International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Telugu text extraction and recognition using convolutional and recurrent neural networks

A. Ram Bharadwaj, A. Venugopal, Ch. Surya Kiran, M. V. Nageswara Rao

Abstract*: Recognizing words form images is the most atomic aspect of an OCR system. Inspired by the recent success of deep-learning based techniques in computer vision & sequence prediction, an end-to-end trainable CNN-RNN based neural network model for recognizing Telugu words from images is presented. This model can also be extended to other languages.

Index Terms: ***Recurrent Neural Networks -RNN, Optical Character Recognition -OCR, Convolutional Neural Networks -CNN***

I. INTRODUCTION

OCR is used to convert typewritten documents into ASCII code. This model enables us to feed images involving various styles and fonts like old literatures, archives, pamphlets and so on. It is common technique used for text to speech conversion, machine translation, key data, cognitive computing and Text Mining. M. S. Rajasree has proposed character recognition on Malayalam that uses multi-layer perceptron and Radial-Basis function. Here, image is preprocessed to a standard size, noise is removed using gaussian noise filter and then individual characters are extracted. It is trained on a total of 715 images whose results are 90.5% accuracy in Gaussian noisy environment [1]. Abhijit dutta & santanu chaudhury suggested character recognition on Bengali. It uses curvature related feature for characterizing the strokes which constitute the characters. The drawback is, it is incapable to perceive composite characters. The achieved accuracy is 85% [2].

Nidhi Kalidas Sawant and Prof. Sangam Borkar discussed the Hindi script for character recognition using histogram of oriented gradients (HOG) feature. The extracted features are given to multiclass Support Vector Machine (SVM) classifier which are then mapped to the characters in English [3]. Baoguang Shi, Xiang Bai and Cong Yao proposed work on image to text recognition using CRNN architecture [5]. This paper presents the extraction of Telugu text from images.

II. OPTICAL CHARACTER RECOGNITION

OCR converts raw images containing text into text data. Tesseract is an OCR engine developed by Hewlett Packard which does the task of image to text conversion and it has been developed for a total of 116 languages. This process can be divided into three major-subtasks. They are recognizing characters, predicting words based on the recognized sequence of characters, predicting sentences from the predicted sequence of words as shown in **Error! Reference source not found.**.

It is often enough to build a word-level OCR-extractor after which the rest of the system can be built. Most OCR-systems differ at this level.

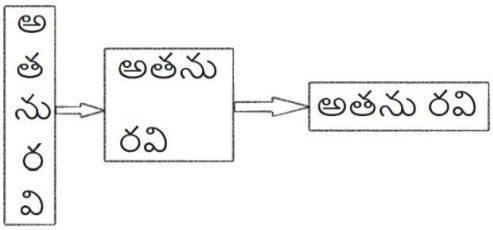

Figure 1. Sentence formation from individual characters.

Recognizing Telugu words, Telugu is an Indian language which is spoken by about 75 million people around the world. Despite, this there is little work in Telugu-language OCR systems largely due to the non-availability of datasets. We present a method to generate a word-image dataset given a large corpus of words in the language, this technique can be used to for other languages as well with little effort.

CRNN is a combination of CNN whose output-vector from the last layer is fed to the input of a bidirectional-recurrent-neural-network. The CNN part recognizes features from the images and transform them into a vector which the RNN uses for sequence prediction. The architecture used is shown in Figure 2.

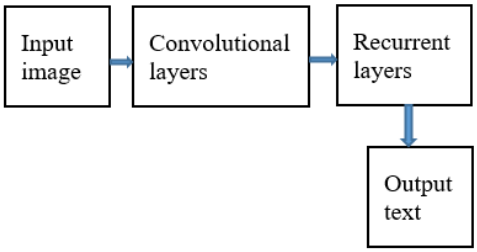

Figure 2. Architecture of CRN

III. CRNN ARCHITECTURE

The architecture of this model is a neural-network-stack which consists of 3 basic blocks. They are CNN block, RNN block, CTC transcription block [5].

Manuscript published on 30 June 2019.
* Correspondence Author (s)
A. Ram Bhardwaj, Dept. of ECE, GMRIT, Rajam, India.
A. Venugopal, Dept. of ECE, GMRIT, Rajam, India.
Ch. Surya Kiran, Dept. of ECE, GMRIT, Rajam, India.
Dr. M. V. Nageswara Rao, Dept. of ECE, GMRIT, Rajam, India.

The CNN block is a typical sequential model convolutional neural network [6] whose last layer produces a flattened hidden-layer vector which is fed to the input of RNN-stack, the layer-wise details are given in the next section. The architecture of CNN is shown in ***Figure 1***.

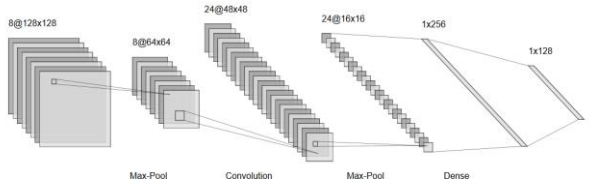

Figure 1. Architecture of CNN

Recurrent Neural Networks are predominantly used in sequential modelling tasks as in ***Figure 2***. However learning long term dependencies through RNNs is very hard due to vanishing and exploding of the gradient flow, this problem is tackled by LSTM [7, 8]. The RNN block is a stack of two bidirectional LSTM units, bidirectionality allows sequence prediction to catch dependencies from both the sides the output vector is modelled in form of a one-hot character-class level encoding with each time-step, which is compared to the actual value and a single scalar value of loss is obtained, the decoding of this matrix to word form is done by beam-searching through character-level probabilities for nearest words, more on this is described in [4].

