A Literature Survey on Handwritten Character Recognition using Convolutional Neural Network

Karishma Verma
Department of Information and Technology
YMCA University of science and Technology, Faridabad, India
Email: karishmaverma555@gmail.com

Dr. Manjeet Singh
Professor
Department of IT & CA
YMCA University of science and Technology, Faridabad, India

Abstract: Automatic handwritten character recognition has many academic and commercial interests. Nowadays, deep learning techniques already excel in learning to recognize handwritten characters. In computer vision handwriting recognition is the core problem and the main challenge in this field is to deal with enormous variety of handwriting style by different writers in different language. The similarities of the neighboring characters make further complicate the character recognition problem. The large variety of writing style of different writers and the complex features of the handwritten characters are very challenging for accurately classifying the handwritten characters. This paper presents a detailed review in the field of handwriting recognition using convolutional neural network.

Keywords: Handwriting recognition, Deep learning, Convolutional neural network, Optical character recognition

I. INTRODUCTION

Image classification is one of the core problems of computer vision, image classification refers to the task of extracting information classes from a multiband raster image. One of the important applications of image classification is the optical character recognition (OCR). OCR is the electronic conversion of scanned images of handwritten text into machine encoded text. In Optical character recognition, an algorithm will be trained on a dataset of known character in order to learn how to classify characters included in test set accurately. A variety of algorithm has been developed for classifying letter and digits from last decades. In the field of optical character recognition there are three methods: template matching, feature extraction and classification, third one is the deep learning method. Template matching is the very first method used in the field of recognition through it is relatively simple than other method but the problem of some illegally written text such as difference of diameters, discontinuity of text or rotation, the recognition accuracy is not satisfactory. In feature extraction and classification, the key point is feature extraction algorithm, it determine the recognition rate and furthermore, this part need researchers to try different kind of feature manually, so a large number of experiment needed in this method. In the early stage of OCR, template matching and structural analysis are used popularly [1].

The late 80s in order to take advantage of massive samples, classification methods such as artificial neural network had been utilized popularly for recognition problems [2]. In the last decade, machine learning methods such as support vector machines (SVMs) have been applied for pattern recognition problems [3]. Neural networks (NNs) are another solution to resolve recognition problems. In this a large number of handwritten letters/digits known as training set are fed into the algorithm in order to infer rules automatically for handwritten character recognition [4].

The coming deep learning led to using the convolutional neural network (CNN) in machine vision problem [5]. Based on Hubel and Wiesel's early work on cat's visual cortex, CNN is a biological inspiration of multilayer perceptron (MLP) [6].

In this paper, section IV describe the method of deep learning to recognize characters, which is like the feature extraction and classification method, this method also first extract the image feature and then does the classification work but like in feature extraction and classification method researchers need not design the feature manually. In deep learning method what we have to do is to define the structure of network and then tune its parameters.

II. MOTIVATION

Many machine learning techniques able to learn and recognize hand written character to a greater extent. Handwritten character recognition (HCR) is a challenging task because of its variety of handwriting style by different type of writers in different type of language, this is the main challenge which is face by many researchers now a day in this field. Since all the information has shifted from handwritten documents to digital file format because of its reliability and durability. However a large amount of old documents are still in hand written forms so the attempt to convert them into digital form using manual typing to copy an existing file takes a lot of time and also require manpower to accurately manage each document and its copy as well to fulfill the task, this method of conversion is necessary for future handwritten documents also to make them digital documents. The main problem arises here is the hand writing styles since every different people has its own approach to handwriting in different language. Furthermore, many languages characters encompass a high range of characters that are morphologically complex. There is some other character also that made up of more than one character known as compound characters which make the task even more challenging. The above stated problem motivated us to build a system that will recognize characters in a way that reduces its manpower cost and also reduce the time to translate them.

III. ARCHITECTURE

A. Biological connection

Convolutional neural network (ConvNet/ CNN) do take a biological inspiration from visual cortex. The visual cortex has a small area of cells which are sensitive to a specified region of the visual field. Two researchers **D.H. Hubel** and **T.N. Wiesel** in 1962 showed that some individual's neuron cells in the brain get fired only in the presence of certain orientation. For example, some neurons fired when exposed to vertical edges and some are fired when exposed to horizontal or diagonal edges. They found out that all these neurons were organized in a columnar architecture and together they produced a visual perception. The idea of having specialized component inside a system having specific task is used by machines as well, and this is the basis behind successful CNNs.

