artificial synthetic data set has been used. In addition to this data set also historic data image sets from many publicly available datasets have been utilized in order to increase the accuracy with respect to text recognition [23]–[26]. And also a minor amount of modern image dataset of handwritten text shall also be used.

F. General Design

The generalized design of the system is given in fig 6. The process of recognition consists of uploading an image which is sent for the pre-processing steps so that the image is normalized to a size which can be accepted by the classifier.

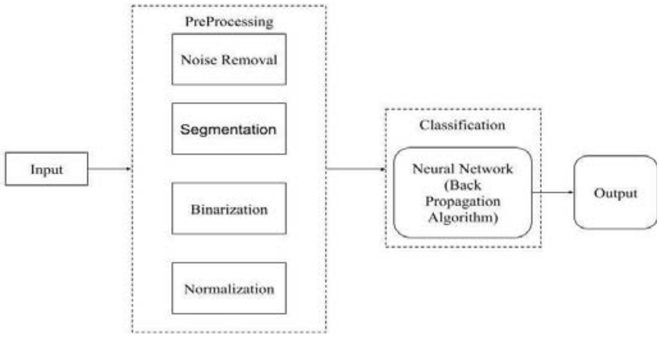

Fig. 6. General Design of the System

IV. RESULTS AND DISCUSSION

The goal of HTR is to recognize handwritten characters using deep learning. Using a recurrent neural network and LSTM model, the handwritten text recognition system is implemented. The original image in this framework is trans- formed into a gray scale image after the gray scale image, after which the text in the image is segmented, accompanied by training and recognition. The fig 7 shows the recognized text. The built model using LSTM deep model achieves a very good accuracy of 94 percent.

Fig. 7. Recognized Text

Table 1 provides a comparison of the accuracy of the proposed method with previously recorded algorithms and also compares the performance. Character Error Rate (CER) and Words Error Rate (WER) obtained by using the 2DLSTM methods is 8.2 percent and 27.5 percent respectively. Character Error Rate (CER) and Word Error Rate (WER) obtained by using the CNN-1DLSTM CTC is 6.2 percent and 20.5 percent respectively. While LSTM is leading in providing word level accuracy, with 2DLSTM the accuracy of the character level is slightly reduced in contrast. It means that the LSTM model appears to make increased spelling errors in the words that have already been mislabeled, but ultimately it produces less spelling errors at the aggregate word level.

Table:1 Comparison of HTR Method

Methods	Character Error Rate	Word Error Rate
2DLSTM	8.2	27.5
CNN-1DLSTM- CTC	6.2	20.5

V. CONCLUSION

This paper presented the approach to build a handwritten text recognition system which is scalable in future. The built model achieves sufficient accuracy for both printed as well as the handwritten text when related to specific handwritten text recognition models. In this paper only English language is considered and the same is reported in this result section of this paper. Experimentation has been performed on the IAM handwritten text data set to evaluate the performance of the two well-known approaches namely, CNN-1DLSTM-CTC and 2DLSTM. This paper reported that the LSTM based model for HTR systems outperforms other methods. In future work many different languages will be considered to evaluate the performance of the presented approach of a handwritten HTR system.

REFERENCES

[1]. Shahbaz Hassan, Ayesha Irfan, Ali Mirza, Imran Siddiqi, ”Cursive Handwritten Text Recognition using Bi-Directional LSTMs: A case study on Urdu Handwriting”, in Deep-ML, 2019.

[2]. Herleen Kour;Naveen Kumar Gondhi,”Machine Learning approaches for Nastaliq style Urdu handwritten recognition: A survey” in ICACCS, 2020.

[3]. A. Graves, S. Fernandez, F. Gomez, and J. Schmidhuber, “Connection- ist temporal classification: Labelling unsegmented sequence data with recurrent neural networks,” in ICML, 2006.

[4]. A. Graves and J. Schmidhuber, “Offline handwriting recognition with multidimensional recurrent neural networks,” in NIPS, 2009.

[5]. Rohan Vaidya;Darshan Trivedi;Sagar Satra;Prof. Mrunalini Pimpale, ”Handwritten Character Recognition Using Deep-Learning”, in ICICCT, 2018

[6]. Gideon Maillette de Buy Wenniger;Lambert Schomaker;Andy Way,”No Padding Please: Efficient Neural Handwriting Recognition”, in ICDAR 2019

[7]. A. Graves, M. Liwicki, S. Fernndez, R. Bertolami, H. Bunke, and J. Schmidhuber, “A novel connectionist system for unconstrained hand- writing recognition,” PAMI, vol. 31, no. 5, pp. 855–868, 2009.

