International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Character Recognition using Convolutional Neural Network

Madhuri Yadav[1], Alok Kumar[2]
[1, 2]*USIC&T, GGSIPU*

***Abstract:** Deep learning has provided solution to many pattern recognition problems. Convolutional neural network architecture based on deep learning is proposed in this work for handwritten character recognition. The learning of Convolutional neural network (CNN) is called End-to-End learning. This paper highlights the differences between end-to-end learning and traditional learning. A CNN architecture giving an accuracy of 86.7% for Hindi handwritten characters is proposed.*
***Keywords:** Deep learning; CNN; Handwritten characters; Character recognition; Convolutional layer.*

I. INTRODUCTION

Automatic character recognition is a process that converts scanned document images into electronically understandable format. Thus, enabling computers to recognize text present in images. The latest advancements in technology have highlighted the need for robust methods of automatic character recognition. There are techniques which have been implemented for hindi character recognition as discussed in next section but there was a need of more complete and modern architecture for recognition. Thus, this article uses deep learning concepts for character recognition.

Artificial intelligence rapidly tackled and solved problems that can be described formally, or by mathematical formulas, but as the technology progressed, people tried to automate their work and solve those problems which were easy to do for humans but difficult to describe formally or mathematically like optical character recognition, speech recognition etc. The solution for this problem was to allow computers to learn from experience and learn in hierarchy or layer by layer. Computers needed to learn from their own experience without much need of human knowledge. The hierarchical learning enables computer to learn complicated concepts using simpler ones. If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers. For this reason, this approach is called deep learning. Deep learning is based on the concepts of artificial intelligence in which a model learns to perform classification task directly from images. Deep refers to the number of layers. Traditional neural networks have two or three layers while deep networks can have more number of layers depending upon the dataset and required architectures.

Many artificial intelligence tasks can be solved by identifying the right set of features, and then providing these features to classifier. For example, estimating the size of speaker's vocal tract is a useful feature for speaker identification, estimating pressure points and pen up and down movements are useful feature for online handwriting recognition. However, for many tasks, it is difficult to identify the right set of features. The solution to this problem is deep learning, also called end-to end learning. It is called end-to-end learning because feature extraction and classification phase is automatically done, unlike traditional machine learning, where features are to be explicitly specified. Deep architectures have provided to solutions to some well known problems of pattern recognition which are mental load classification [8], speech recognition[9], document recognition[11], object detection, scene classification[10], pedestrian detection [12] etc.

The working of deep networks is based on connectionism as shown in Fig. 1. The larger the number of neurons and their connections higher will be the learning. Convolutional networks [1] are deep networks with large number of neurons and layers. They work by extracting simpler features in their initial layers and learn complex features layer by layer. It is important to emphasize that number of neurons and size of network should be large.

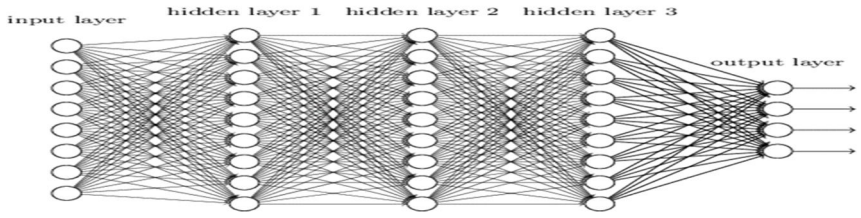

Fig. 1 Basic architecture of deep neural network

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

II. PROPOSED WORK

Convolutional neural network (CNN) are specialized type of neural network with automatic feature extraction. Some of the famous architectures are Alexnet[3], ZfNet[4], VGGNet[5], GoogleNet[6], MicrosoftResNet[7]. The basic architecture of CNN given bt LeCun et,al. [1] is shown in Fig. 2

