Research Paper
Impact Factor: 6.304
Peer Reviewed & Indexed Journal
www.ijbarr.com

IJBARR
E- ISSN -2347-856X
ISSN -2348-0653

AN ACCURATE DEEP-LEARNING MODEL FOR HANDWRITTENDEVANAGARI CHARACTER RECOGNITION

Anuj Bhardwaj*   Prof.(Dr.) Ravendra Singh**

**Research scholar, Dept. of CS & IT, M.J.P. Rohilkhand University, Bareilly (U.P.), India.*
***Professor Dept. of CS & IT, M.J.P. Rohilkhand University, Bareilly (U.P.), India.*

Abstract

Handwritten character recognition is increasingly important in a variety of automation areas, for example, authentication of bank signatures, identification of ZIP codes on letter addresses, and forensic evidence, etc. Handwritten character recognition is the process where a machine detects and recognizes characters from a text image and converts that processed data into a code that is understood by the machine. This is a fundamental but challenging task in the field of pattern recognition. In this paper, we have used a new public image dataset for the Devanagari script character: the Devanagari character dataset (DCD). This considered dataset contains 92 thousand images of 46 different classes of characters in the Devanagari script, fragmented from handwritten documents. The paper also explores the challenges faced in the identification of Devanagari characters. This paper proposes a deep learning based convolutional neural network (CNN) architecture for the recognition of those handwritten characters in an unrestricted environment, with datasets. Deep convolutional neural networks have shown better results than traditional shallow networks in many recognition tasks. While keeping distance from the routine approach of character recognition by Deep CNN, wefocus on the use of dropout and dataset augmentation approaches to improve test accuracy. By applying these techniques to Deep CNN, we were able to increase the test accuracy to about 0.98 percent. The proposed architecture achieved the highest test accuracy of 98.13% on the considered dataset. The results indicate that the proposed model may be a strong candidate for handwritten character recognition and automatic handwritten Devanagari script character recognition applications.

Keywords: Devnagari script character, Handwritten Devanagari Character, Deep Learning.

Introduction

Handwritten character recognition is considered to be one of the most challenging and appealing research areas in the field of pattern recognition and computer vision. It is the natural way of interacting with the computer. (Acharya et al., 2015) Due to the critical factors of differences in writing patterns and cursive text, and the similarity of various characters in Devanagari script characters, recognition research is time-consuming and challenging. It has been a field of great interest for researchers and scientists. Character recognition is the process where the machine detects and recognizes the characters from a text image and converts that processed data into a code which is understood by the machine. It is a fundamental yet challenging task in the field of pattern recognition. Recognition of handwritten Devanagari script characters may be performed online or offline. Online character identification is relatively simple due to the temporal-based character properties such as form, number of strokes, distance, and direction of writing. Offline character recognition implementation is complex due to variations of writers and fonts. Recognition is termed as optical character recognition (OCR) as they deal with characters which are optically processed and not magnetically processed.(Sinha, 2009) The literature shows a high accuracy rate for recognition of characters and isolated words in optical character recognition (OCR) or printed text; however, there is a need for a competent handwritten character recognition system capable of generating a high degree of accuracy in handwritten text recognition.(Fujisawa, 2008) Character classification is an important part of handwritten character recognition that plays an essential role in many computer vision problems like OCR, license Plate recognition, etc. Development of a recognition system is an emerging need for digitizing handwritten hindi documents that use Devnagari characters. Optical Character Recognition systems are least explored for Devnagari characters. (Arica & Yarman-Vural, 2001) present a few approaches for segmentation and recognition of Devnagari characters. The major challenging task while creating a recognizer and classifier for the Devnagari scripts is that they have a large number of symbols as compared to languages like English. English contains mainly 52 characters while a script like Devanagari consists of more than 200 symbols. Almost all the Indic languages are similar to each other. Some of them have a characteristic of allowing characters to be combined together to form another character, generally referred to as 'sayuktakshar' where Sayukt stands for combined and 'akshar' means word. Another challenge is the identification of vowel modifiers or 'matras'. Vowel modifiers change the UNICODE value of that character, and hence, in order to identify the correct character and its Unicode value, we need to write script specific rules. These modifiers may be present at the left, right, bottom or top of the character. In some of the cases, the modifier may be present at two positions on the same character. Hence, it is required to identify these modifiers correctly to reduce the error in classification. So this paper considering a publicly available Devnagari Character Dataset, of 92 thousand images of 46 Devnagari characters. Then, this paper proposing a Deep learning Convolutional Neural Network model to classify the characters in DCD. Introduction of multilayer perceptron network has been a milestone in many classification tasks in computer vision. But, performance of such a network has always been greatly dependent on the selection of good representing features.(Ruck et al., 1990)(Yang et al., 2009) Deep Neural Networks on the other hand do not require any feature to be explicitly defined, instead they work on the raw pixel data generating the best features and using it to classify the inputs into different classes.(Lee et al., 2009) Deep Neural networks consist of multiple nonlinear hidden layers and so the number of connections and trainable parameters are very large. Besides being very hard to train, such networks also require a very large set of examples to prevent overfitting. One class of DNN with comparatively smaller set of parameters and easier to train is Convolutional Neural Network. The ability of CNN to correctly model the input dataset can be varied by changing the number of hidden layers and the trainable parameters in each layer and they also make correct assumption on the nature of images.(Krizhevsky et al., 2012) Like a standard feed forward network, they can model complex non-linear relationship between input and output. But CNN have very few trainable parameters than a fully connected feed-forward network of same depth. CNNs introduce the concept of local receptive field, weight replication and temporal subsampling which provide some degree of shift and distortion invariance. CNNs for image processing generally are formed of many convolution and sub-sampling layers between input and output layer. These layers are followed by fully connected layers thereby generating distinctive representation of the input data. Beside image recognition, CNNs have also been used for speech recognition.(Abdel-Hamid et al., 2012)(Sainath et al., 2013) Although deep convolutional neural networks have a small and inexpensive architecture compared to standard feed forward network of same depth, training a CNN still requires a lot of computation and a large labeled dataset. Training such a network was not so effective and did not produce any superior result to traditional shallow network, until recently. With the availability of large labeled dataset like IMAGENET, NIST, SVHN, development of state of the art GPUs and introduction of unsupervised pre-training phase, CNNs have at present proven to surpass traditional feed forward network in a number of classification tasks. In CNNs, initializing the weight randomly and applying gradient descent and backpropagation to update the weights seems to generate poorer solution for a deep network. So, generally, greedy layer wise unsupervised pre training is applied prior to supervise training. Why such unsupervised training helps is investigated in (Erhan et al., 2010).

