International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-06, Issue-04, July 2020

Deep Learning Based Malayalam Handwritten Character Recognition System

*Athira Sekhar A. S, #Dr. Subu Surendran

*PG Scholar, #Professor & Head of the department (computer science and engineering), SCT College of Engineering, Thiruvananthapuram, India, athirasekharas@gmail.com, subusurendran@gmail.com

Abstract - **Handwritten character recognition (HCR) is the ability of a computer system to recognize the handwritten character inputs. It has been one of the active and challenging research areas in the field of image processing and pattern recognition. It has numerous applications which include, reading aid for blind, bank cheques and conversion of any hand written document into structural text form. HCR is a challenging problem due to huge variation in individual writing styles. Many techniques have been used such as Hidden Markov Model (HMM), SVM (Support Vector Machine), fuzzy membership values etc. for the handwritten character recognition are used in different languages. Another novel technique for handwritten character recognition is by the use of CNN (Convolution Neural Network) which is more accurate than the other methods used so far.**

This paper describes the recognition of Malayalam characters deep learning approach, to execute the task of HCR a convolutional neural network(CNN) model ResNet is implemented. The dataset consist of 105 Malayalam characters (including old script) of 128x128 pixel image format. The characters are mapped using the Unicode values. The threshold value from each class is recorded and is mapped with the Unicode values the character with higher threshold is predicted, also the prediction of characters is compare using Mobile net and Inception models. A graphical user interface is created for user input.

Keywords — **Convolutional Neural Network, Character recognition, Resnet, Inception Network.**

I. Introduction

Handwritten character recognition is an area of pattern recognition which defines an ability of a machine to analyze patterns and identify the character. Pattern recognition make inferences from perceptual data based on either a learned knowledge or on statistical information [1]. The subject of pattern recognition spans a number of scientific disciplines uniting them in the search for the solution to the common problem of recognizing the members of a class in a set containing elements from many patterns in classes. A pattern class is a category determined by some given common attributes. A pattern is the description of any member of a category representing a pattern class. The basic function of pattern recognition system is to detect and extract common features from the patterns describing the objects that belong to the same pattern classes and to recognize the pattern and classify it as a member of the pattern class under consideration.

Deep learning Techniques have attained higher performance in pattern recognition tasks. These include image recognition, human face recognition, human pose estimation and character recognition. These deep learning techniques give better performance than traditional methods for pattern recognition. Deep learning enables automation of feature extraction task. Traditional methods involve feature extraction which is to be done manually. This task of crafting features is time consuming and not very efficient. The features ultimately determine the effectiveness of the system. Deep learning methods outshine traditional methods by automatic feature extraction. Convolutional Neural Networks (CNN) is a popular deep learning method and is state of the art for image recognition. CNN has achieved a breakthrough in the IMAGENET challenge 2011.

The task of handwritten character recognition is difficult to implement since the characters usually has various appearances according to different writer, writing style and noise. Researchers have been trying to increase the accuracy rate by designing better features, using different classifiers and combination of different classifiers. These attempts however are limited when compared to CNN. CNNs can give better accuracy rates but it has some problems that needs to be addressed.

 DOI : 10.35291/2454-9150.2020.0515

ISSN : 2454-9150 Vol-06, Issue-04, July 2020

Malayalam is one among the twenty two scheduled languages in India and is the official language in the state of Kerala [1]. Malayalam characters are complex due to their curved nature and there are characters which are formed by the combination of two characters. These along with the presence of 'chillu' make recognizing Malayalam characters a challenging task. Many works have been done for recognizing characters of different languages Devanagari, Kannada, Tamil, English etc using Hidden Markov Model, fuzzy membership values, Artificial neural networks, machine learning models. Among all these languages Malayalam is a difficult task due it enormous data set also deep learning techniques could provide much more accuracy and precision compared to other machine learning techniques, since learning rate is more for deep learning models.

