Devnagri Handwritten Text recognition using Maximally Stable Extremal Regions algorithm and Cascaded convolutional neural network

Ramanan.B and Nissy Niharika

Indian Institute of Space Science and Technology
Thiruvananthapuram

***Abstract*—This project deals with the recognition of sentences by breaking the sentences into words and then breaking down the words into letters which is done through various methods and alogorithms such as Maximum stable extended region algorithm (MSER) ,similarity index and uses cascaded neural network in recognizing the handwritten digits.**

I. Introduction

In this project it is examined on the hindi literature,So that the words which are splitted into letters are again classified into the letters which have matras and which doesn't have matras. The letters which include matras are then classified into top, bottom and vertical matras.This is done by ResNet architecture trained from scratch,Google Colaboration,GPO,10 fold cross validation by using 120 epochs from training of CNN Convolution Neural Networks.

The result which is obtained has to avoid overfitting and underfitting as well, where this can be done by the help of data augumentation,Image data generator where these two come under Keras neural network package.The input which is to be classified into the top,vertical and bottom which is done by 2 convolution layers ,2 dense layers and 1 output layer where the output layer is of 10 fold CV (Cross Validation) and 15 epochs.

II. Maximally Stable Extremal Regions

Maximally Stable Extremal Regions (MSER) is a feature detector; Like the SIFT detector, the MSER algorithm extracts from an image I a number of co-variant regions, called MSERs. An MSER is a stable connected component of some level sets of the image I. Optionally, elliptical frames are attached to the MSERs by fitting ellipses to the regions. For a more in-depth explanation of the MSER detector, see our API reference for MSER.

A. Extracting MSERs

Each MSERs can be identified uniquely by (at least) one of its pixels x, as the connected component of the level set at level I(x) which contains x. Such a pixel is called seed of the region.

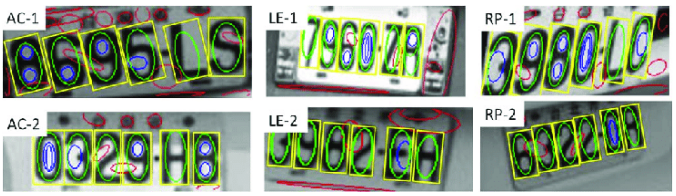

Fig. 1: Mser Detection

A stable region has a small variation. The algorithm finds regions which are "maximally stable", meaning that they have a lower variation than the regions one level below or above. Note that due to the discrete nature of the image, the region below or above may be coincident with the actual region, in which case the region is still deemed maximal.

III. CNN

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

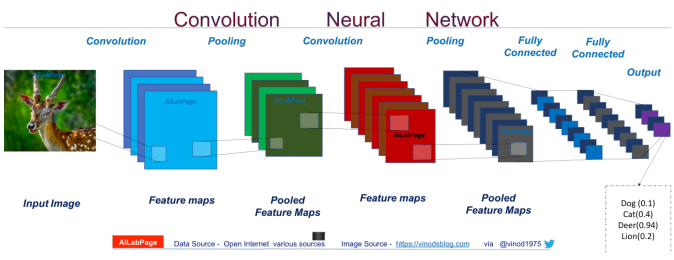

Fig. 2: Convolutional neural network structure

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

IV. Resnet Architecture

ResNet, short for Residual Networks is a classic neural network used as a backbone for many computer vision tasks.The fundamental breakthrough with ResNet was it allowed us to train extremely deep neural networks with 150+layers successfully. Prior to ResNet training very deep neural networks was difficult due to the problem of vanishing gradients.

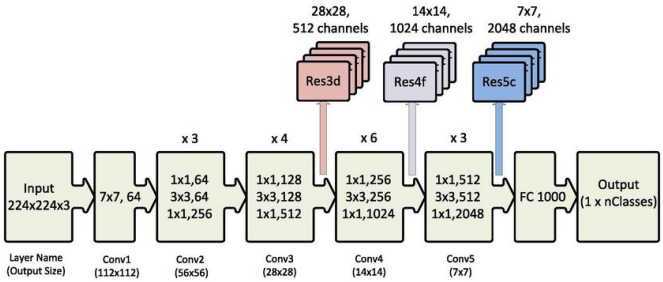

Fig. 3: Resnet-Architecture

We have used Resnet Architecture for recognizing the devnagri handwritten text(without matras) with the help of CALAM dataset

V. Data

Handwritten Vowels and Consonants (with Modifiers) Devanagari database developed at the Department of Computer Science and Engineering of the Malaviya National Institute of Technology as part of research project grant . A database for off-line Hindi handwritten character with matras (modifiers) is developed. Data set is collected from persons of different age, gender, profession and educational qualification. The character images are stored as images in PNG image format for efficient use. The data set consists of more than 23000 images of their original size with pro grammatically segmented consonant, Numerals and Vowels. Data are also collected from person from different geographical locations of India.