The available literature shows unsatisfactory research results compared to other languages. In (Decerbo et al., 2004), the Byblos Pashto OCR system was proposed for script-free OCR using HMMs. This system was also subsequently tested for Chinese, English, and Arabic text with success. As previously mentioned, Devanagari Intelligent character recognition and Optical character recognition (OCR) area are the least explored to date. Based on this discussion, it can be concluded that there is a lack of a Devanagari handwritten character dataset. In addition, a research gap exists for classification of Devanagari handwritten characters based on deep learning techniques, such as the CNN. The current research fills this gap by proposing a CNN model and a Devanagari handwritten character dataset. For this purpose, a model was developed and extensive experimentation conducted.

Devanagari Handwritten Character Dataset Devanagari Script

Devanagari is part of the Brahmic family of scripts of Nepal, India, Tibet, and South-East Asia. The script is used to write Nepali, Hindi, Marathi and similar other languages of South and East Asia. The Nepalese writing system adopted from Devanagari script consists of 12 vowels, 36 base forms of consonant, 10 numeral characters and some special characters.

Devanagari Handwritten Character Dataset

Devanagari Handwritten Character Dataset is created by collecting the variety of handwritten Devanagari characters from different individuals from diverse fields. Handwritten documents are than scanned and cropped manually for individual characters. Each character sample is 32x32 pixels and the actual character is centered within 28x28 pixels. Padding of 0 valued 2 pixels is done on all four side to make this increment in image size. The images were applied gray-scale conversion. After this the intensity of the images were inverted making the character white on the dark background. To make uniformity in the background for all the images, we suppressed the background to 0 value pixel. Each image is a gray-scale image having background value as 0.

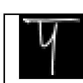
Challenges in Devanagari Character Recognition

There are many pairs in Devnagari script that has similar structure differentiating each with structure like dots, horizontal line etc. Some of the examples are illustrated in Table1. The problem becomes more intense due to unconstrained cursive nature of writings of individuals. Two such examples are shown in Table 2.

Table 1: Structural formation of characters

छ	६	Difference being horizontal line at top
ड	ड़	Difference being presence of single dot on right side
द	ढ	Difference being presence of small circle and small down stroke line

Table 2: Different characters written similarly

	प		य
	ध		घ

Proposed Model for Devanagari handwritten character recognition

This section proposed a means of Devanagari handwritten character classification and recognition model as shown in Figure 1. This recognition model relies on a CNN applied for the Devanagari handwritten character dataset with a feature mapped output layer. This suggested CNN model classifies Devanagari characters into 46 different classes. A detailed explanation of the suggested model is presented in the following subsections.

