Comparison Study of Handwritten Digit Recognition using Artificial Neural Network and Convolutional Neural Network: A Review

[1]Sonia Flora, [2]Anju Kakkad

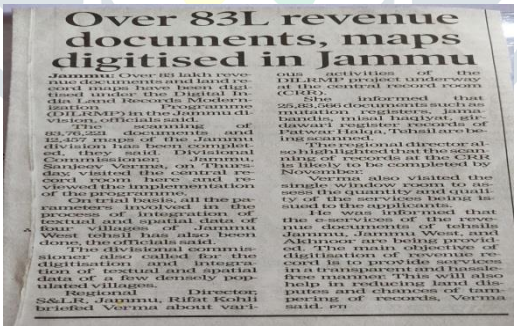
[1]M.Tech Student, [2]Assistant Professor,
[1]Parul Institute of Engineering and Technology,
[1]Parul University, Vadodara, India

Abstract : Handwritten Character Recognition is one of the important application of Computer Vision. There has been a plentiful research done in the Handwritten Character Recognition using various Machine Learning algorithms on various Indian scripts like Devnanagri, Telugu , Tamil, Gurmukhi etc in last 3 decades but optimum recognition rate are yet to achieve in some scripts. In this review paper, we are comparing the classification accuracy of handwritten digit using artificial neural network and using state of the art deep learning model i.e. convolutional neural network.

IndexTerms **– Handwritten Digit Recognition, Artificial Neural Network, Convolutional Neural Network, Deep Learning, Digit Classfication.**

I. INTRODUCTION

Whenever we heard a word Handwritten Character Recognition, the very first term came to our mind is OCR, that is, Optical Character Recognition, it is a process of reading character from a physical paper and convert it into a form that computer can manipulate. This was the only technology used by the researchers since 3-4 decades to convert any physical document into computer editable form. OCR is a process of 5 stages i.e Image Acquisition, Pre-processing, Segmentation, Feature Extraction and Classification. All stages of OCR follows the pipeline concept, successive rate of each stage depends upon the success rate of previous stage. With the advancement of technology, we want machine to perform maximum task. There are abundant applications of computer vision like translation of document, language interpretation, sorting mail, reading checks etc. That is why OCR is an area of interest for several researchers in recent years. Primarily, texts being in different languages and scripts, the efficiency of recognition system largely depends upon the classification methods used. Thus the development of such system was a tough task. Earlier the recognition system was built using machine learning classification techniques, and, one has to extract the features of the image manually and fed it to the classifiers. The recognition results achieved using machine learning algorithms like SVM are quite satisfactory but it includes large amount of mathematics to extract the features of an image manually. But with artificial neural network and deep neural network, the recognition rates are quite desirable and almost no complex mathematics include. The character recognition system using deep neural network for various Indian scripts have already been developed with almost optimum results. Fig 1.1 shows the newspaper cutting in which almost 83 Lacs Land papers of Jammu are being digitised for the ease of the common man.



Over 83L revenue documents, maps digitised in Jammu

Jammu: Over 83 lakh revenue documents and land revenue maps have been digitised under the Digital Jammu and New Modern Land Records Programme (DILRMP) in the Jammu district, officials said.

The scanning of 83,76,258 documents and 12,457 maps of the Jammu division has been completed, they said. Divisional Commissioner, Jammu, Sanjeev Verma, on Thursday visited the central record room here and reviewed the implementation of the programme.

On trial basis, all the parameters involved in the process of integration of textual and spatial data of four villages of Jammu West tehsil has also been done, the officials said.

The divisional commissioner also called for the digitisation and integration of textual and spatial data of a few densely populated villages.

Regional Director, S&LR, Jammu, Rifat Kohli briefed Verma about various activities of the DILRMP project underway at the central record room (CRR).

She informed that 25,83,546 documents such as mutation registers, jamabandis, misal haqiyat, girdawari register records of Patwar Halqa, Tehsildars are being scanned.