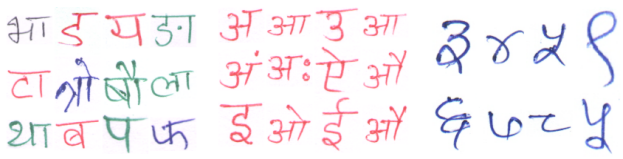

Fig. 4: Sample Images from Calam Dataset

VI. Text Recognition Model

The text area is detected using image thresholding.Then we separate out words from sentences using MSER algorithm.MSER helps us in separating words irrespective of being slant,skewed etc..After separating into words we follow heuristic way of cropping of images that whether a given a threshold width in order to check whether it is a combination of more than one word,if it is greater than the threshold width use a sliding window to crop out words else use the mser detected region as word for our model.

Here are sample images showing regions detected by mser algorithm Detected images are then resized into 64x64x3

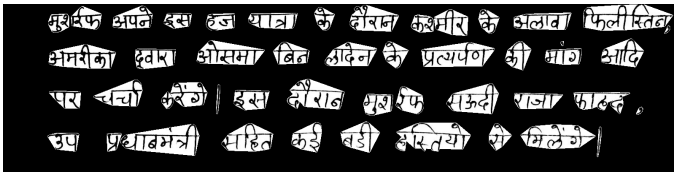

Fig. 5: Seperating text from White spaces using image thresholding

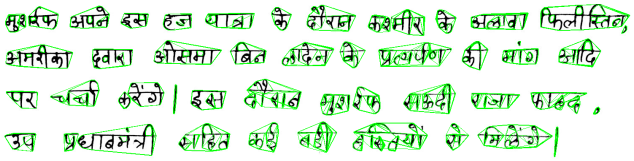

Fig. 6: Detecting words using MSER algorithm

images.Then we take three potential regions where matras can exist.The first one is the top regions where matras like 'ye','ai' etc can exist.First we take 15 rows from as top and re-size it into a new 64x64 image(Top Image)

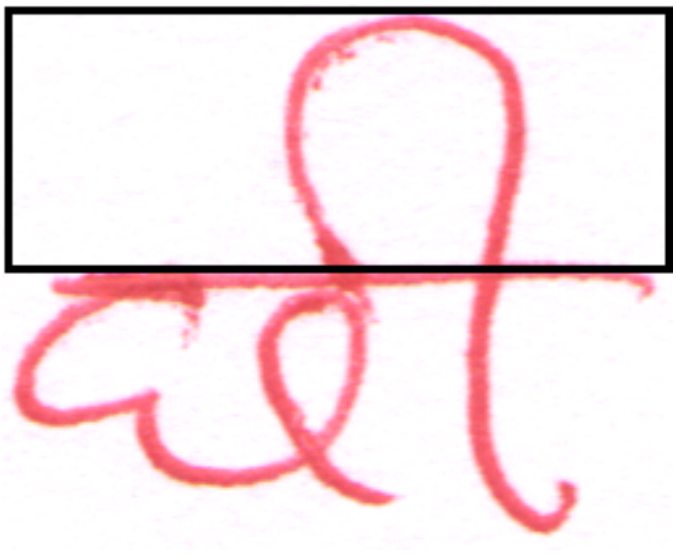

Fig. 7: Cropping top image(Not to scale)

Next we take bottom 25 rows and re-size it into a new 64x64

image(Bottom image)

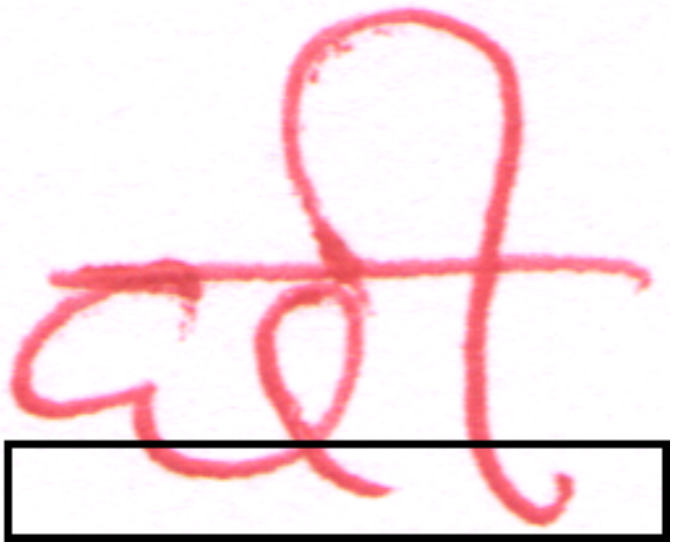

Fig. 8: Cropping Bottom image(Not to scale)