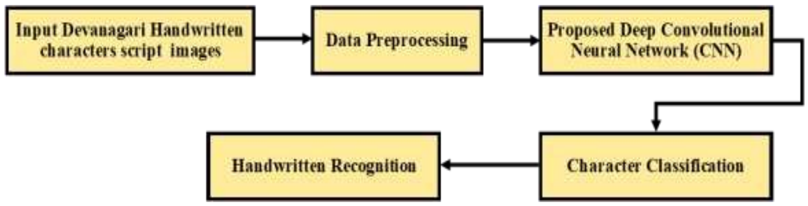


Figure 1: Block diagram of proposed model framework

Devanagari handwritten characters scripts images dataset

A data-enriched dataset plays a vital role in the generation of accurate results in research activities related to deep learning. A concise and precise dataset is required for a true evaluation of mathematical models that are applied to it. Moreover, to achieve benchmark results in deep learning, a standard publicly available dataset is mandatory. During the experimental phase of this research, we considered dataset named Devanagari Handwritten Character Dataset publically available on UCI repository.[1] Detailed description about the data is described in previous section.

Data Preprocessing

Preprocessing steps were applied to the proposed dataset to prepare images for subsequent phases. Image preprocessing consisted of operations on images at the lower abstraction level with the aim of improving the image data. This improvement suppresses undesired distortions in the image dataset, or enhances important details or features that are essential for further processing.

The first step of the preprocessing phase was the removal of noise from images using Gaussian blur. The digital scanner induced spikes of noise into the scanned images, as shown in Figure 2(a). These spikes of noise were removed using the notable work of Soman.(Ganga Gowri & Soman, 2018) The second step was smoothing of the image using the Gaussian function. This can be considered to be low-pass filtering in a non-uniform manner, which conserves the low frequency, and decreases the noise and insignificant details in an image. This was accomplished by convolving the Gaussian kernel with an image.

Figure 2: Image dataset denoising steps: (a) scanner-induced noise; (b) denoised image; (c) Binarized Image; (d) application of Gaussian blur filter

Convolutional Neural Networks

Convolutional Neural Network (CNN or ConvNet) is a biologically inspired trainable machine leaning architecture that can learn from experiences like standard multilayer neural networks. ConvNets consist of multiple layers of overlapped tiling collections of small neurons to achieve better

Research Paper
Impact Factor: 6.304
Peer Reviewed & Indexed Journal
www.ijbarr.com

IJBARR
E- ISSN -2347-856X
ISSN -2348-0653

representation of the original image. ConvNets are widely used for image and video recognition. There are three main types of layers used to build a ConvNet architecture.

1. **Convolution Layer:** The convolution layer is the core building block of a convolutional neural network. It convolves the input image with a set of learnable filters or weights, each producing one feature map in the output image.
2. **Pooling Layer:** The pooling layer is used to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network. The pooling layer takes small rectangular blocks from the convolution layer and subsamples it to produce a single output from that block. There are several ways to do this pooling, such as taking the average or the maximum, or a learned linear combination of the neurons in the block.
3. **Fully-Connected Layer:** The fully-connected layer is used for the high-level reasoning in the neural network. It takes allneurons in the previous layer and connects it to every single neuron it has. Their activations can be computed with a matrix multiplication followed by a bias offset as a standard neural networks.

A simple convolutional neural network similar to the one used in our recognition system is shown in Figure 3. The input layer consists of the raw pixel values from the 32x32 grayscale image and has no trainable parameters. The first convolution layer has 4 feature maps with 784 units/neurons each (28 x 28). Each feature map is shown in figure as 2D planes and they have different set of weights. All the units in a feature map share the same set of weights and so they are activated by the same features at different locations. This weight sharing not only provides invariance to local shift in feature position but also reduces the true number of trainable parameters at each layer. Each unit in a layer receives its input from a small neighborhood at same position of previous layer. So the number of trainable weights associated with each unit in a convolutional layer depends on the chosen size of the neighborhood of previous layer mapped to that unit. Since all the units are activated only from the input taken from a local neighborhood they detect local features such as corners, edges, end-points. This concept of local receptive field is inspired from study of the locally-sensitive, orientation selective neurons in the cats visual system.(Hubel & Wiesel, 1962)

