IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 5 (May. 2019), ||S (XI) || PP 27-30

Enhancing the Recognition of Handwritten Scripts Using CNN-RNN Hybrid Networks

Pranati Paidipati[1], Dr. Sachin Choudhari[2], Mr. Ashish Kumbhare[3]
M.Tech, Department of Computer Science & Engineering,
Jhulelal Institute of Technology, Nagpur, India

Abstract : *Deep learning is the authoritative reach in AI that utilizes neural models inside and out to mirror the elements of human mind, neural systems work in comparable example comprising many concealed layers. Hence, handwritten text recognition is the capacity to transliterate the content info encased in documents or pictures into digital content. The content example can differ from language to language. Handwritten content includes wide set of variations, for instance, couple of dialects have characters disconnected from one another while couple of dialects incorporate cursive organizations. Along these lines, making it profoundly challenging to precisely distinguish transcribed contents. In this paper, we find the factors that emphasize on improving handwritten original copies for scanned offline images. The proposed CRNN hybrid network has its point of convergence on effectual training of datasets by (i) efficient pre-processing of data, (ii) image normalization with integration of data augmentation, (iii) classification and clustering of data to get exact lexicon word with respectable dimensions of accuracy. The cutting edge approaches focus on extracting features by eliminating distortions plus noise, and later predict the possible outcomes of that specific character, be it in any fluctuated arrangement.*

Keywords: *Handwritten Text Recognition, Machine Learning, CNN-RNN networks, Deep Learning, Data Augmentation, Image pre-processing,*

I. Introduction

Handwritten text recognition is the process of automatic conversion of handwritten text into machine-encoded text. It has been a popular research area for many years due to various applications such as digitizing handwritten manuscripts, postal automation, etc. Extracting textual information from natural images is a challenging problem with many practical applications. As a result, many text detection and recognition systems rely on cleverly hand-engineered features to represent the underlying data.

With better algorithms and technologies, we move a step closer to address the problem of content level access to ancient historical books and manuscripts which were written by hand and digitized in the form of scanned images as part of modern digital library projects. Apart from historical databases, if modern age handwritten scripts are considered then it is observed that the major challenges in recognizing text from handwritten images comes from the inherent variability in data [8]. Every individual has a different style of writing and moreover, depending on the various underlying factors, even the style of a single person also changes in different instances of writings. In this paper, we approach the challenges using a hybrid network, i.e., the architectures of convolutional neural networks (CNN) and recurrent neural networks (RNN) with large annotated data and increased computational capacity of GPU's. The Convolutional Recurrent Neural Network (CRNN) acquires the advantages from Convolutional Neural Network (CNN) for local feature extraction of trained data and Recurrent Neural Networks (RNN) for chronological summarization of identified features. This neural network model depicts decent levels of efficiency, and exemplifies strong performance with huge datasets.

II. Related Work

With regards to current systems, center is around acquiring great capable outcomes for perceiving written by hand contents. In the course of the most recent decade, AI has been esteemed to show striking arrangements regarding transcribed content recognition. The fundamental motto following the examination of neural systems was the sharp want to structure a machine that works undifferentiated from human mind.

In this range of research, it has been seen that the characters like digits or letter sets are treated as parallel designs while the human mind can without much of a stretch recognize them , as an outcome making it a need to produce calculations that can undoubtedly order characters that are human-composed [7].

Keeping in tune with judicious technology update, a research by [11] depicts that multidimensional recurrent neural networks have been widely adopted by Handwritten Text Recognition (HTR) community. Though being computationally very expensive but are successful in line level HTR.

Studies have observed that a mixture of three different architectures namely, (i) Convolutional Neural Network, (ii) Recurrent Neural Network and (iii) Connectionist Temporal Classification (CTC) output layer help in unequivocal segmentation free recognition of online manually written contents autonomous of a particular vocabulary. [13]

A research by [10] states that the CNN-RNN hybrid network has a few well defined favorable benefits over customary neural system models as, it doesn't require point by point explanations and requires just height normalization in both preparing and testing stages.

