Grenze International Journal of Engineering and Technology, Jan Issue

Hand Written Text Recognition using Convolutional Neural Network and Recurrent Neural Network

Mrs. Shankari N[1] and Mr. Sunil Kumar Aithal[2]
[1]Assistant Professor, NITTE (Deemed to be University), Dept. of ECE, NMAM Institute of Technology, Nitte, India
Email: shankarimn@nitte.edu.in
[2]Assistant Professor, NITTE (Deemed to be University), Dept. of CSE, NMAM Institute of Technology, Nitte, India
Email: sunilaithal@nitte.edu.in

***Abstract*—Character recognition is one of the developing research fields in Machine Learning. Most of the ancient famous literatures written by many famous writers are in the form of hand written text. Most challenging work of researcher is to extract these hand written texts and store digitally. In hand written character recognition, the characters are extracted and computers can be able to read and edit texts. The character recognition improves the man-machine interface for various applications which in turn helps for the development of fully automated system. Traditional approach of and written character recognition involves Support Vector Machine (SVM) classifiers to differentiate between the writers and some of the features like shape of the letters, spacing between the letters were extracted. In this paper Convolutional Neural Network and Recurrent Neural Network (CNN-RNN) model is designed to identify the features. CNN-RNN model is a deep learning based approach which uses a small patch of hand written images. The feature is extracted in CNN-RNN model and trained with a generalized binary form of logistic regression classifier.**

***Index Terms*— Anaconda, Python, Convolution Neural Network, Support Vector Machine, Recurrent Neural Network, Long Shot Term Memory.**

I. INTRODUCTION

Character recognition is one of the emerging research fields as today. Most of the ancient literatures are hand written and difficult to store and access the data efficiently. Hand written character recognition software helps to store and access the data in easier way. It also provides security to data. Handwritten character recognition is one of the rapidly developing research areas in the field of deep learning. The deep learning field is highly changing due to different methods, models, algorithms and technologies that are introduced every now and then. It is important to have an understandable, faster and efficient model that is used to perform deep learning tasks like pattern recognition, speech recognition and character recognition etc. Thus development in the models of data learning is very important and one such model is using CNN-RNN hybrid model for Hand written text recognition. CNN –RNN hybrid model is one of the efficient models in Deep learning. The proposed model classifies the word in the image of the scanned word and produces equivalent text in a machine readable digital form. The digitally converted data can be presented anywhere, it can be used to read bank cheques or postal cards, it is also used in proper documentation and content management. The aim of this paper is to design software that can identify the hand written character using the concepts of deep learning. CNN-RNN model is

Grenze ID: 01.GIJET.9.1.534

selected which provide better accuracy in feature extraction. The proposed hybrid model calculates the data dependency and continuity features of the Intermediate Layer Output of the Convolutional Neural Network Model using Recurrent Neural Network and links the features of these middle layers to the final full-connection network for predicting the classification behavior.

II. LITERATURE SURVEY

Safna K M et al. presented the importance of image processing and pattern recognition in optical character recognition which can be used for office automation. Here a handwritten character recognition system for English language using Convolution Neural Network (CNN) is proposed with six layers of CNN. It uses the basic LeNet-5 architecture which has three convolution layer, two average pooling layer and a fully connected layer. CNN has better accuracy in character recognition and also the method itself extract features from the character during training phase. The final result will be in an editable format. CNN is also compared with Artificial Neural Network (ANN) to show which one has better accuracy in character recognition. The result shows that CNN have better accuracy than ANN [1].

S.Selvi et al. proposed a combination of Convolutional Neural Network (CNN) and Principal Component Analysis (PCA) with segmentation technique to improve the accuracy of hand written character recognition from text document. In this paper, the characters were classified in CNN and features are extracted using two dimensional principal component analysis method. Since the data set selected has both letters and digits, predictable results are produced by recognizing each segments of digits and letters by segmentation techniques [2].

