DESIDOC Journal of Library & Information Technology, Vol. 40, No. 5, Sept 2020, pp. 268-271, DOI : 10.14429/djlit.40.5.16336

Recognising Devanagari Script by Deep Structure Learning of Image Quadrants

Seba Susan* and Jatin Malhotra

Department of Information Technology, Delhi Technological University, Delhi – 110 042, India
**E-mail: seba_406@yahoo.in*

ABSTRACT

Ancient Indic languages were written in the Devanagari script from which most of the modern-day Indic writing systems have evolved. The digitisation of ancient Devanagari manuscripts, now archived in national museums, is a part of the language documentation and digital archiving initiative of the Government of India. The challenge in digitizing these handwritten scripts is the lack of adequate datasets for training machine learning models. In our work, we focus on the Devanagari script that has 46 categories of characters that makes training a difficult task, especially when the number of samples are few. We propose deep structure learning of image quadrants, based on learning the hidden state activations derived from convolutional neural networks that are trained separately on five image quadrants. The second phase of our learning module comprises of a deep neural network that learns the hidden state activations of the five convolutional neural networks, fused by concatenation. The experiments prove that the proposed deep structure learning outperforms the state of the art.

Keywords: Digitisation; Devanagari script; Handwritten character recognition; Structure learning; Part learning; Deep neural network; Convolutional neural network

1. INTRODUCTION

Digitisation of ancient Indian scripts is a part of the language documentation and digital archiving initiative of the Government of India. Ancient Indian languages were written in the Devanagari script that was in use in the 7th century C.E. It is considered as the origin of writing systems of several modern-day languages of the Indian sub-continent, such as Nepali, Hindi and Marathi[1]. Digitisation of handwritten scripts found in photographs is complicated since it is a multi-stage process involving several AI components. The first task is segmenting the text area from the image which is a challenging task by itself due to the presence of noise and artifacts such as the presence of foreign objects in the photograph[2]. The next step is to isolate individual letters from the text area by applying image processing techniques such as line detection and vertical and horizontal projection profiles as in[3] that segmented Brahmi letters from a script. After isolating individual characters, the next step is to employ an AI model that would recognise the character as resembling one of the alphabets of the language. This last and crucial step to digitisation and curation of documents, is well researched in literature, with several good surveys available on the topic[4-5]. In late 1980s, the neural network was used for learning directly from handwritten numeral images that were linearised before being given as input to the neural network[6]. From early 2000s, the focus was more on feature engineering. Specialised features for digit recognition were developed such as the box-partitioning method that divides the image into grids or windows, and computes measurements based on the area of the pixels occupied by the digit and the gradient of the image inside the window[7]. In yet another example, the mean normalised distance of each digit pixel from the bottom right corner of the window is computed[8]. In both these cases, the feature vector was formed by concatenating the values from all image sub-boxes, and learnt by a suitable classifier. In later years, a variety of popular image features such as HOG, SIFT, SURF, ORB and LBP were applied for the recognition of handwritten digits or characters[9-10].

In recent years, especially since 2012, deep learning[11] has become the trend in computer vision, with the focus being on 2D image processing using convolutional neural networks (CNN)[12], and increasing the number of layers to form deeper networks[13]. Constructing deeper and wider neural networks has led to an improvement in the classification of handwritten text as well, as seen from examples of Saha and Jaiswal who performed segmentation of Devanagari characters based on contours followed by classification with CNN[14], fine-tuning of pre-trained deep convolutional neural networks[15], capsule networks[16], deep belief networks[17] and deep convolutional neural network by Acharya *et al.*[18]. The deep CNN in[20] contained three repetitive layers of the combination of convolutional layer-rectified linear unit layer-normalisation layer-pooling layer. The problem with deep neural networks is that it lacks interpretability[21] and it is difficult to dissect the decision made by the deep CNN based on image content. Some possible solutions to the interpretability issue are structure learning[22] and part-wise semantic learning of the image[23]. We incorporate both structure learning and part-based learning in our work based on deep neural networks.

