2018 International Conference on Information, Communication, Engineering and Technology (ICICET)
Zeal College of Engineering and Research, Narhe, Pune, India. Aug 29-31, 2018

Handwritten Devanagari Character Classification using Deep Learning.

Prasad K. Sonawane
Dept. of Electronics and Telecomm.
College of Engineering, Pune
Pune, India
prasadsonawane1992@gmail.com

Sushama Shelke
Dr. D. Y. Patil Institute of Engineering,
Management and Research,
Pune, India
shelke.sushama@gmail.com

***Abstract*—Since past few years, deep neural networks, because of their outstanding performance, is getting highly used in computer vision and machine learning tasks such as regression, segmentation, classification, detection, pattern recognition etc. Recognition of handwritten Devanagari characters is challenging task, but Deep learning can be effectively used as a solution for various such problems [1]. Person to person variations in writing style makes handwritten character recognition one of the most difficult tasks. In this Experiment, we successfully tried to classify handwritten Devanagari characters using transfer learning mechanism with the help of Alexnet. Alexnet, a convolutional neural network, is trained over a dataset of around 16870 samples of 22 consonants of Devanagari script which shows impressive results. The transfer learning helps to learn faster and better even if the data samples are less as compared with the training a CNN from scratch.**

***Index Terms*—AlexNet, CNN, Transfer Learning**

I. Introduction

Handwritten character recognition is a challenging task due to so many reasons and few are, writing style variations, a lack of ordering information of strokes, overlapping of wide strokes, pen-type and also the large similarities between the characters [2]. Even though it is one of the most difficult tasks, it has many useful applications like postal zip code detection [3], and recognition, signature verification, converting handwritten documents into digital documents. Because of numerous such tasks, researchers are more attracted towards handwritten character recognition.

There are many traditional methods to recognize handwritten characters such as Fuzzy based classification [4], Hidden Markov Model [2], Support Vector Machine. In these conventional methods, one has to perform preprocessing and extraction of features from the data and then using those extracted features, a classifier is trained. After that the trained classifier is used to classify the testing data into number of classes. The accuracy of these conventional methods depends highly on the features that have been selected for training. For printed character recognition [5], it is easy to calculate distinct features from the data, but it gets more difficult for handwritten data. Some of the features are chain codes, Discrete wavelet transform coefficient, Discrete Fourier Transform coefficient, Gabor wavelet features, Morphological features. Sometimes, it becomes difficult to select appropriate features which makes traditional methods less robust.

From the early 2000s, Deep Neural Networks are used for object detection, segmentation and recognition in image processing field with immense success. It is also widely used for many other pattern recognition problems such as handwritten digits and characters recognition, speech recognition, drug discovery and genomics [1].

Deep neural networks make use of a common property of many natural signals, that the signals are composed of hierarchies, where higher level structures can be formed by combining two or more lower level structures. For example, in Image processing, small edges with different orientation and position form contours and different contours form objects. A Deep neural network tries to find small edges first and then their combination at next levels to detect more complex structures and from those structures finally the objects are detected [1]. Deep learning is also helping to solve many problems which were faced by artificial intelligence community. Moreover, its use in natural language processing has shown tremendous success [6].

The main big advantage of Deep neural network is that, it requires less engineering by hand. It learns the features from the applied data on its own using different algorithms. The learning gets better and better with the increase in data. It also requires variations in data. If we provide the samples with variations in data, it will learn those variations, else will misclassify the new samples.

A convolutional neural network is also used for document analysis purpose. Many researchers have used Convolutional neural network for printed as well as handwritten characters and word segmentation recognition for Tamil [7], Chinese [8], Bangla [9] and many more. The Convolutional neural network can also be used to recognize handwritten Devanagari character recognition [10]. In this paper, the pretrained convolutional neural network is trained for handwritten Devanagari character recognition. this technique is called transfer learning [11]. Both Convolutional neural network and transfer learning is discussed in further sections.

II. DEVANAGARI

A. Devanagari Script

The epigraphical evidences show that there are roots of Devanagari script present in the ancient Brahmi script family. It is also called as Nagari. Hindi, Marathi, Nepali and also many languages of Asia uses Devanagari script with some modification for writing purpose. The Devanagari script has 47 primary characters out of which 33 are consonants and 14 are vowels. Fig.1 and Fig.2 shows Devanagari vowels and consonants respectively.