Convolutional Neural Networks (CNN) is a popular deep learning method and is state of the art for image recognition. CNN has achieved a breakthrough in the IMAGENET challenge 2011. CNN is very suitable to deal with image structures. The properties of CNN that makes this possible are the local connectivity strategy and the weight sharing strategy. Here we attempt to use CNN model ResNet to achieve better accuracy rate in Malayalam handwritten character recognition. Also other two models Inception and mobile net is also used for a comparison.

The paper is organized as follows, Section I contains the introduction of Character recognition, Section II explains the related works, Section III problem statement, Section IV describes about datasets, Section V explain about CNN architecture, Section VI describes proposed method, Section VII include its results and discussion. Finally, the models used are summarized and concluded in section VIII.

II. RELATED WORKS

The technique by which a computer system can recognize characters and other symbols written by hand in natural handwriting is called handwriting recognition system. Handwritten character processing systems are domain and application specific, like it is not possible to design a generic system which can process all kinds of handwritten scripts and language. Lots of work has been done on European languages and Arabic (Urdu) language whereas domestic languages like Hindi, Punjabi, Bangla, Tamil, and Gujarati etc. are very less explored due to limited usage. Offline handwritten character recognition has been implemented in earlier days onwards, but modern techniques able to learn and recognize hand written character to a greater extent, since they focus on machine learning techniques which are able to learn visual features, avoiding the limiting feature engineering previously used in traditional techniques. Here we review various methods used in the field of handwritten character recognition.

Salvador España-Boquera et al [5], in this paper hybrid Hidden Markov Model (HMM) model is proposed for recognizing unconstrained offline handwritten texts. In this, the model has been constructed with Markov chains, and a Multilayer Perceptron is used to calculate the emission probabilities. Here different techniques are used to remove slopes and slants from handwritten text and to normalize the size of text images with supervised learning methods. The key features of this recognition system were to develop a system having high accuracy in preprocessing and recognition, which are both based on ANNs.

T. Som [10] have discussed fuzzy membership function based approach for HCR. Here images of characters are normalized to 20x10 pixels. From each characters around 10 images are fused to form average image. Around characters bounding boxes are constructed by using vertical and horizontal projection of character. After cropping image to bounding box, it is resized to 10x10 pixels size. After that, thinning is performed and thinned image is placed in one by one raw of 100x100 canvas. Similarity score of test image is matched with fusion image and characters are classified.

Yu Weng, Chunlei Xia et al. [4] proposed a method in which an image processing module is design to collect the data, then process the data, create a dataset. Finally a network structure is created for optical character recognition of the data set.

Shailesh Acharya et al.[8] proposed a Deep Learning Based Large Scale Handwritten Devanagari Character Recognition that used convolutional neural network for classifying Devanagari handwritten characters. Apart from the convolution layer a dropout layer is added to avoid the over fitting of the model. They implement two models of the network, model A consisted of three convolution layers and one fully connected layer and model B was a shallow network. The highest testing accuracy for Model A was 0.98471 and for model B was 0.982681.

III. PROBLEM STATEMENT

The work can be taken into consideration for analyzing handwritten character in different users so that it can be later, after further enhancement used in document verification in post offices, banks etc.

The problem statement is to recognize the individual Malayalam letters as well as sentences from a scanned handwritten document using deep neural network.

IV. DATASETS

Handwritten character recognition has a goal of recognizing handwritten input from different users. It is a challenging task due to the varying writing style of the people, Malayalam character recognition becomes more challenging

 DOI : 10.35291/2454-9150.2020.0515

ISSN : 2454-9150 Vol-06, Issue-04, July 2020

due to the enormous character dataset. As compared to the traditional methods new deep learning methods become a better platform for recognition tasks, they required huge data set for extracting features from the data of their own. The dataset used for the task is image dataset which consist for 105 different Malayalam character with 500 samples each of 128x128 size. Among them 48 character images were downloaded from the github site and remaining characters are inserted into the dataset by collecting from different users.