Pranati Paidipati et al., have tried to find thefactors that emphasize on improving handwritten original copies for scanned offline better classification accuracy. The proposedCRNN hybrid network has its point of convergence on effectual training of datasets by efficient pre-processing of data, image normalization with integration of data augmentation, classification and clustering of data to get exact word with respectable dimensions of accuracy. IAM images and Devanagari datasets were used in which the subjective after-effects of the CNN-RNN hybrid architecture were quite decent. Calibrating the system layers and custom datasets has helped in improvingthe consequences of manually written content recognition regardless of 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 [3].

Sanjay Mor et al., used concepts of computer vision along with deep learning to implement hand written character recognition Here Emnist data set was used which contain numbers from 0-9, 26 alphabets in upper case and 26 alphabets in lower case. In total the dataset has 62 classes. The android application was developed using android studio and tensor flow libraries are used to link the hand written text recognition program. Text classification and detection was performed using traditional CNN approach [4].

Joseph James S, et al., used CNN classifier with XGBOOST model as prediction model. It evaluates the loss function and regularization, helps to decide particular objective function on the data set. The data set used is NIST data base 19 with 8,10,000 character images including numbers and alphabets [5].

R.Reev Ingle et al., focussed on the development of online handwritten text recognition system with the knowledge of neural networks. Proposed line recognition model achieved better accuracy and parallelism in data training compared to LSTM model [6].

Batuhan Balci et al., used CNN-LSTM approach for data classification and segmentation on IAM data set. Data training was done using ADAM Optimizer and fine tuned Tesseract model was used for data segmentation, compared the accuracy of text recognition using VGG-19, RESNET-18, RESNET-34 and character level classification techniques [7].

F.Such et al. In their work shorten the pre processing process by using fully convolutional model which takes random texts of unknown length, skips symbol alignment, correction, dictionary matching etc. They used probabilistic error rate to correct the erroneous word detection [8].

Alex Graves et al., applied Recurrent Neural Network algorithm on unsegmented data of TIMIT data base to identify the sequences of labels. The proposed technique used Connectionist temporal classification which gave better result compared to HMM and Hybrid HMM [9].

Karen et al. In their research work varied the weight of convolutional network and compared the accuracy in large scale image recognition system[10].

Lisa Yan applied Faster R-NN algorithm for objet detection and c classification in handwritten character recognition. Printed, non recursive words are identified by simple object detection approach. Performance of the system for objet detection is measured by the metric called Average Precision [11]. Curtis Wigington et al. , proposed the combination of CNN with LSTM incorporated with grid based and test side data augmentation

techiques to siginficantly reduce word error and character error rate on standard READ, RIMES and IAM datasets [12].
Tao Wang et al., suggested for accurate and speedy data representation techniques based on unsupervised feature extraction along with CNN is dominant approach to overcome the drawbacks of conventional handwritten text detection and recognition approaches which were immensely dependent on manual hand engineering based features recognition[13].
S. Hochreiter and J. Schmidhuber, presented a famous recurrent based learning approach termed as long short-term memory (LSTM) which is a gradient oriented technique to overcome backward flow of error related signals in conventional techniques. LSTM nowadays plays a significant role in various applications such as speech recognition, handwritten recognition and robotics[14].
Alex Graves et al., proposed a novel method for accurately labelling input data from speech recognition using hybrid model which works directly on sequencing of labels process. Temporal and framewise classification is done along with probability distribution technique for classification task [15].
P. Voigtlaender et al. , in their research work highlited the limitations of traditional CPU based frameworks for handwritten recognition and suggested GPU based frameworks with multidimensional network support achieves impressive recognition results. Multidimensional layers performance on standard datsets achieved better performance than single dimensional models [16].
TN Sainathet al., demonstrated a novel unified architecture for speech reconginition applications. LSTMs, CNNs combined with DNN yields a significant improvement over WER compared to earlier individual LSTMs for the same task [17].
U.-V. Marti & H. Bunke., experimented on IAM database consisting of English texts converted to standard corpus form. The corpus form is text extracted and segmented later using available image processing techniques word and line distribution is assessed for recognition accuracy [18].