अ	आ	इ	ई	उ	ऊ	ए	ऐ	ओ	औ	अं	अः

Fig. 1. Vowels of Devanagari

क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट	ठ
ड	ढ	ण	त	थ	द	ध	न	प	फ	ब	भ
म	य	र	ल	व	स	ष	श	ह	क्ष	त्र	ज्ञ

Fig. 2. Consonants of Devanagari

B. Handwritten Devanagari Dataset

The required Devanagari dataset is collected by scanning documents written by different individuals from different fields. The different characters are then cropped manually. each character image is of 220x220 pixels. The character can be of any size and can be at any position in the image that means it is not at the center of the Image. Each Image is Grey-Scale image in which the character is represented by grey values and the background is white. The database contains 16870 images of 22 more frequently used consonants. Fig.3 shows some handwritten Devanagari characters and variations in its writing style. In that figure, two Marathi characters tha and gha each written in three different ways by three different people are shown.

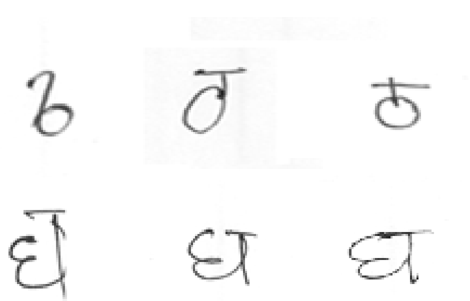

Fig. 3. Handwritten Devanagari characters with variations in writing style.

III. METHODS

A. Convolutional Neural Network

Convolutional neural network is used to process the data which occurs in the form of multiple arrays. For example, A grey-scale Image has 2-dimensional data and a color image has 3-dimensional. There are four key ideas behind the success of convolutional neural network. These are: multiple layers, shared weights, local connection and pooling.

There are mainly three types of layers in convolutional neural network which can be used number of times in a network.

1) Convolution Layer: This is the main building block of the network. It uses number of filters called filter banks and it convolves input image with it. The training steps updates filter weights which gives feature maps from input images.

2) Pooling Layer: When an input image is convolved with the number of filters in convolution layer, it generates stack of images. In this process, the data increases by huge amount. The pooling layer helps to reduce the data easily. It divides newly generated images into number of blocks and subsamples it to produce single value for each block. Max pooling and average pooling are the types. The combination of convolution layers and pooling layers can be regarded as feature extractor.

3) Fully-connected Layer: In fully-connected layer, each neuron takes input from the activations of every neuron of the previous layer. The fully-connected layer can be regarded as classifier.

B. Transfer Learning

In transfer learning, the weights of the model which is trained to classify different data is used as initial weights for training the target data. Transfer learning gives way better and faster results than training the network from scratch with smaller training set. There are many pretrained neural networks publicly available like AlexNet, GoogleNet, Inception, VGGNet, LeNet which can be used to train to perform new tasks. In ImageNet LSVRC-2010 contest, The AlexNet was successfully used to classify on 1.2 million high-resolution images into 1000 classes. The AlexNet containes 650,000 neurons, 60 million parameters. There are 5 convolution layers. After some convolution layers max pool layers are also present to reduce data size. At the end of the network, three fullyconnected layers are used with soft-max layer which has output size of 1000. AlexNet is trained on the ImageNet dataset of centered color images. The ImageNet dataset is downsampled to 227x227x3 and then used to train the network. For Devanagari handwritten character recognition, the pretrained Alexnet is used to classify 22 different consonants. AlexNet requires color images and each image of size 227x227 at its input stage. Our Dataset has Grey-scale images of size 220x220, so to bring an input image to the required size, we added a convolution layer which is having 3-dimensional filter at the start of the network. The last Fully connected layer from the AlexNet is also changed as we are only classifying the data into 22 classes.