It has been seen in a research [9] that for effective preparing of deep learning design, the accessibility of immense training data is essential, as any ordinary architecture comprises of a millions of parameters.

However, a comprehensive literature study over various methodologies for handwritten text recognition research papers evaluates and funnels the facts that there are many algorithms that have been simulated to recognize handwritten text scripts of varied languages individually either online or offline, and apparently making it necessary to create a generalized model to recognize human written scripts with better levels of accuracy and efficiency.

III. Current Implementation

In this paper, we employ Convolutional Neural Network – Recurrent Neural Network (CNN –RNN) , a hybrid architecture that combines the advantages of both the neural systems.

A. Convolutional Neural Networks

A CNN comprises of various convolutional and sub-inspecting layers alternatively pursued by completely associated layers. The information is a m x m x r picture where m is the tallness of picture, n is the width of the picture and r is the quantity of channels. Any convolutional layer will have k channels (or bits) of size n x n x q where n is littler than the component of picture and q can be like the quantity of channels 'r' or littler, 'q' may differ for each channel. The measure of channels offer ascent to privately associated structure which are then convolved to produce a k include guide of size m-n+1. Each element map is then sub examined regularly with mean or max pooling over p x p infectious areas where p goes between 2 for little pictures and under 5 for bigger pictures of transcribed contents.

A ConvNet gainfully catches the spatial and fleeting conditions in a picture through the utilization of significant channels. The real goal of a ConvNet is to remove abnormal state highlights, for example, edges from an information picture.

We have sustained the input image into the CNN layers. These layers are prepared to extricate important highlights from the picture. Each layer comprises of three activities. In the first place, the convolution task, this applies a channel piece of size 5×5 in the initial two layers and 3×3 in the last three layers to the info. At that point, the non-straight RELU work is connected. At last, a pooling layer condenses picture districts and yields a cut back form of the information. While the picture tallness is scaled back by 2 in each layer, highlight maps (channels) are included; with the goal that the yield includes guide (or grouping) has a size of 32×256.

B. Recurrent Neural Networks

A recurrent neural can thought of as a system where various duplicates of same system being passed as a message to the successor. RNN works on sequences of subjective lengths, crossing from start to end. Moreover, RNN back-propagates error differentials to its input, for example the Convolutional layer, enabling us to together train the repetitive layers and the convolutional layers in a brought together network. The LSTM is a specific sort of RNN that joins proficient back proliferation elements.

In the present usage of handwritten text recognition framework, the element succession contains 256 highlights for each time-step, the RNN propagates important data through this sequence. The well known Bidirectional Long Short-Term Memory (BLSTM) implementation of RNNs is utilized, as it can proliferate data through longer separations and gives increasingly powerful training characteristics. The RNN output sequence is mapped to a framework of size 32×80. The IAM dataset comprises of 79 unique characters, further one extra character is required for the CTC activity (CTC clear name), and consequently there are 80 passages for every one of the 32 time-steps.

C. Connectionist Temporal Classification

The CTC layer converts the predictions generated by the BLSTM output layer as a maximum probable label sequence for the input. One of the key advantages of the above framework is that the input images need not be resized to a fixed size, thus avoiding distortion in the aspect ratio, since both convolutional and recurrent layers can operate with variable size images and feature sequences respectively.

D. Data

Input: It is a gray value image of size 128×32. Generally, the pictures from the dataset don't have precisely this size, hence we resize it (without bending) until it either has a width of 128 or a stature of 32. At that point, we duplicate the picture into a (white) target picture of size 128×32. At last, we standardize the dim estimations of the picture which disentangles the errand for the NN. Information growth can without much of a stretch be incorporated by replicating the picture to irregular positions as opposed to adjusting it to one side or by arbitrarily resizing the picture.

CNN Output: the output of the CNN layers which is a succession of length 32. Every section contains 256 highlights. Obviously, these highlights are additionally handled by the RNN layers, nonetheless, a few highlights as of now demonstrate a high relationship with certain abnormal state properties of the info picture: there are highlights which have a high connection with characters (for example "e"), or with copy characters (for example "tt"), or with character-properties, for example, circles (as contained in transcribed "l"s or "e"s).