III. SYSTEM DESIGN

System design involves the process of defining the architecture or flow diagram. The proposed method of hand written character recognition system involves Conolutional Neural Network with VGG-19 Classifier. The system uses Long Short Term Memory Network (LSTM) for character segmentation. Fig.1 shows the flow diagram of hand written character recognition system. The data set used is IAM hand written data set. This data set has 657 samples of hand written texts of various writers, counts more than 100K+ data for training, validation and testing. Before training the modelwith dataset, pre-processing and data augmentation are done to make it more robust. Pre-processing involves resizing of images to make all images of same size and converting coloured images to gray sale.

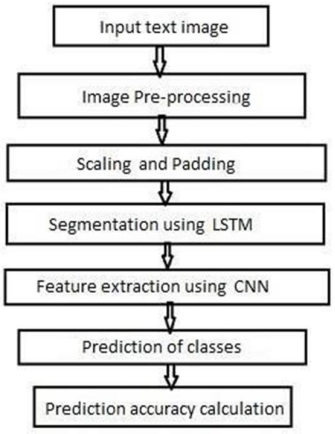

Fig 1: Flow Diagram

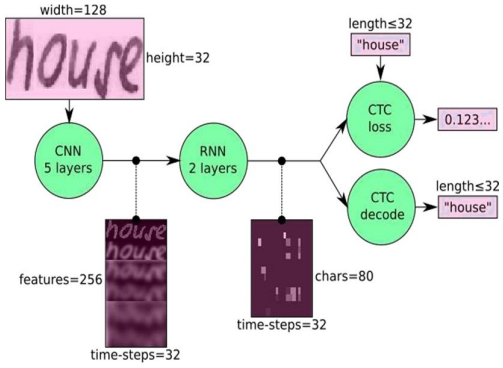

Fig.2: Overview of Hand Written Text Recognition System

Data augmentation involves de-slanting of images toincrease accuracy. Implementation of this idea requires a server on which the model will be executed. Pre-processed data are then subjected to perform segmentation and feature extraction using LSTM and CNN respectively. Final stage of implementation is prediction of classes and the prediction accuracy calculation. Model of implementation contains 5 convolution layers with each layer followed by pooling layer, followed by two LSTM layer with 256 hidden unit and lastlya decoder which will decode the output from LSTM layer.

The overview of Hand written character recognition system is shown in Fig. 2. The model has CNN with 5 layers, with each layer followed by a pooling layer, RNN with 2 LSTM layer which has 256 hidden units and a final CTC layer. The input image such as an image of hand written text is initially processed in CNN layers. CNN will train the image to extract various features. There are three basic operations performed by CNN which includes Convolution operation, RELU operation and Decimation. Convolution operation is performed in each layers of CNN. The image is filtered with filter kernel size 5×5 in first 2 layers and 3×3 in last 3 layers. Then RELU operation is performed which in non linear in nature. Finally the image is decimated in the pooling layer. Here image size is reduced by two in each layer. The size of output from CNN is 32×1×256 which indicates 256 features are extracted per time-step. There are 256 channels and each channel represents distinct features. The output of CNN block is fed too RNN block which uses two layers of STM to train the data. These two layers of LSTM have 256 hidden words which help for predicting the correct word from the sequence of data. The output of RNN is in the form of matrix. The last block of implantation involves CTC word bean search decoder with loss calculation feature. It has a matrix of ground truth text. The output matrix from RNN is compared with ground truth text. That is the predicted words are compared with original words and loss function is calculated accordingly. The advantage of using CTC word bean search decoder is this will limit the guessing of words only to the predefined set of words.

IV. IMPLEMENTATION

The series of steps involved in the Implementation of Hand Written Text recognition system with its functionality is represented in a flow diagram as shown in Fig.3. The data acquisition involves collection of IAM hand written data set. Selected text from the data set is then preprocessed to maintain same size and converted it to gray scale images. The text is then de-slanted and subjected to feature extraction and segmentation by NN and LSTM respectively. Finally selected text is recognized and accuracy is calculated.