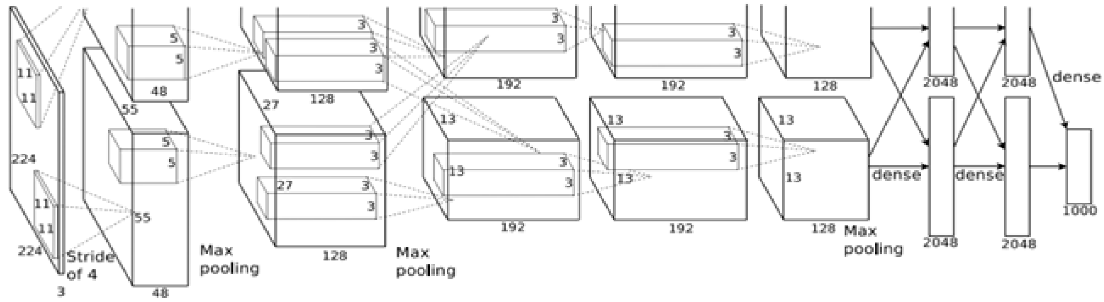

Fig. 4. The detail architecture of AlexNet showing two layers designed for processing on two GPUs. the upper layer is for one GPU and bottom layer is for second GPU [12]

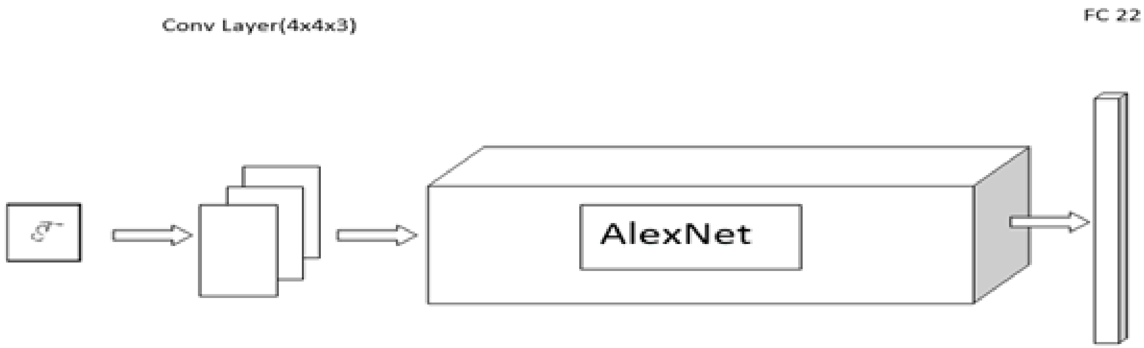

Fig. 5. Architecture of the Convolutional neural network. Where the values shown at conv layer indicates size of the layer and FC 22 means Fully-connected layer of size 22 .

C. Architecture

Fig.4 shows the architecture of AlexNet for 2 GPUs [12] and Fig.5 shows the architecture of the network used. The first convolutional layer, which is not a part of AlexNet, has 3 filters of size 4x4 with strides of [1,1]. The input to this layer is an image with 220x220 pixels resolution. Before feeding this image to the first layer, zero padding of 5 is done from all the four directions. Zero padding will bring the size to 230x230 and when three 4X4 filters are applied, for each image in the dataset we will get output of 227x227x3 from first convolutional layer. Now this output is applied to the AlexNet. The AlexNet has 5 convolutional layers with filters of size 11x11x3 (96), 5x5x48 (256), 3x3x256 (384), 3x3x192 (384), 3x3x192(256) sequentially, where the numbers in the brackets shows the number of filters at that layer. Each convolutional layer uses ReLU, a rectified linear unit, as an activation function.

$$F(x) = max(0, x) \tag{1}$$

The ReLU gives faster training with Stochastic Gradient Descent method as compared to other non-linearities like

$$F(x) = tanh(x) \tag{2}$$

or

$$F(x) = 1/(1 + e^x) \tag{3}$$

After five convolution layers, there are 3 Fully-connected layers in AlexNet of size 4096, 4096 and 1000. Now for transfer learning we need to remove the last fully-connected layer and replace with the other as per requirement. In our case, where the number of classes are 22, we replaced it with the Fully-connected layer of size 22. At the end, SoftMax function is applied to the output of Fully-connected layer to normalize all the values in the range (0,1).

$$Softmax(x) = e^x/sum(e^x) \tag{4}$$

The concept of drop-out is present in last two FC layers. In drop-out, some random neurons get temporarily removed from the network with each iteration. The output of these neurons is set to zero, so they do not take part in forward as well as backward pass. The drop out adds regularization in the fully-connected layer to reduce the over fitting.

IV. Experiment and Results

For this experiment, MATLAB 2017b and Neural Network toolbox is used to implement CNN. To train the network, the dataset is divided into 3 parts (Training 56%, Validation 24%, testing 20%).