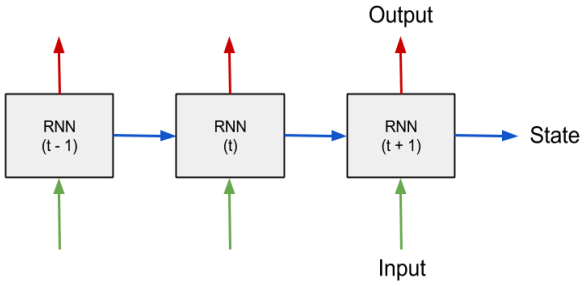

Figure 2. Architecture of RNN

The Neural network training will be accompanied by the CTC matrix [4]. The output from Recurrent Neural Network layer is a matrix which is to be trained initially and then decode the matrix. After training the Neural Network, it is used to recognize text in unseen images. This is done by calculating the best path by taking the most likely character per time-step and removing the redundancies as in **Figure 3**.

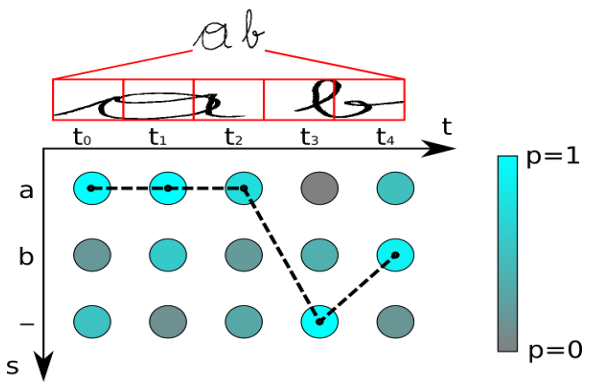

Figure 3. Output matrix of NN. The thick dashed line represents the most probable output.

IV. IMPLEMENTAION

The CNN block is a stack of 5 Convolutional layers each followed by Relu activations and 2x2 Maxpooling at the end the obtained vector is flattened to a 512x1x16 dimensional vector which is fed to a stack of two bidirectional LSTM units whose output vector is reshaped to resemble the dimensions of the CTC matrix of the output and Softmax function is applied to obtain the character-level class probabilities at each time step. This CTC matrix is decoded by running beam search following the path of classes with maximum probabilities. All the model building is done in python using the pytorch module, the training process is done using Adam optimizer with a learning rate of 0.0001 for 40 epochs.

V. RESULTS

The proposed model is trained on data set of 11000 generated printed text-words, the word-distance of wrongly classified words is not included since character-error rates are misleading. Out of randomly chosen 100 words from the validation set the developed model is able to predict 81 words correctly. Thus achieving 81%-word prediction rate where tesseract-OCR-system gets 57%. When compared with tesseract OCR system the developed model outperforms it on our dataset. However, this model fails to recognize words in varied backgrounds and other font styles. Few of these results are shown in Table 1.

Table 1.

ORIGINAL IMAGE	PREDICTED WORD
	కండరాల
	గంపెపాశతో
	గళన
	గడిచేవారు
	గుణాత్మక
	సందర్భాలు
	సందర్భోచిత

VI. CONCLUSION

Developed the model for Telugu text extraction and recognition using convolutional and recurrent neural networks .The model developed is confined to developed dataset. By using additional LSTM networks this model can be extended to predict at sentence level. It can also be extended to other languages by training on their datasets. Based on the results obtained this model is able to predict 81 words of randomly chosen 100 words from the validation set.

REFERENCES

1. M Abdul Rahiman and M S Rajasree. Printed Malayalam Character Recognition Using Back-propagation Neural Networks, IEEE, IACC 2009.
2. Abhijit dutta and santanu chaudhury. Bengali alpha-numeric character recognition using curvature features, Pattern Recognition, 26(12): pp. 1757-1770, 1993.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

3. Nidhi Kalidas Sawant and Prof. Sangam Borkar. Devanagari Printed Text to Speech Conversion using OCR.IEEE ISBN:978-1-5386-1442-6.
4. Alex Graves, Santiago Fernandez, Faustino Gomez and Jurgen Schmidhuber. Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks International Conference on Machine Learning, 2009.
5. Baoguang Shi, Xiang Bai and Cong Yao. An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition.arxiv.org/abs/1507.05717
6. Yann LeCun, Patrick Haffner, Leon Bottou and Yoshua Bengio. Object recognition with Gradient-Based Learnin. IEEE, 2278 – 2324, 1998.
7. Y. Bengio, P. Y. Simard, and P. Frasconi. Learning longterm dependencies with gradient descent is difficult. NN, 5(2):157–166, 1999
8. Sepp Hochreiter and Jurgen Schmidhuber. Long Short Term Memory, Neural Computation :1735-1780,1997

AUTHORS PROFILE

A. Ram Bhardwaj, B.Tech, Dept. of ECE, GMRIT, Rajam, AP. Interested in theoretical computer science and Computational Complexity.

A. Venugopal, B.Tech, Dept. of ECE, GMRIT, Rajam, AP. Interested in Mixed Signal VLSI design

Ch. Surya Kiran B.Tech, Dept. of ECE, GMRIT, Rajam, AP. Interested in signal processing.

Dr. M. V. Nageswara Rao, received Ph.D from Andhra University 2013. Presently, he is professor, Department of ECE, GMRIT, Rajam, A.P; India. His research interests are VLSI and signal processing.