B. Convolutional neural network

Human brain is a very powerful machine. We see multiple images every second and process them without realizing how the processing is done. But this is not the case with machines. The first step in image processing or character recognition is to understand, how to represent an image so that a machine can read it. In simple term, image is a collection of dots (pixels) arranged in a order. If you change the order or colour of a pixel, the image would change as well. All CNN models follow a similar architecture, as shown below. There is an input image that we are working with and then perform a series of convolution, pooling operations followed by a number of fully connected layers.

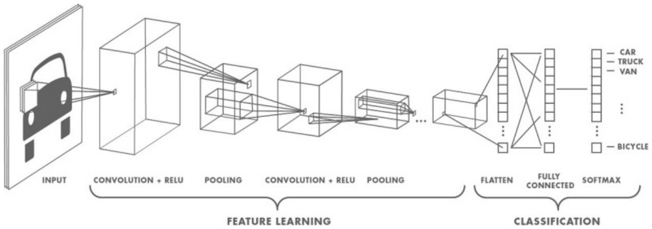


Figure 1: Convolutional neural network

The LeNet Architecture (1990s)

LeNet was one of the very first CNN. This pioneering work by Yan LeCun was named LeNet5 after much previous successful iteration since the year 1988. At that time this architecture used mainly for character recognition task such as reading zip codes, digits etc. However several new architecture proposed in the recent years which are improvements over the LeNet, but all uses the main concept from the LeNet and are relatively easier to understand. There are mainly four operation needed to built a simple ConvNet: Convolution, Non linearity, Pooling or Subsampling and classification (fully connected layer). These operations are the basic building block of every convolutional neural network channel is a term which used to refer a certain component of an image. A standard image will have three channels Red, Green, Blue. Each having pixel value in the range of 0 to 255. A gray scale image on the other hand has just one channel.

The convolution step**:** Convolution neural network derive their name from the operator "convolution". The primary purpose of this operator in case of CNNs is to extract features from the input image. This is the first layer in CNN, the input to this layer is a 3D array (32*32*3) of pixel value. Convolution is a mathematical operation to merge two set of information in other our case, first is input image and the second set is the filter.

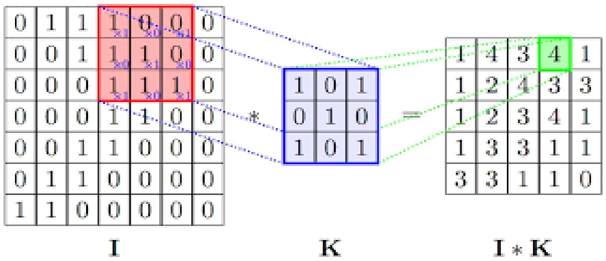
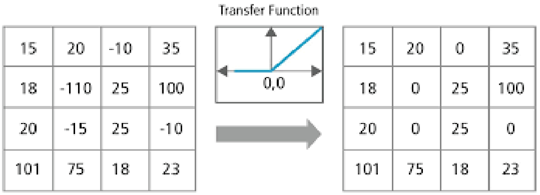


Figure2: Convolution process

In figure 2, on the left side we have input image, convolution filter is located in the centre which is also called kernel, detector or feature whereas on the right side of figure 2 we have the output of the convolution process called activation map also called feature map, convolved feature. We perform convolution operation by sliding this filter over the input image. At every location we do an element wise matrix multiplication and sum the result, this resultant matrix is called Convolved feature. CNN learn the value of these filters on its own during training process.

Non linearity (ReLu): This is the addition operation called ReLu has been used after every convolution operation. ReLu is called rectified linear unit and is a non linear operation. It is given by: Output = Max(zero, input) as indicated in figure 3. ReLu is an element wise operation and replaces all negative pixel values in the feature map by zero.

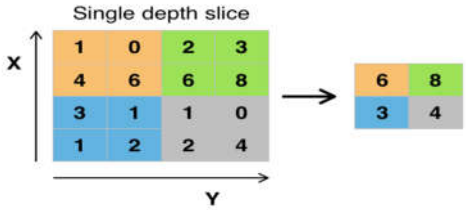


15	20	-10	35
18	-110	25	100
20	-15	25	-10
101	75	18	23

15	20	0	35
18	0	25	100
20	0	25	0
101	75	18	23

Figure 3: ReLu operation

Pooling Step: Pooling (also called subsampling or downsampling) it reduces the dimensionality of each feature map but retain the most important information. This layer makes the input representation smaller and more manageable. It also reduces the number of parameters and computations in the network and controlling from over fitting. The output of pooling layer acts as an input to the fully connected layer which is the next layer in the CNN after this. Figure 4 shows the pooling operation in CNN architecture.