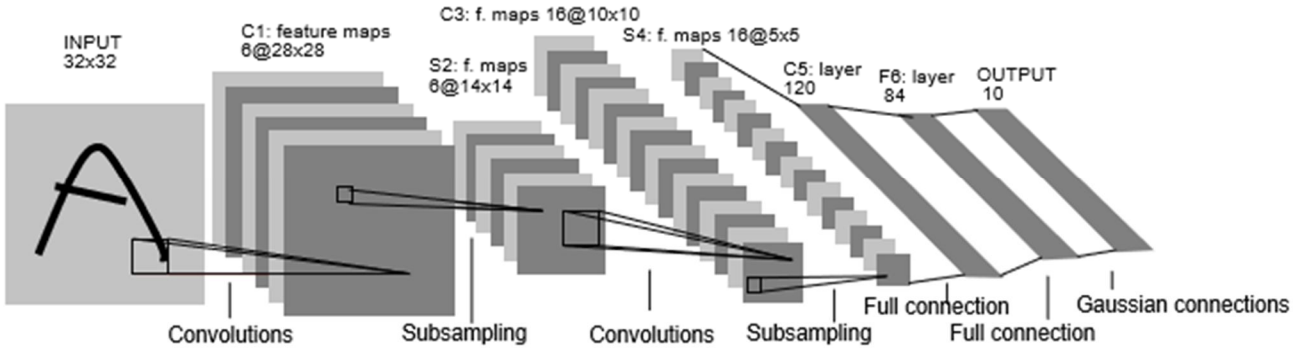

Fig. 2 CNN architecture given by [2]

The figure shows different types of layers present in CNN: The first layer i.e. Input layer, second layer is called Convolutional layer, Intermediate layers are pooling layers (subsampling) and convolutional layers, second last layer of the network is fully connected layer, and finally, the output layer.

A. Input Layer

The input layer is the layer of images which is fed as input to the network. In this work, the input is the Hindi character image as shown in Fig. 3. The input layer can be a grayscale image or colored image (RGB values). Depending upon the type of input image, the input layer can have dimension WxHxD. WxH is the width and height of the image and D is the depth of image which is 1 for grayscale images and 3 for RGB images. Thus, the input layer has dimension 32x32x1.

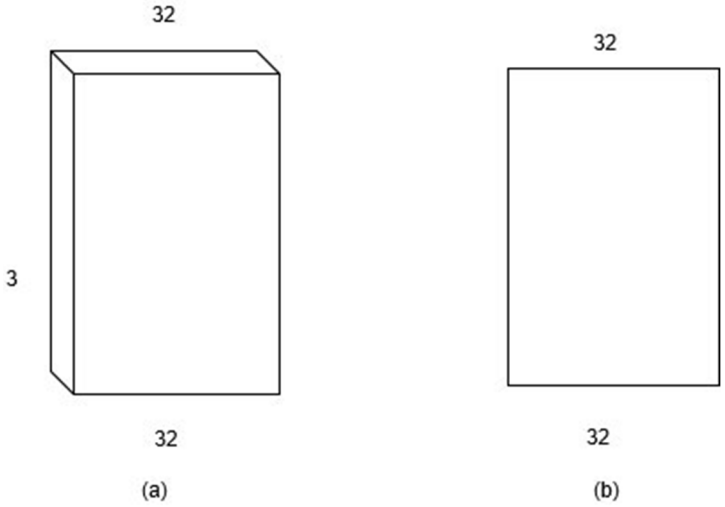

Fig. 3. Input images (a) Input layer image dimensions for RGB image (b) Input layer image dimensions for grayscale image

B. Convolution Layer

This is the building block of entire network because most of the computational work is done in this layer. The convolutional layer consists of a set of learnable filters which are called parameters of this layer. Every filter is a square matrix of small width and height spatially but extends through the full depth of the input volume. For example, a typical filter on a first layer of a network can have size might size 5x5x3 (i.e. 5 pixels width and height, and 3 because images have depth 3, the color channels. In our case, filters will have depth of 1, thus size 5x5x1). During the forward pass, each filter is slided or convolved across the width and height of the

input volume and compute dot products between the entries of the filter and the input at any position. As filter is convolved over the width and height of the input volume, a 2-dimensional activation map is generated which gives the responses of that filter at every spatial position. Intuitively, the network will learn filters that activate when they see some type of visual feature such as an edge of some orientation or at horizontal gradients etc.