Received : 21 September 2020, Revised : 25 September 2020
Accepted : 29 September 2020, Online published : 4 November 2020

We focus on learning the Devanagari handwritten letters that includes the 36 alphabets and 10 numerals, using a two-phase classification module. In the first phase, five image quadrants are isolated and trained on separate Convolutional Neural Networks (CNN). The features extracted from the last dense layer of the convolutional network, for all five image quadrants, are concatenated, and fed as linearised input to a deep neural network (DNN) classifier. The classification scores are proved to be higher than the state of the art. The organisation of this paper is as follows. Section 2 introduces the proposed method. Section 3 discusses about the experimentation and the results. Section 4 gives the final conclusions and the future scope of this work.

2. PROPOSED METHOD

This work extends the idea of part-based learning introduced in a recent work of the authors on recognition of handwritten numerals from the MNIST dataset containing 60,000 English numeral images for training[19]. The 28 x 28 image was divided into five quadrants (a quarter of the image) -top left, top right, bottom left, bottom right and center. Each image quadrant was first linearised into a 1D input vector and was trained on a 196-35-35-35-10 Multi-layer perceptron (MLP) neural network. The hidden state activations (35-dimensional) of the last hidden layer was extracted from each quadrant, and concatenated to form a 175-dimensional feature vector that was trained on a support vector machine classifier. Our work was successful for the ten classes of English numerals. However, for the larger number of classes found in any language script, we have to incorporate deep learning for better text recognition results. This work advances our previous work in the following ways:

- The MLP neural networks used to train the five quadrants are substituted by five CNNs. The configuration of a single CNN is shown in Table 1. This ensures 2D processing of the image quadrant rather than 1D processing as in[19]
- The decision module in the second phase of our experiment comprises of a deep neural network (DNN) that is of configuration 2560-1000-1000-1000-46. This implies that there are three hidden layers each having 1000 neurons. The input layer is of dimension 512 x 5=2560. The number of output neurons is 46 since the number of classes of Devanagari characters are 46
- Our present work thus advances the state of the art by incorporating not only 10 digits, but the entire alphabet and 10 numerals of an ancient Indic script often found in Indian historical documents. We achieve this by incorporating deep learning into the five-quadrant system in[19] which is constricted for fewer classes (10 classes of numerals only).

Table 1. CNN architecture used for training each image quadrant

Layer	Kernel size	Output channels	Output shape	Stride	Activation function
Convolutional (2D)	3x3	16	12x12	1	Sigmoid
Max pooling (2D)	2x2	16	6x6	1	--
Convolutional (2D)	3x3	32	4x4	1	Sigmoid
Max pooling (2D)	2x2	32	2x2	1	--
Flatten	--	--	128		--
Dense (feature extracted)	--	--	512		--
Output	--	--	46		--

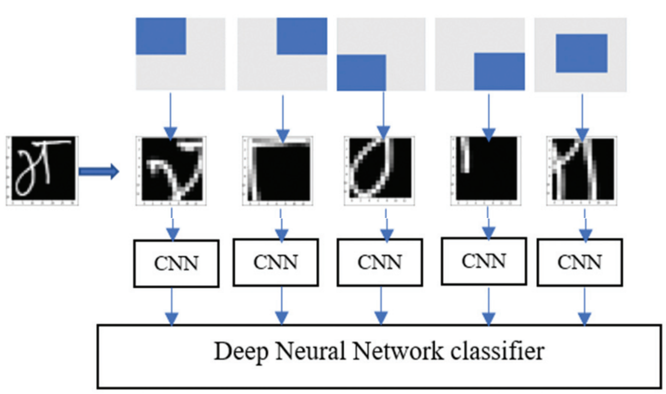

Figure 1. Proposed method.

The proposed approach is illustrated with the help of a block diagram in (Fig. 1). The figure shows the five quadrants -top left, top right, bottom left, bottom right and center. Each image is of size 28 x 28 and each image quadrant of size 14 x 14. Each 2D quadrant is trained on a CNN and the 512-dimensional feature vector (hidden structure) is extracted from the dense layer of each CNN, and concatenated to form the 2560-dimensional feature vector that is given as input to the DNN. The size of the DNN is 2560-1000-1000-1000-46.