1	0	2	3
4	6	6	8
3	1	1	0
1	2	2	4

6	8
3	4

Figure 4: Pooling process

Fully connected layer: The fully connected layer is a traditional multilayer perceptron, it uses softmax activation function in the output layer. This term implies that every neuron in the previous layer is connected to every neuron in the next layer. The output from convolution layer and pooling layer represent high level features of image. The function of fully connected layer is to classifying these features of input image into various classes based on the training dataset. The sum of output probabilities from fully connected layer is 1. This is ensured by using softmax as the activation function in the output layer of fully connected layer.

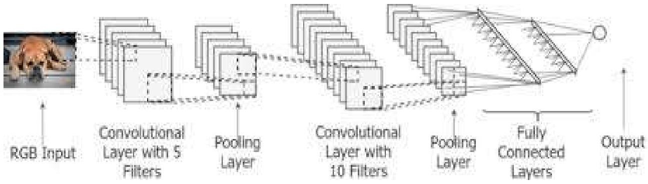

RGB Input
Convolutional Layer with 5 Filters
Pooling Layer
Convolutional Layer with 10 Filters
Pooling Layer
Fully Connected Layers
Output Layer

Figure 5: Fully connected layer

Training using back propagation: the overall process of training of CNN is summarized as below:
1. We initialize all filters weight with random value.
2. The network takes a training image as input, goes through forward propagation step (convolution, ReLu and Pooling operations) and finds the output probabilities of class.
3. Calculate total error at the output layer.
4. Use back propagation to calculate gradient of error and use gradient descent optimizer to update all filter values.
5. Repeat step 2 and 4 with the training set of all images.

IV. RELATED WORK

Advent of deep learning encourages us to use convolutional neural network in machine learning problem. Convolutional neural network (CNN) is a multi layer perceptron model proposed by D.H.Hubel and T.N.Weisel [1] that is inspired by biological neuron and animal vision. The basic design of CNN is inspired from the scientist Yann le Cunn who conducted certain experiments regarding mammal's perception and vision. Since the day of propose of CNN different researchers have created different version of the CNN and developed different type of applications like character recognition, object detection, face recognition and sound detection. Some of the works are presented here:

Bishwajit purkaystha, Tapos Datta,Md Saiful Islam [7] *Et al* , this paper proposes a convolutional neural network model for recognizing handwritten Bengali characters. They proposed a convolutional neural network using two convolution layers and then using three densely connected layers, on the final dense layer Softmax Function is used. The first layer scans for 5*5 receptive field throughout the image. After scanning it is then passed through ReLu activation function. The ouput from this layer is then passed to max pooling layer having size 2*2. This output then again passed to second convolutional layer having 64 kernels of size 3*3. As in first layer same function ReLu and pooling of size 2*2 is applied. The output through this layer is actually used for classification process. They achieved 98.66% accuracy on numerals, accuracy of 94.99% on vowels, accuracy of 91.23% on alphabets and 89.93% accuracy on all Bengali characters. Some of the error in their prediction was due to misleading of the dataset.

Samad Roohi, Behnam Alizadehashrafi [8] *Et al*, the study behind this paper is to investigate the performance of deep convolutional neural network on the Persian handwritten character dataset. To determine the outperformance of conventional method in PHCR problem, in this paper two type of CNN have been implemented. First network is single convolutional neural network which is based on the simple structure of CNN for example (LeNet-5). To extend it into ensemble CNN the bagging paradigm applied on CNN with a variety of network parameters. This paper concluded that the performance is 97.1% when ensemble CNN model was implemented and 96.3% accuracy when single CNN was implemented.

Mahesh jangid and sumit srivastava[9] *Et al,* this paper proposes the implementation of devanagari character using deep neural convolutional network (DCNN). They used the database provided by ISI (information sharing index) Kolkata, ISIDCHAR database and V2DMDCHAR database. They built six different architectures of DCNN to determine the performance of the network and also uses six different adaptive gradient method they use Adagrad optimizer, RMSProp optimizer, Adaptive Momentum estimation optimizer (Adam) etc. they concluded that AdaDelta, AdaMax and RMSProp optimizer outperformed the accuracy of Adam optimizer and got 97.30% accuracy on ISIDCHAR dataset by using DCNN. Similarly on V2DMDCHAR they achieved 97.65% accuracy by using layer wise DCNN and 96.45% accuracy by using DCNN and on the combination of both dataset they gained accuracy of 98.00% using layered wise DCNN and got 96.53% by using DCNN.