RNN Output: the RNN output framework for a picture containing the content. The lattice contains the scores for the characters including the CTC clear name as its last (80th) passage. The other grid passages, through and through, relate to the accompanying characters: " !"#&'()*+,./0123456789:;
ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz".

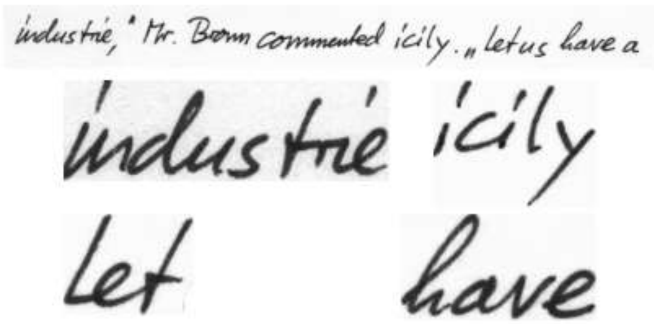

Figure 1: Few sample dataset images from IAM datasets that have been used to train data using the proposed CNN-RNN hybrid neural network.

IV. Results and Discussions

The subjective aftereffects of the CNN-RNN hybrid architecture on IAM datasets and Devanagari contents are quite decent. Calibrating the system layers and custom datasets have helped in improving the consequences of manually written content recognition regardless of the vocabulary utilized. The individual blend of points of interest of both the neural system has likewise helped in improving the outcomes at both line and word levels.

References

[1]. Kian Peymani, Mohsen Soryani , *From machine generated to handwritten character recognition; a deep learning approach* , IEEE 2017 3rd *International Conference on Pattern Recognition and Image Analysis (IPRIA)*

[2]. Meiyin Wu, Li Chen, *Image recognition based on deep learning, 2015 Chinese Automation Congress*

[3]. Mahmoud M. Abu Ghosh, Ashraf Y.Maghari, *A comparative Study on Handwritten Digit Recognition using Neural Network, 2017 International Conference on Promising Electronic Technologies*.

[4]. Martin Rajnoha, Radim Burget, Malay Kishore Dutta, *Handwritten Comenia Script Recognition with Convolutional Neural Network,* IEEE 2017 *40th International Conference on Telecommunication and Signal Processing.*

[5]. T.K Das, Asis Kumar Tripathy and Alekha Kumar Mishra, *Optical Character Recognition using Artificial Neural Network, 2017 International Conference on Computer Communication and Information (ICCCI-2017)*

[6]. Md Zahangir Alom, Paheding Sidike, Mahmudul Hasan, Tarek M.Tanha and Vijayan K. Ansari, *Handwritten Bangla Character Recognition using State-of-the-Art Deep Convolutional Neural Networks, Hindawi Computational Intelligence and Nanoscience*, Volume 2018.

[7]. Sunil kumar, Krishan Kumar, Rahul Kumar Mishra, *Scene Text Recognition using Artifical Neural Network: A survey, International Journal of Computer Applications*, March 2016.

[8]. Kartik Dutta, Praveen Krishnan, Minesh Mathew, C.V. Jawahar, *Offline Handwritten Recognition on Devanagari using a Benchmark Dataset, 2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*

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

[10]. Baoguang Shi, Xiang Bai, Cong Yao, *An End-to-End Trainable Neural Network for Image Based Sequence Recognition and its Application to Scene Text Recognition,2017 IEEE Transactions on Pattern Analysis and Machine Intelligence*

[11]. Saniya Ansari, Udaysingh Sutar, Devanagari *Handwritten Character Recognition using Hybrid Features Extraction and Feed Forward Neural Network Clasifier, Nov. 2015, International Journal of Computer Applications*

[12]. Partha S. Mukherjee, Ujjwal Bhattacharya, Swapnil K.Parui, Bappaditya Chakraborthy, *A Hybrid Model for End to End Online Handwriting Recognition, 2017 14th IAPR International Conference on Document Analysis and Recognition*.