Fig 2.1 a

Fig 2.1 b

Fig 2.1 a, b shows the characters in the dataset

V. Convolutional Neural Network

A convolution neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imagery. They are regularized versions of multi-layer perceptron's. Multi-layer perceptron usually refer to fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "fully-connectedness" of these networks make them prone to over-fitting data. It is a subclass of neural-networks which have at least one convolution layer. CNN is a type of deep learning model for processing data that has a grid pattern, such as images, which is inspired by the organization of animal visual cortex and designed to automatically and adaptively learn spatial hierarchies of features, from low- to high-level patterns. A convolutional neural network consists of a number of layers such as the input layer, the output layer, and multiple hidden layers. The hidden layers of a CNN usually include a sequence of convolutional layers that extract the feature from input contents. It has a RELU layer for applying an activation function and followed by additional layers like pooling layers, fully connected layers and normalization layers, these layers are known as hidden layers in CNN because their inputs and output values are covered by the activation function and last convolution. Convolution and pooling layers, perform feature extraction, whereas the third, a fully connected layer, maps the extracted features into final output, such as classification. Lenet, Alexnet, ResNet, VGG, mobile net, Inception are some of the CNN networks. Among them here we use Resnet model and compare using mobile net and Inception networks.

VI. Proposed Method

A character recognition system is constructed based on the convolutional neural network. Initially a mapping file which include the character and their corresponding Unicode values are generated for mapping the characters with their Unicode values during the prediction phase. Once the dataset creation and preprocessing is completed data's are fed to the model for training, as a result a model file is generated. When an input is given for prediction the probability values are generated in the output layer corresponding to each node (represent the characters), index value of the node with higher probability value are extracted, with the help of the mapping file and the index value we get the Unicode of the corresponding character.

The proposed system consists of 4 different modules namely Data set Creation and Augmentation, Defining CNN architecture, Training the CNN Model, Deploy the Model, the below Fig.3.2 shows the detailed system architecture. The data set creation module consists of segmented Malayalam handwritten character samples that are used for training the model. The data samples are collected from different individuals and then collected samples are augmented. Then the augmented dataset is separated into training and testing set. The data get pre-processed at the pre-processing module in which quality of data set is improved by removing noise, elimination of unwanted spaces, etc. After that these data's are fed to the convolutional neural network for training, finally the model gets created. From a Graphical user interface (GUI) the word inputted from the user after subjected to segmentation is fed to the trained model for classification.

DOI : 10.35291/2454-9150.2020.0515

International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-06, Issue-04, July 2020

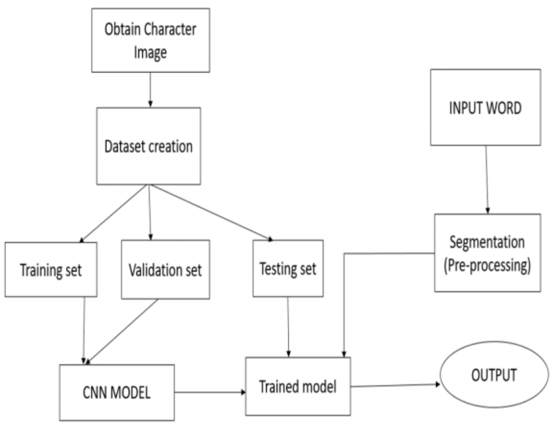

Fig 6.1 shows the proposed model.

A. PREPROCESSING THE DATASET

In the preprocessing stage, the character image is processed for removing all the undesirable entities from an image to make the process of recognizing easier. The input images are resized to a suitable format because it must not be too large or too small. Initial step is to converting the character images into images with background since the image inputting for prediction contains the background otherwise model learns the incorrect data. The second step is to find the contours in the image, Contour tracing is a technique that is applied to digital images in order to extract their boundary, for finding the contours grey-scaling, binarization and cropping of the image should be done. Third step is to resize the image, it is done to reduce the size of characters to a particular minimum level. The images given to the model should be of same size. Forth step is augmenting the image, random elastic distortions are performed to the images and expand the characters to 500 samples each. The final step is to partitioning the dataset into train, test, and validation, here 55 percent of the dataset is partitioned as training set, 25 percent is partitioned as validating set and finally 20percent is partitioned into testing set.