Jia Xiadong, Gong wedong, Yuan Jie [10], CNN is used by many researchers for character recognition task but they concluded two main problem which they face while using CNN.

(i)The first problem in CNN is the manual training of data is very time consuming and it is labor intensive.

(ii)Second problem is the design and parameter problem all depend on experiences.

(iii) To address these problem stated above they use density based clustering algorithm which is used as remedy for the first problem that means it is effective in data labeling. In this paper the handwritten Yi character recognition is divided into two parts. In the first part they describe how to construct the Yi characters database which uses improved density based clustering algorithm and in the second part they explain CNN architecture. In the construction of dataset they scan the Yi character documents then they binarize and normalize them and finally an image slice is used as the input in the clustering algorithm.

The CNN is formed by using four convolutional layers, four max pooling layers and one fully connected layer. The size of input image is 52*52 and the size of kernel or filter is 5*5 and 20 kernels are used therefore the feature map is 20 and the output size of feature map is 48*48 then max pooling layer was applied which reduces the output size into 20*20. In the second convolution layer they used 60 kernel of size 5*5 the feature obtained are 20*20 in size and 60 in number after pooling 0f size 2*2 the size of the output feature map is 10*10. In the third layer they use 120 kernel of size 3*3 after pooling the size 4*4. In the last layer kernel size is 2*2 they got 2*2 feature map and after that they applied logistic regression layer whose output is 100 because of 100 clusters in the database. Finally they achieved 99.65% accuracy on test set. They also concluded that if we pay more attention towards the combination of CNN with different techniques then there are chances of higher accuracy.

Ashok kumar pant, Prashnna kumar Gyawali Shailesh Acharya [11] they propose system for recognizing devnagari characters. They propose their own dataset of devnagari characters. In the dataset, 92 thousand images of 46 different classes of characters are segmented from handwritten documents. Along with the dataset they also propose the CNN architecture which was based on simple CNN. They built two models, model A and model B. In model A, the image in the dataset is rescaled from 36*36 to 28*28 which was convolve with 5*5 filter and a 2*2 size pooling layer was used which gives a output of 14*14 feature map. In the second convolutional layer this 14*14 feature map got convolved with 5*5 filter after that 2*2 max pooling layer was applied on it and the final output was of size 5*5 feature map after that Softmax Function was applied. The model B was derived from LeNet family. It has shallow architecture that means it consist of fewer number of parameters. They used mini batch size 200 and 50 epochs with learning rate 0.005 for model A and for model B it is 0.001 then stochastic gradient descent optimizer with momentum 0.9 is applied. The higher test accuracy obtained for model A is 0.98471 and for model B the higher test accuracy is 0.982681 addition of dropout show better result in model B.

Mathew Y.W. Teow [12] presented a work on how to bridge the gap on understanding the mathematical structure and the computational implementation of a CNN using minimal model they proposed a minimal CNN which consist mainly two computational networks. They are feature learning network and the classification network. The objective of fully connected network is to perform unsupervised image feature learning. The learned image feature will be synthesized by feature learning with high level representation. Finally this high level image will be input to the classification to perform image recognition. They used MNIST dataset for handwritten digit recognition which consist of 70,000 different handwritten digits images where 60,000 digits are used for training and 10,000 images used for testing. Each digit image is of 28*28 in size and in grey scale. 20 trainable convolution kernels with kernel size 9*9 and used with input image, convolution perform feature extraction and generate 20 convolved feature map. This convolved feature rectified by ReLu activation function finally pooling layer perform mean pool to rectify feature map. Another proposed model is the extended version of the minimal model where they used 2 pooling layer; 2 convolution layer and 2 ReLu layer they called this model as extended minimal model. And then compare this with LeNet 5. The performance of the minimal model is 97.3% and extended minimal is 98.50% with epoch 10.