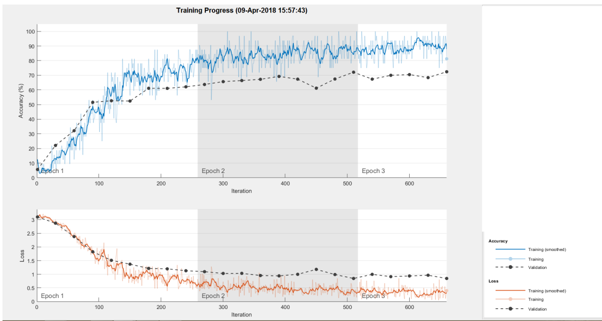

Fig. 6. Training Curves vs epochs. upper window shows accuracies, Bottom window shows loss curve.

Fig. 6 shows training curves in which blue is training accuracy while black dotted curve is validation curve against number of epochs. We can see that, as number of epochs increases accuracy increases and loss decreases. The training and validation loss curves are the decaying curves. Due to MATLABs default stopping criterion on validation, the training terminated when validation reached to 72.46%. when the same trained network is given to training again, the curves are shown in Fig. 7 which shows saturation of accuracy and loss.

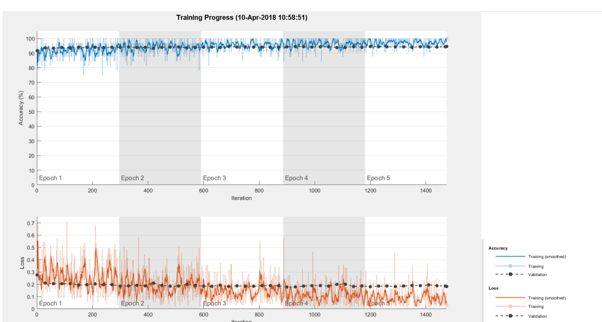

Fig. 7. Training curves when accuracy and loss remains almost constant.

V. Conclusion

This paper presents fine-tuning of CNN with the help of transfer learning. The CNN achieved 94.49% validation accuracy and 95.46% test accuracy. In the training curves, we can see that the training accuracy went above 90% in 3 epochs only. The results show that for less input data and faster and better training the transfer learning is a better option.

References

[1] Y. LeCun, Y. Bengio and G. Hinton , “Deep Learning,” Nature, vol. 521, pp.437- 444, May 2015.

[2] B. Shaw, S. Parui and M. Shridhar, “Ofine Handwritten Devanagari Word Recognition: A holistic approach based on directional chain code feature and HMM,” International Conference on Information Technology, pp.203-208, 2008.

[3] Y. Le Cun et al., “Handwritten ZIP Code Recognition with Multilayer Networks,” Proc. 10th Intl Con$ Pattern Recognition, Vol. 2, IEEE CS Press, Los Alamitos,Calif., Order No. 2063,pp. 35-40, 1990.

[4] S. Shelke and S. Apte, “A Fuzzy based Classification Scheme for Unconstrained Handwritten Devanagari Character Recognition ” International Conference on Communication, Information & Computing Technology (ICCICT), Jan 2015.

[5] T. Karayil,A. Ul-Hasan and T. Breuel, “A Segmentation-Free Approach for Printed Devanagari Script Recognition,” 13th International Conference on Document Analysis and Recognition (ICDAR), pp.946-950,2015.

[6] Collobert, R., et al., “Natural Language Processing (Almost) from Scratch,” Journal of Machine Learning Research 12, pp.249-2537, 2011.

[7] P. Vijayaraghavan, M. Sra, “Handwritten tamil recognition using a convolutional neural network, ”2014.

[8] Yuhao Zhang, “Deep Convolutional Network for Handwritten Chinese Character Recognition, ”Stanford University.

[9] M. Rahman et al., “Bangla Handwritten Character Recognition using Convolutional Neural Network, ” I.J. Image, Graphics and Signal Processing,8, pp. 42-49 , July 2015.

[10] S. Acharya, A. Pant, P. Gyawali, “Deep Learning Based Large Scale Handwritten Devanagari Character Recognition, ”9th International Conference on Software Knowledge Information Management and Applications (SKIMA), 2015.

[11] Y. Tang, L. Peng, Q. Xu, Y. Wang, A. Furuhata, “CNN based transfer learning for historical Chinese character recognition ” IAPR Workshop on Document Analysis Systems, pp. 25-29, 2016.

[12] A. Krizhevsky, I. Sutskever, G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks, ” Proc. Neural Information and Processing Systems, 2012.