In this way, each filter is convolved over the entire image and the generated output after convolution are called activation maps as shown in Fig. 4. The size and number of filters depends upon the experimental rules. There is no well-defined procedure to identify the size and number of filters. Initially, filters can contain any small random values as they are learnable parameters and their values will be updated with each learning of the network.

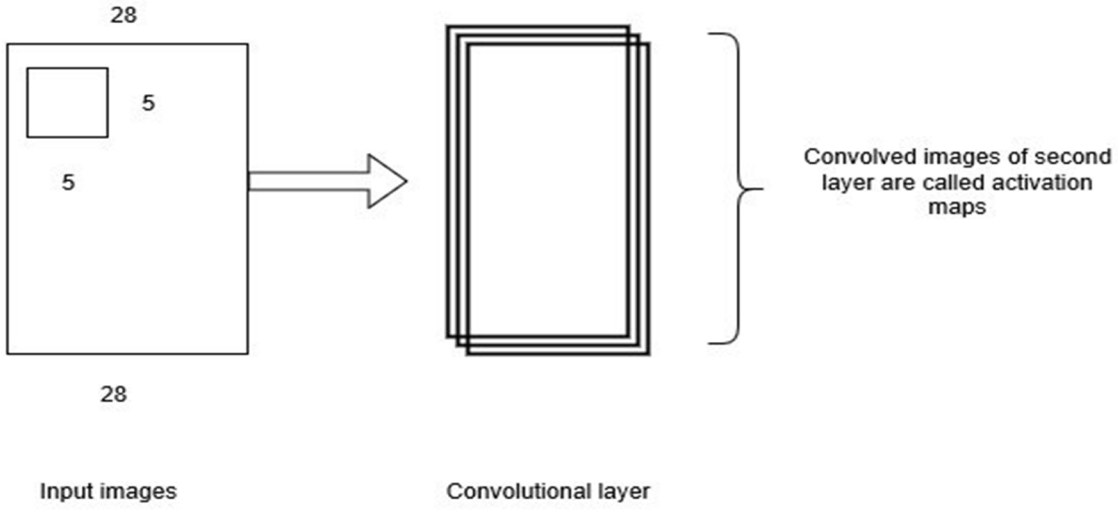

Fig. 4 Activation maps of second layer

The layer accepts input of dimension WxDxH, and using two hyperparameters i.e. filter size(F) and stride(S), generates input for another layer with dimensions $W_1xH_1xD_1$ where W_1 and H_1 are given by the equation (1) and (2) respectively. Depth remains same i.e. D_1=D. Here, P is padding. It introduces new row and column of zeros on each side of image.

$$W1 = (W - F + 2P)/S + 1 \quad (1)$$

$$H1 = (H - F + 2P)/S + 1 \quad (2)$$

In the proposed work, the number of filters was chosen as 32 of size 5x5x1, P=0, and S=1. Thus, the dimensions of second layer image become 28x28x32 according to equation (1) and (2).

C. Pooling Layer

Pooling layers [2] are placed in-between convolutional layers in a convolutional architecture. Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network, and hence to also control overfitting. The Pooling Layer operates independently on every depth slice of the input and resizes it spatially. Commonly used operation is MAX. The most common form is a pooling layer with filters of size 2x2 applied with a stride of 2 down-samples every depth slice in the input by 2 along both width and height, discarding 75% of the activations. Every MAX operation would select a maximum value over 4 numbers (little 2x2 region in some depth slice). The depth dimension remains unchanged.

In the proposed work, in first pooling layer max filter of size is 3x3, P=1 and stride of 2 is used. Thus, the output dimensions of this layer will remain same as 14x14x32. Other filters such as averaging, min, L2 norm pooling can be used in pooling layer.

D. Fully connected Layer

Neurons in a fully connected layer have full connections to all activations in the previous layer, as in regular Neural Networks. Their activations can hence be computed with a matrix multiplication followed by a bias offset. There can be multiple fully connected layer depending upon the application architecture. The last contains neurons equivalent to the number of classes in problem domain.