The regional director also highlighted that the scanning of records at the CRR is likely to be completed by November.

Verma also visited the single window room to assess the quantity and quality of the services being issued to the applicants.

He was informed that the services of the revenue documents of tehsils Jammu, Jammu West, and Akhnoor are being provided. The main objective of digitisation of revenue record is to provide services in a transparent and hassle-free manner. This will also help in reducing land disputes and chances of tampering of records, Verma said. PTI

Fig 1.1 Digitising Jammu Land Papers

II. RELATED WORK

In recent years attempts have been made to develop text recognition systems in almost all the languages and scripts of the world. Many researchers have worked on the problems of recognizing the offline characters or digits. Archana.N.Vyas et al [1] showed the comparison of different classifers like KNN, SVM, and Neural Network with different feature extraction techniques including Fourier descriptors and discreet cosine transform and achieved accuracy of 85.67 %, 93.60 % and 93 % respectively. Akanksha Gaur et al [2] used the K Means Clustering for feature extraction and SVM with linear kernel as classifier and attained accuracy of 95.86 %. Neha Sahu et al [3] showed the recognition of devanagri character with the help of neural network with two hidden layers and 244 epochs and got accuracy of 75.6 % on noisy characters. Manoj Kumar Mahto et al [4] have reported that by using combined and vertical projection feature extraction & SVM as classifiers on gurmukhi character recognition achieved an accuracy of 98.06 %.Gunjan Singh et al [5] projected a approach based on back propagation neural network for the recognition of handwritten hindi character and attained accuracy of 93 %. Ashutosh Aggarwal et al [6] proposed a recognition system for offline handwritten Gurmukhi Characters and Numerals using Gradient information as mode of feature extraction technique, two ways of feature extraction using gradient information are explained. They have used the SVM as classifier. Adwait Dixit et al [7] proposed a feature extraction & classification approach named as wavelet transform. The proposed technique achieve a accuracy of 70%.Wavelet transform is used in order to extract wavelet features. Generally two wavelet coefficients are determined named as approximation and directional coefficients. Now for each and every wavelet

coefficient, various parameters like mean, deviation will be found. Author has used artificial neural networks (ANN) for the classification stage. Gurpreet Singh [8] projected a recognition system for offline handwritten Gurmukhi character recognition .The author has used ANN and with the proposed system accuracy of 98.96% is achieved.

III. SCRIPT AND DATASET

In India, there are 29 states and 22 different languages have been spoken by the people of India living in different states. Each language has some rules to read and write that particular language, and these are mentioned in the Script of that language. Likewise Hindi language is written in Devanagari Script, Punjabi Language is written in Gurmukhi Script and English Language is written in Roman Script. One can create its own dataset for experimentation or can download the various dataset available online. MNIST and CIFAR-10 are reliable dataset available online easily, if one wants to start from scratch. MNIST dataset has 60,000 handwritten digit images available. It is available online in different formats like image form or in .csv file (Comma Separated Values). One can download the dataset by using Google tab of Datasets. In order to realize the performance of deep learning, one needs a huge amount of data, may be in lacks because the models of deep learning gives their best in case of large amount of data only otherwise performance difference cannot be figured out accurately.

IV. THE RECOGNITION MODEL

To create a recognition model of offline handwritten character recognition, one has to pass through the following 5 stages. Each stage has its own importance and success rate of each state depend on success rate of previous state. Fig 4.1 shows the sequence of stages.

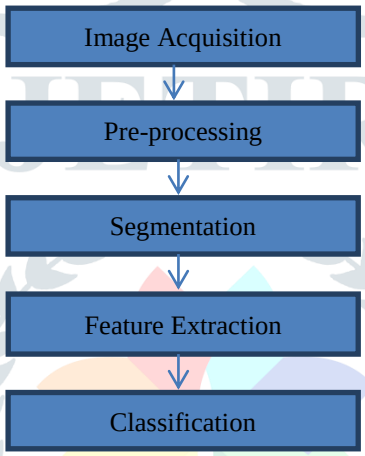


Fig 4.1 Character Recognition Stages

4.1 Image Acquisition: It is a process to digitize the handwritten character document using a capturing device i.e. scanning the document with proper resolution and save it as a 'jpg' or 'bitmap' image format.