After training the model to an acceptable accuracy then move to the testing phase. In this phase an android application is built to capture an image of handwritten text written on a sheet of paper, and then this image is sent to the model using Python flask connection. The model recognizes the word from the imageand returns the result. The Android app will receive the result and prints the recognized word on the app screen. The Android app has two layouts. The firstlayout has one image view and 3 buttons for the camera, crop and save the image. In the next layout has 5 views- Two buttons one for selecting the image other for establishing the connection to send the image to the model, then there are 3 text fields for receiving IP address and port number from the user and last field for displaying message for connectionestablishment and final result. Android app uploads an image to the model using Flask (server-based application in Python) from the Android storage (Phone Gallery). OkHttp library in android will help to handle the HTTP requests and responses. The URL is built for serer side flask by accessing the data entered by the user in text field provided. These data involves IP address and port number. On clicking the Connect Server Button a request is built consisting two kinds of information. First the URL and Second the HTTP message to be delivered which will be inthe form of an image. This selected image is read as a Bitmap which will be converted into a byte array. The HTTP response is also accessed using two callback methods: Firstwhen the result is successfully returned from the model, next when result couldn't be executed due to network error, cancellation or timer run out. HTTP server is built on a simple framework called flask, created as an instance of flask class. The constructor of this call acceptsa single argument representing the name ofthe app's module [8]. The route function associates the URL which refers to the location of the model where the image is sent. The function supports both GET and POST type of messages. Model will recognize the word which will be sent as a return responseto Android app.

V. RESULT AND DISCUSSIONS

The proposed hybrid model of Hand Written Character Recognition was tested for various sized data set and is discussed in this section. Fig.5. shows the output accuracy obtained by simple CNN and RNN model by training on 25 thousand images having batch size of 50 images. The accuracy obtained was 55.24%.

Accuracy was increased to 10% when the size of data set increased to 1lakh with bat size of 150 images which is shown in Fig. 6. In the above trial the image was not de-slanted.

Fig. 7 (a) and 7 (b) shows the slanted and de-slanted version of image text. The de-slanted data sets are then trained on the same CNN-RNN model and the accuracy obtained was 69.77% as shown in Fig 7(c). The size of de-slanted data sets is doubled and accuracy was calculated. The obtained accuracy is 75.24 % as shown in Fig 8.