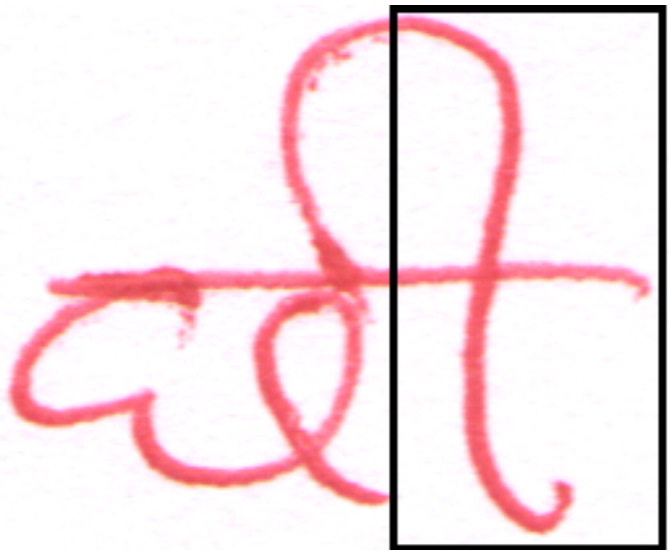

Fig. 9: Cropping vertical image(Not to scale)

we take leftmost 45 columns and re-size it into a new 64x64 image(Vertical Image)

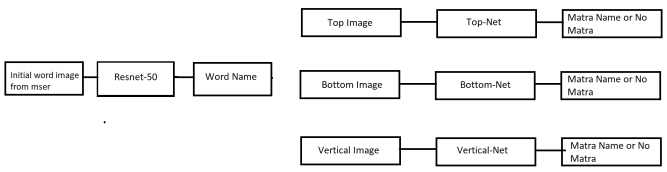

Fig. 10: Model structure

We have our cascaded neural network model as shown above.

We use Resnet convolutional neural network to detect the word and all the top net,bottom net,vertical net follow same architecture shown but with different number of output nodes depending on number of matras present in top,bottom,vertical parts of the word.We train Resnet neural network using CALAM dataset with 10 fold cross validation and the other three using the cropped matras images from words of CALAM dataset with 10 fold cross validation.Data augmentation has been used in all the process in order to prevent overfitting. We use a relatively simple architecture for our neural nets processing modifiers.Architecture of top net,bottom net and vertical net can be seen at Fig.11 in the last page

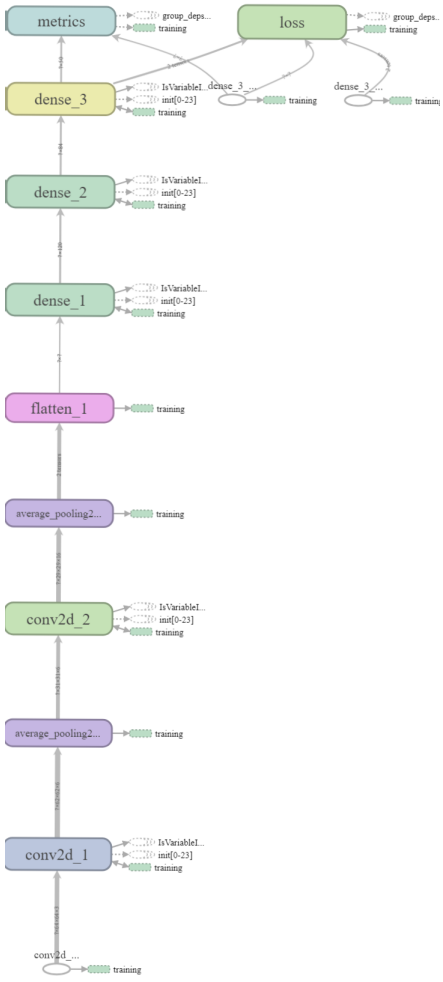

Fig. 11: Neural network architecture

VII. Observations

We compare our Resnet model performance with that of Lenet. Then we have following accuracies for our neural nets

.

VIII. Conclusion

Thus in this project we were able to detect words using the above mentioned model but still we face problems of aligning

TABLE I: Test Accuracies of neural network acheived after training

Neural Net	Accuracy
Resnet	0.86
Lenet	0.20

Neural Net	Accuracy
Top net	0.80
Bottom	0.80
Vertical	0.77

the words in the right sequence as per the sentence and lack of training data for top-net,bottom-net and vertical net neural network as calam dataset provides only words and does not provide data with respect to matras.Thus performance of the above model can be improved by having more data for top-net,bottom-net and vetical net that is the images of the matras.

ACKNOWLEDGMENT

I would like to thank Dr.Deepak Mishra Associate Professor,Department of Avionics,Indian Institute of Space Science and Technology for his guidance on this project.

REFERENCES

[1] Deep Convolutional Neural Network for Image Deconvolution-Li Xu Lenovo Research Technology Jimmy SJ. Ren Lenovo Research Technology Ce Liu Microsoft Research celiu@microsoft.com Jiaya Jia The Chinese University of Hong Kong

[2] A Cascaded Convolutional Neural Network for Single Image Dehazing-Chongyi Li, Jichang Guo, Fatih Porikli, Huazhu Fu, Yanwei Pang

[3] Handwritten Devanagari Script Recognition: A Survey -Aradhana A Malanker ,Prof. Mitul M Patel

[4] LeNet-5 in 9 lines of code using Keras-Medium.co.in