4.2 Pre-Processing: It includes the operations which will be performed on the image format. To make all the scanned images effective to provide the better results and removal of any noise and filling the holes, skew detection and correction etc techniques applied as per requirement. The output of this stage is Binary Image.

4.3 Segmentation: This step is used to segment the characters in case of multi character. This can be done using cropping the image and make all images of uniform size. Segmentation can be done in three ways i.e. Line segmentation, word segmentation and Character Segmentation depending upon the requirement of the recognition system.

4.4 Feature Extraction: This is the heart of the recognition system. It finds the interest features of an image and these interest features may include height, width, density, loops, lines, stems and other character traits. Each character is judged based on these characteristics and basic features [3]. There are different feature extraction algorithm which are used to fetch the interest features. It is not necessary all the extracted feature of an image are useful so it is minimized by saving only the key information. This will give us a vector with scalar values is called feature vector. This stage was mandatory and of high importance when machine learning algorithms were used for classification.

4.5 Classification: The classification stage is the decision making part of a recognition system and it uses the features extracted in the previous stage. There are various classifiers like Support Vector Machine (SVM), K nearest neighbour (KNN) etc .These classifier will be trained on 80% of the dataset and then tested on 20% of the dataset. The evaluation metric will determine the accuracy of the classifier. In this paper we have decided to compare the accuracy of two classifiers Artificial Neural network and Convolutional Neural network.

V. CLASSIFIERS

5.1 Artificial Neural Network: It emulates the human brain's ability to learn and recognize patterns. Neural network is mainly used for classification problems. The feed forward artificial neural network, or also known as multilayer perceptron, is a fully connected network model that maps the input data sets into the corresponding output sets. The multilayer neural network consists of three layers – input layer, one or more hidden layer and output layer. Each node performs the simple computation and connections between nodes are labelled with a number called weight or "connection strength". The main advantage of using the neural network in pattern recognition is that it will remove the need of external feature extraction techniques. If the neural network is properly designed with good number of hidden layers and nodes then there is no

requirement of any manually feature extraction techniques. Back propagation is the most common neural network learning algorithm. It uses the gradient descent technique to update the weight and bias while learning.

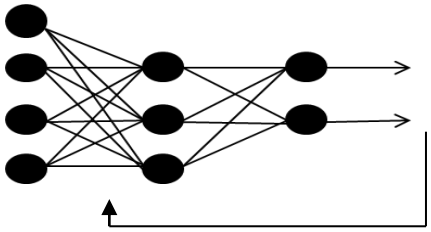


Fig 5.1 A Multi-Layer Neural Network

5.2 Convolutional *Neural Network (ConvNet or CNN)*: This is one of the deep learning models which is specifically used for the processing of computer vision applications. ConvNet is very much similar to the regular neural network consisting of weights and biases and more number of layers unlike regular neural network. Each neuron receives input and performs dot product with corresponding weight and adds bias. And uses different activation functions to get the output of single neuron or to get the output of different layer and that output can become the input to the next layer. The artificial neural networks do not perform well in case of large images. Convolutional Neural Networks take advantage of this fact and gives the best performance results in case of any size of images. But the performance of the convolutional neural network model can be well judged with huge amount of data. Generally, Convolutional Neural network consists of three main layers - Convolutional Layer, Pooling Layer, and Fully-Connected Layer. The sequence of layers are shown in fig 5.2. Every layer of a ConvNet transforms one volume of activations to another through a differentiable or activation function. It uses the reLU(Rectified Linear Unit) activation function to get the output of a layer and give as input to the next layer. Convolution layer applies the convolution function with the help of kernel on each and every part of an image to fetch the feature to distinguish between visually different images. Max Pooling or Average Pooling layer is used to downsample the image data. Fully Connected Layer works like the artificial neural network. One has to use the flatten function to convert two dimensional data into single dimensional to give the input to the Fully Connected Layer. Fully connected layer are also called Dense Layer [9].