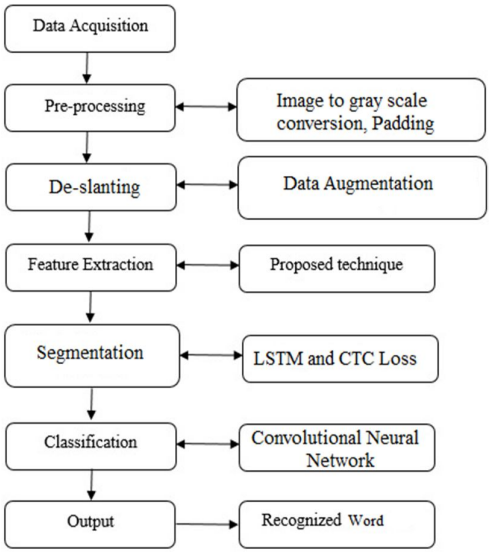

Fig.3. Flow Diagram of Hand Written Text Recognition System

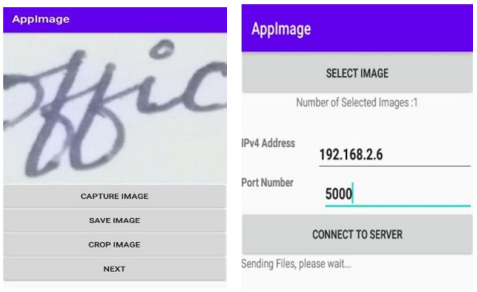

Fig.4. Android App Layout

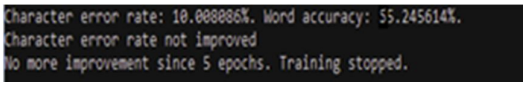

Fig.5: Accuracy obtained on 25 thousand word images

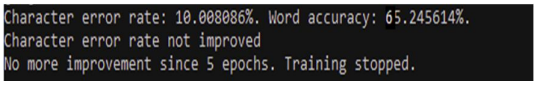

Fig.6. Accuracy obtained on 1 lakh word images

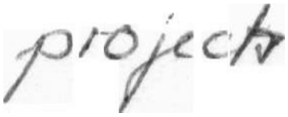

Fig. 7(a)

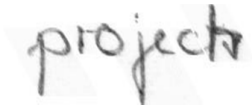

Fig. 7(b)

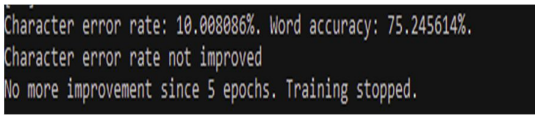

Fig. 8: Accuracy obtained on doubled dataset

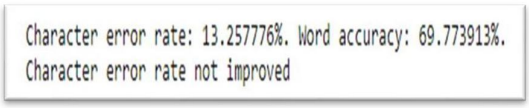

Fig. 7(c)

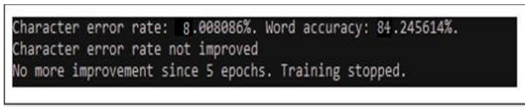

Fig. 9. Final Accuracy obtained during training

Fig.7.(a) Slanted image (b) De-slanted image (c) Accuracy obtained for de-slanted image

Then the “Beam Search” decoder isreplaced by “Word Beam search” decoder which results in a hike in accuracy of about 9% from 75% to84%. From the above images it has been observed that themodel is able is correctly classify the images with a 84.2% accuracy with a character error rate of only 8%. Input word-image is then fed to the model to perform the classification. The model classifies the word correctly and displays the recognized word in digital format and also display the CTC loss value along with the digital format of the word. Fig 10(a) and 10(b) shows two such examples for correct hand written text recognition by CNN-RNN approach.

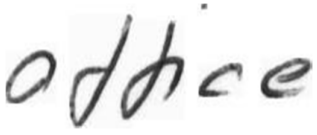

Fig.10.(a) Input image

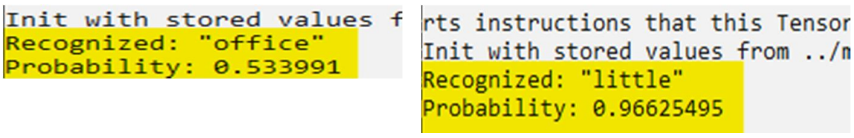

Fig.10 (b) Recognized output word

VI. CONCLUSION

Handwriting recognition is the ability of a computer to receive and interpret intelligiblehandwritten input from sources such as paper documents, pictures, etc. This stands as one of the pillars of the digital platform to recognize characters without human effort. An ultimate of handwriting recognition is to simulate the human reading capabilities so that the computer can read, understand, edit and work as human do with the text. It contributes immensely to the advancement of automation process and improves the interface between man and machine in numerous applications. Handwritten data converted into digital text can be stored electronically and can bepresented anywhere. The purpose of this project is to legitimize and implement the usage of Convolution neural networks (CNN) in parallel with long short term memory (LSTM) model to improve handwriting Recognition systems. The learning modelproposed in this project is based on CNN as a feature extraction and classification tool and LSTM as an segmentation tool. Finally, as seen in our project we integrated our model with a mobile application for betterunderstandability and to improve ease of use of the project. The output of our project is a digital representation of a handwritten word. This project can help reduce manual labour needed in maintaining handwritten documents since data can be stored and accessed easily on computers. It can also ensure security of data and makes sure no data is misplaced easily. It can also allow data to be presented anywhere.