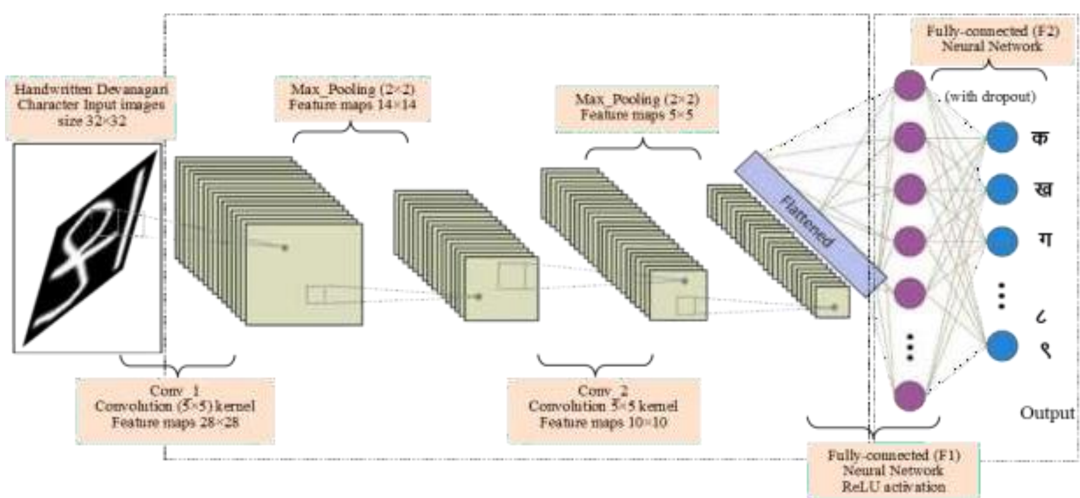


Figure 3: Proposed Convolutional Neural Network

IJBARR
E- ISSN -2347-856X
ISSN -2348-0653

For a 5x5 kernel as shown in Figure 3 the number of input weights for each unit is 25. In addition the units also have a trainable bias. The total number of units in a layer depends upon the size of kernel in the previous layer and overlap between the kernels.

Experiments and Result

We tested the dataset with different architectures by varying depth, width and number of parameters of network. The results of two of those experiments are presented in the coming sections. The first model is very wide and deep and consists of a large number of parameters. It will be referred to as model A in the coming section. It consists of three convolution layers and one fully connected layer. The sequence of the layers in model A is shown in Figure 4., where model consist of three convolution layer, three Rectified Linear Unit Layer, two Normalization layer implementing Local Response Normalization, two pooling layer implementing max pooling, one Dropout layer and one Fully Connected Layer and a accuracy layer for test set and a Softmax Loss layer that computes multinomial logistic loss of the softmax of its input.

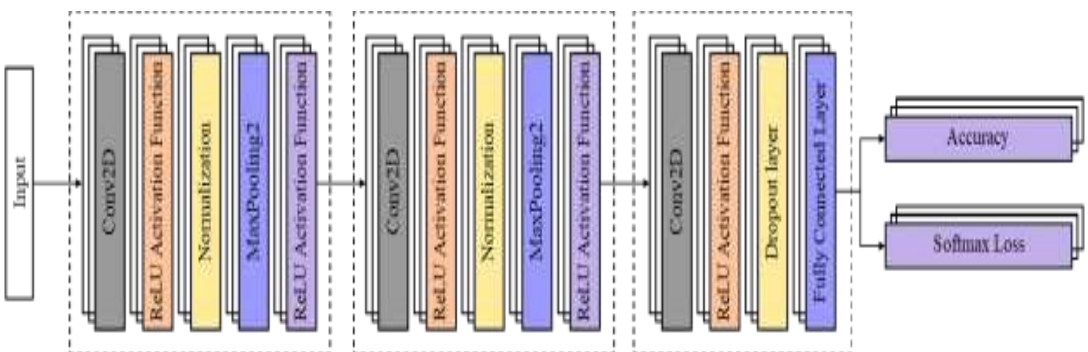


Figure 4: Architecture of model A- deep CNN model for handwritten character recognition