B. DEFINING THE CNN ARCHITECTURE

Convolutional neural network is efficient architecture among deep learning models for image data set, since here we use character images CNN architectures are more convenient. Among CNN architecture ResNet architecture is chosen for training, the architecture of ResNet50 has 4 stages. The network can take the input image having height, width as multiples of 32 and 3 as channel width. Here the input size as 128x128x3. Every ResNet architecture performs the initial convolution and max-pooling using 7x7 and 3x3 kernel sizes respectively. Afterward, Stage 1 of the network starts and it has 3 Residual blocks containing 3 layers each. The size of kernels used to perform the convolution operation in all 3 layers of the block of stage 1 are 64, 64 and 128 respectively, The convolution operation in the Residual Block is performed with stride 2, hence, the size of input will be reduced to half in terms of height and width but the channel width will be doubled. As we progress from one stage to another, the channel width is doubled and the size of the input is reduced to half. Finally, the network has an Average Pooling layer followed by a fully connected layer having 1000 neurons (Image Net class output). This architecture is trained on more than a million images from the Image Net database, it can classify images into 1000 object categories. Since here only 105 classes are required we have to change the classification part of the network. In the classification part of the network we add an average pooling layer and then a fully connected dense layer with an activation functions of Relu and softmax.

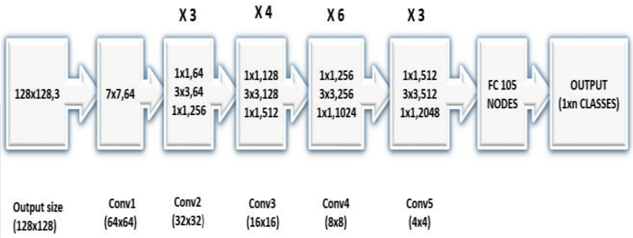

Fig 6.2 shows the representation of the Resnet model.

C. MODEL TRAINING

Training a model simply means learning (determining) good values for all the weights and the bias from labeled data's. Training a deep neural network that can generalize well to new data is a challenging problem. A model with too little capacity cannot learn the problem, whereas a model with too much capacity can learn it too well and over fit the training dataset. Both cases result in a model that does not generalize well. Here back-propagation learning method is used to learn the weights from the training dataset. The epochs (indicates the number of passes of the entire training dataset the machine learning algorithm has completed) used is 50, 70,100, batch size (refers to the number of training examples utilized in one iteration) used is 128, optimizers (update the weight parameters to minimize the loss function.) used are Adam and sgd, learning rate used for training has to be mentioned for proper training of the model and its 0.001.

D. PREDICTION

Prediction refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data. Here new (or external data) data is inputted by a graphical user interface (GUI). Here a segmentation threshold value range is provided to apply suitable threshold value for different images, since the threshold value applied for the words for segmenting into characters may vary depending upon the paper materials, writing patterns etc. The output

 DOI : 10.35291/2454-9150.2020.0515

from the final layer is the probabilities of the 105 classes, the node with maximum value will be detecting the character, index value of the node with higher probability value are extracted, with the help of the mapping file and the index value we get the Unicode of the corresponding character.

VII. RESULTS AND DISCUSSION

In this section, the results of our experiment are analyzed and discussed. The dataset contains 105 character each has 500 samples. Among them 55% data is used for training, 25% used for validation and remaining 20% for testing. We trained the dataset by using various models such as Resnet, Inception, mobile net with 120 epochs, optimizer used here is Adam. Accuracy is determined based on the training and validation data for deep learning technique.

Table 7.1: Performance of different CNN architecture

Model	Training accuracy	Testing accuracy	Parameters
Resnet	99.3	97.2	24,690,665
Inception	98.4	96.1	22,905,737
Mobile net	88.5	82.2	3,807,529

The above table shows the performance summary of the CNN architectures, among them the highest accuracy is achieved by Resnet by 99.3% and the least accuracy is achieved by mobile net by 88.5 accuracy.