Mujjadded AL Rabbani Alif, Sabbir Ahmed, Muhammad Abdul Hasan [13] *et al,* in this paper they proposed a modified ResNet-18 architecture for Bangla handwritten characters. In order to measure the performance they used two recently introduced large dataset called Bangla lekha-Isolated dataset and CMARTER db dataset. The image size of this dataset vary from 110*110 to 220*220 pixels. Handwriting of this dataset is collected from 4 to 27 years age group. In the first experiment, to measure the performance three optimizers were used namely RMS prop, Adam and SGD on 110*110 inputs then using Adam they achieve 0.4% and 0.1% performance. In Second experiment they investigate the performance of proposed method with different dropout rates. Then investigate the performance of bangla character recognition using several states of art CNN models. They use VGC Net-16, VGC Net-19, ResNet-18, ResNet-34 and achieve 91.0%, 92.11%, 94.52%, 94.59%.

V. CONCLUSION

In this paper, an overview of convolutional neural network and working of all layers is presented. This survey concludes that the different parameters like dropout, optimizer, learning rates greatly influenced the efficiency of the CNN architectures. Preprocessing of dataset also consider a major factor behind the accuracy of the models. Training of network requires a supportive hardware to implement large dataset efficiently. In some language the characters are differ by only a single dots for that characters it is necessary to train the network with large amount of dataset so that the network easily recognize the character for that there is a requirement of high memory and high processing speed to achieve a efficient network. The articles provided in the literature survey contributes different networks with different tuning of parameters on different types of dataset which help in achieving efficient network in future that can even recognize a whole sentence of different handwritten languages.

VI. REFERENCES

[1] S. Mori, C. Y. Suen, and K. Yammamoto, “historical review of OCR research and development,” Proc. IEEE, vol. 80, no. 7, pp. 1029-1058, 1992.

[2] A. Rajavelu, M. T. Musavi, and M. V. Shirvaikar, “A neural network approach to character recognition,” neural network, vol. 2, no. 5, pp. 387-393, 1989.

[3] H. Byun and S. W. Lee, “Applications of support vector machines for pattern recognition: A survey,” in pattern recognition with support vector machine, springer, 2002, pp. 213-236.

[4] P. D. Gader, M. Mohamed, and J. H. Chiang, “Handwritten word recognition with character and inter-charcter neural networks,” IEEE Trans. Syst. Man Cybern. Part B Cybern., vol. 27, no. 1, pp. 158 -164, 1997.

[5] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” nature, vol.521, no.7553, pp. 436-444, 2015.

[6] D. H. Hubel and T. N. Wiesel, “Receptive fields and functional architecture of monkey striate cortex,” J. phyiol., vol. 195, no. 1, pp. 215- 243, 1968.

[7] Bishwajit Purkaystha, Tapos Datta, Md Saiful Islam, “Bengali Handwritten Character Recognition Using Deep Convolutional Neural Network”, 2017 20th International Conference of Computer and Information technology(ICCIT), 22-24 December, 2017

[8] Samad Roohi, Behnam Alizadehashrafi, “Persian Handwritten Character Recognition Using Convolutional Neural Network,” 10th Iranian Confernce on Machine Vision and Image Processing, Nov, 22-23, 2017, Isfahan Univ. of Technology, Isfahan, Iran.

[9] Mahesh Jangid and Sumit Srivastava, “Handwritten Devnagari Character Recognition Using Layer Wise Training of Deep Convolutional Neural Networks And Adaptive Gradient Methods”, 2018 journal of imaging.

[10] Jia xiaodong, gong wednog, yuan jie, “Handwritten Yi Character Recognition with Density Based Clustering Algorithm and Convolutional Neural Network”, 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference Embedded and Ubiquitous Computing (EUC).

[11] Ashok Kumar Pant, Prashnna Kumar Gyawali, Shailesh Acharya, “Deep Learning Base Scale Handwritten Devanagri Character Recognition”,

[12] Matthew Y.W. Teow “Understanding Convolutional Neural Network Using A Minimal Model for handwritten Digit Recognition”, 2017 IEEE 2nd International conferences on automatic and intelligent system, kota kinabalu, sabah, Malaysia.

[13] Mujjadded AL Rabbani Alif, Sabbir Ahmed, Muhammad Abdul Hasan, “Isolated Bangla Handwritten Character Recognition with Convolutional Neural Network”

[14] Karishma Verma, Manjeet Singh, “Hindi handwritten Character recognition using Convolutional neural network”, International journal of computer sciences and Engineering, vol.6, Issue.6, pp.909-914, 2018.