In this work, there are 41 character classes; hence the last layer has 41 neurons. The second last fully connected layer has 256 neurons. The number of neurons in this layer is chosen experimentally. Fig. 5 displays the complete architecture as proposed in this work.

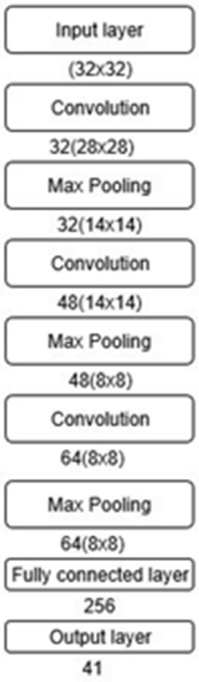

Fig. 5 CNN architecture proposed in this work.

III. DIFFERENCE BETWEEN TRADITIONAL AND DEEP LEARNING

Traditional learning or machine learning follows basic pipeline of pre-processing, feature extraction, classification and post-processing (optional). It requires humans to correctly identify the viable set of features for their problem. In contrary, deep learning such as CNN does not depend upon human for feature extraction. The network learns by backpropogation learning by training parameters. Deep learning has solved various complex problems where machine learning algorithms fail. The great power of problem solving of deep techniques comes with few lapses. Techniques working on deep architectures require high performance hardware and graphics processing units (GPU). They demand high computational power and memory. Deep architectures give good results with large amount of data.

IV. CONCLUSIONS

Convolutional neural networks have provided solution to almost every pattern recognition problem. They were among the first working deep networks trained with back-propagation. It is not entirely clear why convolutional networks succeeded when general back-propagation algorithms fail. It may simply be that convolutional networks work in hierarchy and solve complex structures by simpler ones. This work proposed convolutional network architecture for Hindi handwritten characters. The CNN architecture has thousands of parameters and hyper-parameters to tune. The authors would like to optimize the proposed CNN architecture in near future.

REFERENCES

[1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, "Backpropagation applied to handwritten zip code recognition", Neural computation, vol. 1, no. 4, pp. 541-551, 1989.
[2] A. Hyvarinen, U. Kouster: "Complex cell pooling and the statistics of natural images Network", Journal of Network: Computation in Neural Systems, vol. 18, no. 2, pp. 81–100, 2007.
[3] Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", C. J. C. Burges, L. Bottou, et al., (Eds): In Advances in Neural Information Processing Systems , pp. 1097–1105, 2012.
[4] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks", CoRRabs/1311.2901, 2013.
[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", CoRR abs/1409.1556, 2014.

[6] C. Szegedy, W. Liu, Y. Jia, et al.,"Going deeper with convolutions", International conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA,pp. 1–9, June 2015.
[7] K. He, X. Zhang, S. Ren, et al., "Deep residual learning for image recognition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, pp.770–778, June 2016.
[8] Zhicheng Jiao, Xinbo Gao, Ying Wang, Jie Li, Haojun Xu, "Deep Convolutional Neural Networks for mental load classification based on EEG data", Pattern Recognition, vol. 76, pp. 582–595, 2018.
[9] G. Hinton, L. Deng, D. Yu, G.E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T.N. Sainath,"Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups", IEEE Signal Process. Mag., vol. 29, pp. 82–97, 2012.
[10] Y. Yuan, J. Wan, Q. Wang, "Congested scene classification via efficient unsupervised feature learning and density estimation", Pattern Recognit., vol. 56, pp.159–169, 2016.
[11] Xiao-Xiao Niu, Ching Y. Suen, "A novel hybrid CNN and SVM classifier for recognizing handwritten digits", Pattern Recognition, vol. 45, no. 4, pp. 1318–1325, 2012.
[12] W. Ouyang, X. Wang, "Joint deep learning for pedestrian detection", International Conference on Computer Vision, Sydney, Australia, pp. 2056–2063, Dec 2013.