Accuracy graph of the ResNet, Inception models and mobile net are shown in figures from 7.1 to 7.3 among the 3 models ResNet provide more accuracy. From the plot of accuracy we can see that the model had been trained properly as the curves for accuracy on both datasets (training and validation set) has been approach to one in last few epochs. We can also see that the model has not yet over-learned the training dataset, since there is no variations on both datasets. From the plot of loss, we can see that the model has comparable performance on both train and validation datasets (labeled test). If these parallel plots start to depart consistently, it might be a sign to stop training at an earlier epoch.

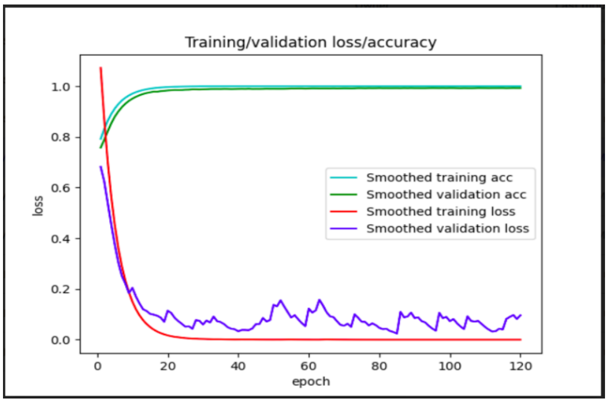

Fig 7.1 Accuracy and loss plot of Resnet.

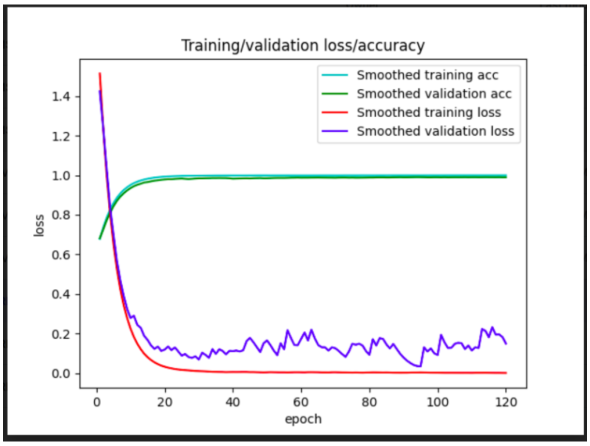

Fig 7.2 Accuracy and loss of Inception network.

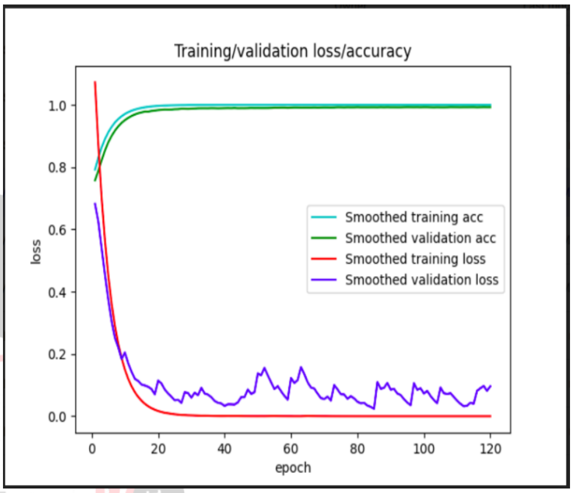

Fig 7.3 Accuracy and loss plot of mobile net.

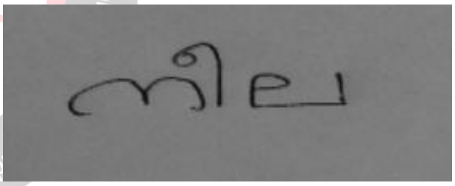

Fig 7.4: A word inputted to the model.