[8]. V. Pham, T. Bluche, C. Kermorvant, and J. Louradour, “Dropout improves recurrent neural networks for handwriting recognition,” in ICFHR, 2014.

[9]. P. Voigtlaender, P. Doetsch, and H. Ney, “Handwriting recognition with large multidimensional long short-term memory recurrent neural networks,” in ICFHR, 2016.

[10]. T. Bluche and R. Messina, “Gated convolutional recurrent neural net- works for multilingual handwriting recognition,” in ICDAR, vol. 01, 2017.

[11]. J. Puigcerver, “Are multidimensional recurrent layers really necessary for handwritten text recognition?” in ICDAR, 2017.

[12]. D. Castro, B. L. D. Bezerra, and M. Valena, “Boosting the deep multidi- mensional long-short-term memory network for handwritten recognition systems,” in ICFHR, 2018.

[13]. D. Keysers, T. Deselaers, H. A. Rowley, L. Wang, and V. Carbune, “Multi-language online handwriting recognition,” PAMI, vol. 39, no. 6, pp. 1180–1194, 2017.

[14]. V. Carbune, P. Gonnet, T. Deselaers, H. A. Rowley, A. Daryin, M. Calvo, L.-L. Wang, D. Keysers, S. Feuz, and P. Gervais, “Fast multi-language lstm-based online handwriting recognition,” ArXiV, 2019.

[15]. J. Walker, Y. Fujii, and A. C. Popat, “A web-based ocr service for documents,” in DAS, Apr 2018, pp. 21–22.

[16]. S. Ghosh and A. Joshi, “Text entry in indian languages on mobile: User perspectives,” in India HCI, 2014.

[17]. T. M. Breuel, “Tutorial on ocr and layout analysis,” in DAS, 2018. [18]Y. Fujii, K. Driesen, J. Baccash, A. Hurst, and A. C. Popat, “Sequence-to-label script identification for multilingual ocr,” in ICDAR, 2017

[18]. M. Kozielski, P. Doetsch, and H. Ney, “Improvements in rwth’s system for off-line handwriting recognition,” in ICDAR, 2013.

[19]. P. Doetsch, M. Kozielski, and H. Ney, “Fast and robust training of re- current neural networks for offline handwriting recognition,” in ICFHR, 2014.

[20]. P. Voigtlaender, P. Doetsch, S. Wiesler, R. Schlter, and H. Ney, “Sequence-discriminative training of recurrent neural networks,” in ICASSP, 2015.

[21]. T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, “Convolutional, long short-term memory, fully connected deep neural networks,” in ICASSP, 2015.

[22]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in CVPR, 2015.

[23]. J. Wang and X. Hu, “Gated recurrent convolution neural network for ocr,” in NIPS, 2017.

[24]. M. Liang and X. Hu, “Recurrent convolutional neural network for object recognition,” in CVPR, 2015, pp. 3367–3375.

[25]. J. Sa ́nchez, A. Toselli, V. Romero, and E. Vidal, “ICDAR 2015.Competition HTRtS: Handwritten Text Recognition on the tranScriptorium Dataset,” in ICDAR, 2015.

[26]. A. Toselli, V. Romero, M. Villegas, E. Vidal, and J. Sa ́nchez, “ICFHR,2016 Competition on Handwritten Text Recognition on the READ Dataset,” in ICFHR, 2016.

[27]. J. Sa ́nchez, V. Romero, A. Toselli, M. Villegas, and E. Vidal, “ICDAR2017 Competition on Handwritten Text Recognition on the READ Dataset,” in ICDAR, 2017.

[28]. Folger Shakespeare Library, “Early Modern Manuscripts Online (EMMO).” [Online]. Available: https://emmo.folger.edu

[29]. K. Chen, L. Tian, H. Ding, M. Cai, L. Sun, S. Liang, and Q. Huo, “A compact cnn-dblstm based character model for online handwritten chinese text recognition,” in ICDAR, 2017.

[30]. A. Graves, “Generating sequences with recurrent neural networks,” ArXiV, 2013.

[31]. Adhesh Garg;Diwanshi Gupta;Sanjay Saxena;Parimi Praveen Sa- hadev,”Validation of Random Dataset Using an Efficient CNN Model Trained on MNIST Handwritten Dataset”, in SPIN 2019.

[32]. K. Asha;H.K. Krishnappa,”Kannada Handwritten Document Recogni- tion using Convolutional Neural Network” in CSITSS, 2018

[33]. Thomas M. Breuel, ”High Performance Text Recognition Using a Hybrid Convolutional-LSTM Implementation”, in ICDAR 2017.