FUTURE WORK

The system that has currently developed has multiple features that benefit the user while interacting with the application. Further one can focus more on making the system more user-friendly and appealing to the users of all ages. In future overall performance of the system can be improved. This can be done in a various number of ways. Starting with pre-processing, accuracy of the model can be improved if we implement different pre-processing techniques. The samecan be done with data augmentation. Several research papers have shown that the larger the size of the dataset can achieve better accuracy. Increasing the accuracy of themodel can reduce in loss and hence word can be recognized more efficiently. Once this model works efficiently for word recognition wecan move on from word recognition to sentence recognition.

REFERENCES

[1] Safina K M, “English Handwritten Character Recognition using Convolution Neural Network (CNN)”, International Journal for scientifics research and development, ol.6, issue 2, pp. 3391-3398, 2018.
[2] Safina K M, “English Handwritten Character Recognition using Convolution Neural Network (CNN)”, International Journal for scientifics research and development, ol.6, issue 2, pp. 3391-3398, 2018.
[3] Pranati Paidipati, Dr. Sachin Choudhari, Ashish Kumbhare “Enhancing the Recognition of Handwritten Scripts Using CNN-RNN Hybrid Networks.”, IOSR Journal of Engineering (IOSRJEN), Vo.9, issue 5, pp. 27-30,2019.
[4] Shubam Sanjay Mor, Shivam Solanki, Saransh Gupta, Sayam Dhingra, Monika Jain, Rahul Saxena. ,“Handwriting Text Recognition withDeep Learning and Android” , International Journal of Engineering and Advanced Technology, vol.8, issue 3, 2019.
[5] Joseph JamesS ,C.Lakshmi, UdayKiran P, Parthiban “An Efficient Offline Hand Written Character Recognition using CNN andXgboost”, International Journal of Innovative Technology and Exploring Engineering , vol.8, issue 6, 2019.
[6] R. R. Ingle, Y. Fujii, T. Deselaers, J. Baccash and A. C. Popat, "A Scalable Handwritten Text Recognition System," 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 17-24, 2019.
[7] Batuhan Balci, Dan Saadati, Dan Shiferaw, “Handwritten Text Recognition using Deep Learning” , IJRAT, 2019.
[8] Felipe Petroski Such, Dheeraj Peri, Frank Brockler, Paul Hutkowski, RaymondPtucha, “Fully Convoluted Network for HandWriiten Recognition.”, International Conference on Frontiers in Handwriting Recognition, 2019.
[9] Alex Graves, Santiago Fernndez, Faustino Gomez, Jrgen Schmidhuber, “Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks”, Proceedings of the 23rd international conference on Machine learning. 2006.
[10] K. Simonyan, A. Zisserman “ Very Deep Convolutional Networks Large-Scale Image Recognition”, arXiv technical report, 2014
[11] LisaYan.”Recognizing Handwritten Characters”. CS 231N Final Research Paper, 2019.

[12] Curtis Wigington, Seth Stewart, Brian Davis, and Bill Barrett, ”Data Augmentation for Recognition of Handwritten Words and Lines using a CNNLSTM Network”,14th IAPR International Conference on Document Analysis and Recognition, 2017.
[13] T. Wang, D. Wu, A. Coates, A. Ng. ”End-to-End Text Recognition with Convolutional Neural Networks” ICPR2012.
[14] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” NeuralComputation, vol. 9, no. 8.
[15] A. Graves, S. Fern´andez, F. Gomez, and J. Schmidhuber, “Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks,” in ICML, 2006.
[16] P. Voigtlaender, P. Doetsch, and H. Ney, “Handwriting recognition with large multidimensional long short- term memory recurrent neural networks,” in ICFHR, 2016.
[17] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, “Convolutional, long short-term memory, fully connected deep neural networks,” in ICASSP, 2015
[18] U. Marti and H. Bunke, “The IAM- database: An english sentence database for offline handwriting recognition,” IJDAR, vol. 5, pp. 39– 46, 2002.