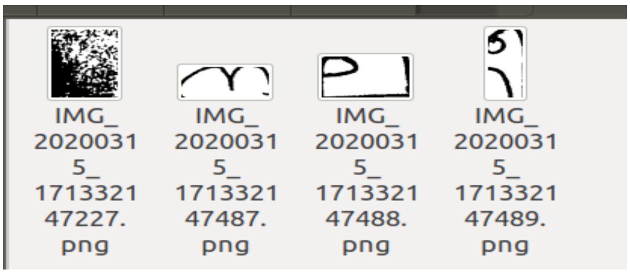

Fig 7.5: The contours detected when the inputted word is given

When a given word is inputted first of all the contours are detected which is used to separate each characters from the word, figures 7.4 and 7.5 shows the word inputted by the user and the corresponding contours detected, these contours are stored in a sorted order in the list to maintain the order of the characters in the word. Here 4 contours are detected among them one is the incorrect contours, we eliminate the incorrect contours by calculating the average

 DOI : 10.35291/2454-9150.2020.0515

ISSN : 2454-9150 Vol-06, Issue-04, July 2020

area of the contours. These contours are inputted to the model and probabilities values are generated in the output layer corresponding to each node (represent the characters), index value of the node with higher probability value are extracted, with the help of the mapping file and the index value we get the Unicode of the corresponding character, the process is repeated for all the contours in the list and finally displaced each characters in the word.

During the prediction of handwritten words ResNet provide more accurate prediction compared to mobile net and inception network models, in these two models a lot of miss-predictions has been occurred.

VIII. CONCLUSION

Among the methods used in literature survey it was found that deep learning methods shows more accuracy than the existing methods for the recognition of word. Since the dataset deals with image data, which is mostly dealt by convolution neural network was able show a high prediction accuracy. The analysis of result shows an accuracy of 99 percent for the trained and validation data. Also we compare Mobile Net and Inception Network models for prediction and also we plot the performance graph of the models in the analysis section. When an external input is provide to the model almost 97% of correct prediction has been occurred. Also we could find that due to the limitation in the dataset and writing style of users, some characters got misclassified.

An attempt was made to avoid the dataset creation task by calculating inter and intra class distances of the variants (characters of similar types) of some characters, to find the similarity between them, since the threshold value of the distances are varying during each iteration this method was not effective.

The work can be further extended in cooperation with the real time data, also an application can be developed for recognizing old scripts in the document verification scenarios. The recognition can also be done in extended classes that is more classes can be included for the analysis of user data.

REFERENCES

[1] Pravan P Nair, Ajay James, C Saravanan “Malayalam Handwritten Character Recognition Using Convolutional Neural Network" ICICCT (2017).

[2] Manju Manuel, Saidas S, R “Handwritten Malayalam Character recognition using curvelet and ANN", IJCA (2015).

[3] K. Gaurav and Bhatia P. K., “Analytical Review of Preprocessing Techniques for Offline Handwritten Character Recognition”, 2nd International Conference on Emerging Trends in Engineering & Management, ICETEM, (2013).

[4] Yu weng, Chunlei kai, “A New Deep Learning Based Handwritten Recognition System on mobile computing devices (2019).

[5] Salvador España-Boquera, Maria J. C. B., Jorge G. M. and Francisco Z.M., “Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 4, April 2011.

[6] Savitha Attigeri, “Neural network based handwritten character recognition system”,22 march (2018).

[7] Rajashekararadhya S. V. and Vanaja Ranjan P. (2008), ‘Isolated handwritten Kannada and Tamil numeral recognition: A novel approach’, in Proceedings of the First International Conference on Emerging Trends in Engineering and Technology, pp.1192-1195.

[8] Shailesh Acharya, Ashok Kumar Pant, Prashnna Kumar Gyawali “Deep Learning Based Large Scale Handwritten Devanagari Character Recognition”, 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), (2015).

[9] M. M. Rahman, M. Akhand, S. Islam, P. C. Shill, and M. H. Rahman, “Bangla handwritten character recognition using convolutional neural network, “International Journal of Image, Graphics and Signal Processing, vol.7, no.8,pp.42–49, (2015).

[10] T.Som, Sumit Saha, "Handwritten Character Recognition Using Fuzzy Membership Function", International Journal of Emerging Technologies in Sciences and Engineering, Vol.5, No.2, pp. 11-15, Dec (2011)

 DOI : 10.35291/2454-9150.2020.0515