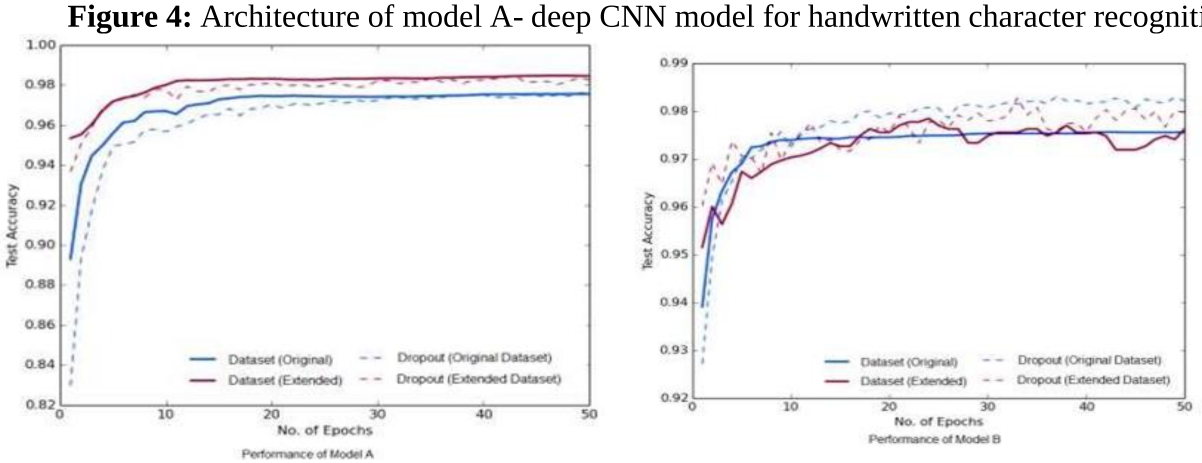


Figure 6: Accuracy testing of the models

The result of training for 50 epoch is presented in Figure 6. Test Accuracy remained nearly constant after 50 epochs. For model A, Extending Dataset showed superior result in Test Accuracy. So, increasing number of training sample is effective to increase performance of wide and deep network with large bank of parameters. The highest testing accuracy obtained for Model A is 0.98471. For model B, addition of dropout showed better improvement in Test accuracy. However, extending dataset also resulted slight improvement in Test accuracy. The highest value of Testing Accuracy obtained for this model is 0.981326.

Conclusion

The presented research work proposed a deep learning CNN model that provide a high accuracy rate ≈ 98.13% to recognizing handwritten Devnagari Character, the model is trained using the publically available Devnagari Character Dataset for any researcher. It consists 92 thousand images of 46 different characters of Devnagari script. The proposed model explored the challenges in classification of characters in Devnagari Dataset. The challenges result due to the fact that the dataset consists many characters that are visually similar or written in a similar way by most people. Also, In Devnagari script, the base form of consonant characters can be combined with vowels to form additional characters which is not explored in this research. For recognition, we proposed two deep learning models to train the dataset. We also analyzed the effect of dropout layer and dataset increment to prevent overfitting of these networks. The experimental results suggested that Deep CNNs with added Dropout layer and Dataset increment technique can result in very high test accuracy even for a diverse and challenging dataset like ours.

References

1. Acharya, S., Pant, A. K., & Gyawali, P. K. (2015). Deep learning based large scale handwritten Devanagari character recognition. 2015 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), 1–6.
2. Sinha, R. M. K. (2009). A journey from Indian scripts processing to Indian language processing. IEEE Annals of the History of Computing, 31(1), 8–31.
3. Fujisawa, H. (2008). Forty years of research in character and document recognition—an industrial perspective. Pattern Recognition, 41(8), 2435–2446.
4. Arica, N., & Yarman-Vural, F. T. (2001). An overview of character recognition focused on off-line handwriting. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 31(2), 216–233.
5. Ruck, D. W., Rogers, S. K., & Kabrisky, M. (1990). Feature selection using a multilayer perceptron. Journal of Neural Network Computing, 2(2), 40–48.
6. Yang, J.-B., Shen, K.-Q., Ong, C.-J., & Li, X.-P. (2009). Feature selection for MLP neural network: The use of random permutation of probabilistic outputs. IEEE Transactions on Neural Networks, 20(12), 1911–1922.
7. Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2009). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Proceedings of the 26th Annual International Conference on Machine Learning, 609–616.
8. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 1097–1105.
9. Abdel-Hamid, O., Mohamed, A., Jiang, H., & Penn, G. (2012). Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4277–4280.
10. Sainath, T. N., Mohamed, A., Kingsbury, B., & Ramabhadran, B. (2013). Deep convolutional neural networks for LVCSR. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 8614–8618.
11. Erhan, D., Courville, A., Bengio, Y., & Vincent, P. (2010). Why does unsupervised pre-training help deep learning? Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 201–208.
12. Decerbo, M., MacRostie, E., & Natarajan, P. (2004). The bbn byblos pashto ocr system. Proceedings of the 1st ACM Workshop on Hardcopy Document Processing, 29–32.