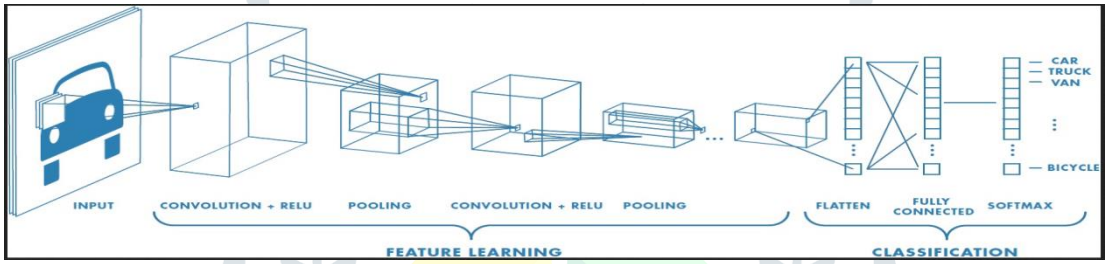


Fig 5.2 Convolutional Neural Network architecture [10]

VI. COMPARISON RESULT

6.1 *Implementation using Artificial Neural Network:* The classification of handwritten digit 0-9 has been implemented using Python3.6 with PyCharm developing environment. The preprocessing of the scanned images is performed and made all images of uniform size. The dataset is load using the Numpy library of python. Keras library is used to create the neural network. The neural network consists of three main layers, input layer, hidden layer and output layer. The number of neurons in the input layer is equal to the size of the input image [height x width] 28 x 28 i.e. 784 pixel, therefore the number of nodes in input layer is 784. To get the output of each layer, reLU activation function is used, and its output is given as input to the next hidden layer. The hidden layer consists of same number of neurons as there are in input layer i.e 784. The output layer consists of 10 neurons as there are 10 (0-9 digits) different class lables. The softmax activation function is used at the output layer to get the output as the probability like values of 0-9 digits, which further help to select maximum one of the value from 10 probability values that will be the class label. Neural network model has been trained on 60,000 images over 10 epochs with batch size as 200. Categorical cross entropy (Logarithmic) function is used as the loss function and ADAM (Adaptive Moment Estimation) optimization algorithm is used to learn the weights and biases. For the evaluation of the model, baseline error is calculated using the accuracy metric on the test images. And the value of Baseline error achieved as 1.91%.

6.2 *Implementation using Convolutional Neural Network (CNN or ConvNet):* Now with the same dataset, trained the CNN model. CNN are bit more complex to train than regular neural network. In this case we have trained two dimensional convolutional neural networks. The first layer i.e Convolutional layer, expecting images as input in the form of [height][width][channel], gray scale image will have channel value of 2 , i.e. 0 and 1. In this layer, we have to pass the number of filters required and the size of filter as parameters. These are hyper-parameters so usually we take the number of filter as 32 ,64 and so on and sizes of filter as 3x3, 5x5 etc. The model learns the value of filter basically i.e filter size and number of filter used in convolutional layer decides the number of parameters learned by the model. The output of this layer will be calculated using the reLu activation function. Next is a pooling layer that takes the max called MaxPooling2D of pool or filter size 2×2. To avoid the overfitting, one can use the drop out function. This layer can be called as regularization layer or Dropout layer which randomly excludes 20% of neurons in the layer in order to reduce over fitting. Until now the output which we were getting from previous layer was in the two dimensional form, to give this input to the fully connected layer, we first have to convert it into one dimensional using Flatten function. Then give the output of flatten function to the first dense layer with 128 neurons and reLu activation function. And finally, the second dense layer (output layer) has 10 neurons for the 10 classes and a softmax activation function is used to get output probability-like values for each class. The CNN model is trained on 60,000 images over 10 epochs with a batch size of 200. Logarithmic function which is also called categorical cross entropy in keras,

is used as the loss function and ADAM (Adaptive Moment Estimation) optimization algorithm is used to learn the weights and biases. For the evaluation of the model, baseline error is calculated using the accuracy metric on the test images. And the value of Baseline error achieved as 1.02%.

It is clearly seen that baseline error in case of convolutional neural network is less than the baseline error in artificial neural network. The key difference in both are number of layers and learnable parameters. In case of artificial neural network, all layers are fully connected and learnable parameters are more but performance is less and in convolutional neural network, initial two layers i.e. convolution layer and max pooling layer are processing the image with the help of filter followed by the fully connected layers and number of learnable parameters are less but performance is better.

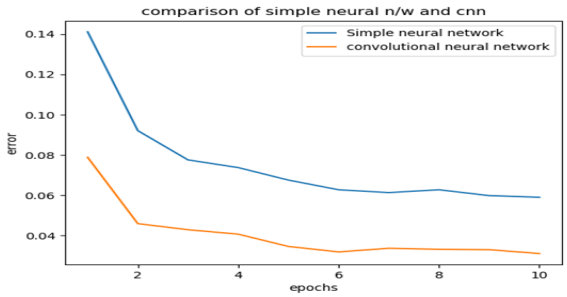


Fig 6.1 Comparison of Artificial Neural Network and ConvNet Results

VII. CONCLUSION

A dataset has been trained on artificial neural network and on convolutional neural network. The average error of both networks was calculated using the accuracy evaluation metric. The average error of CNN is less than artificial neural network on the CPU. Though while training CNN over CPU took more time than artificial neural network. But image classification is better performed in case of CNN. It can be concluded that as the model is trained with CNN, the accuracy of recognition increases respectively but if trained on GPU, one can get optimum result for the classification with CNN in good timing.

REFERENCES

[1] Archana N.Vyas, Mukesh m.Goswami, "Classification of Handwritten Gujarati Numerals", 978-1-4799-8792-4/15/$31.00 ©2015 IEEE

[2] Akanksha Gaur, Sunita Yadav, "Handwritten Hindi character Recognition using K-Means Clustering and SVM", 978-1-4799-5532-9/15/$31.00 ©2015 IEEE

[3] Neha Sahu, Nitin Kali Raman,"An Efficient handwritten Devanagari character Recognition System Using Neural Network", 978-1-4673-5090-7/13/$31.00 ©2013 IEEE

[4] Manoj Kumar Mahto, Karamjeet Bhatia, R.K.Sharma, "Combined Horizontal and Vertical Projection feature Extraction Technique for Gurmukhi Handwritten Character Recognition", 2015, International conference on Advances in Computer Engineering and Applications (ICACEA).

[5] Gunjan Singh et al, "Recognition of Handwritten Hindi Characters using Backpropagation Neural Network" 2012 / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (4) ,4892-4895\

[6] Ashutosh Aggarwal ,Karamjeet Singh, kamalpreet Singh,"Use of Gradient Technique for extracting features from Handwritten Gurmukhi Characters and Numerals", 2014, International Conference on Information and Communication Technologies (ICICT 2014).

[7] Adwait Dixit , Ashwini Navghane, ogesh Dandawate,"Handwritten Devnagari character Recognition using Wavelet Based Feature Extraction and Classification Scheme", 2014, Annual IEEE India Conference(INDICON).

[8] Gurpreet Singh, Manoj Sachan, "Multi-Layer Perceptron(MLP) neural Network Technique for Offline Handwritten gurmukhi Character Recognition", 978-1-4799-3975-6/14/$31.00 ©2014 IEEE

[9] Convolutional Neural Network [Online] Avaialble- http://cs231n.github.io/convolutional-networks/

[10] Convolutional Neural Network [Online] Available:- https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148