3. EXPERIMENTATION AND RESULTS

The experiments are conducted in Python software on an Intel Core processor clocked at 2.8 GHz. The dataset is the UCI Devanagari handwritten character dataset[18] that has 46 categories of characters; samples from each of the 46 classes are shown in (Fig. 2). These include 36 alphabets and 10 numerals.

Overall, there are 92,000 images, with 2000 images for each of the 46 Devanagari characters. The dataset is segregated into 85 per cent training images and 15 per cent testing images at the source. This makes available 78,200 images for training and 13,800 images for testing. The CNN training and the subsequent feature extraction took 30 mins and the second learning module of deep neural network training took 15 mins for training on the entire dataset. The results are summarised in Table 2. As observed, the proposed network gives a high classification score of 98.39 per cent that is higher than the state of the art listed in Table 2. The only other method (Acharya *et al.* (2015))[18] that scores higher is trained on a larger training dataset, compiled by augmenting the data synthetically, by

Figure 2. Samples from 46 classes of Devanagari dataset[18].

Table 2. Performance comparison on the Devanagari test dataset

Method	Test accuracy (%)
Ours	98.39
Saha and Jaiswal (2020)[14]	93
Leathart *et al.* (2019)[20]	79.3
Acharya et al. (2015)[18] (*with five-fold increase of training data—Extended dataset*)	98.47
DenseNet pre-trained network[15]	90
VGG16 pre-trained network[15]	98
AlexNet pre-trained network[15]	98

five-fold. For the presently available training data of 78,200 images, our results show the highest accuracy.

4. CONCLUSIONS

A novel convolutional neural network based approach is proposed in this paper for recognizing Devanagari script from images. Each image is divided into five image quadrants, and each quadrant is trained on a CNN. The output of the dense layer in CNN is extracted and fed as input to a deep neural network. The results yield the maximum recognition scores for the Devanagari script. The contributions of our work are summarised as:

- The MLP neural networks used to train the five quadrants are substituted by five CNNs. This ensures 2D processing of the image quadrant rather than 1D processing as in[19]
- The decision module consists of the second phase of our experiment and comprises of a deep neural network (DNN) that is of configuration 2560-1000-1000-1000-46. This implies that there are three hidden layers each having 1000 neurons. The input layer is of dimension 512 x 5=2560. The number of output neurons is 46 since the number of classes of Devanagari characters are 46
- Our present work thus advances the state of the art by incorporating not only 10 digits but the entire alphabet and 10 numerals of an ancient Indic script often found in Indian historical documents. We achieve this by incorporating deep learning into the five-quadrant system in[19] which is constricted for fewer classes (10 classes of numerals only).

Application of our two-phase deep network module to other ancient Indic scripts such as Brahmi and Indus, forms the future scope of our work.

REFERENCES

1. Setlur, S. Guide to OCR for Indic Scripts. Edited by Venu Govindaraju. Springer London, 2009.
2. Susan, S. & Devi, K.M.R. “Text area segmentation from document images by novel adaptive thresholding and template matching using texture cues.” *Pattern Anal. Appl.*, 2019, 1-13.
3. Singh, A.P. & Kushwaha, A.K. “Analysis of Segmentation Methods for Brahmi Script.” *DESIDOC J. Libr. Inf. Technol.,* 2019, **39**(2).
4. Kumar, M.; Jindal, M.K.; Sharma, R.K. & Jindal, S.R. “Character and numeral recognition for non-Indic and Indic scripts: A survey.” *Artif. Intell. Rev.,* 2019, **52**(4), 2235-2261.
5. Jayadevan, R.; Kolhe, S.R.; Patil, P.M. & Pal, U. “Offline recognition of Devanagari script: A survey.” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2011, **41**(6), 782-796.
6. Le Cun, Yann.; Jackel, L.D.; Boser, B.; Denker, J.S.; Graf, H.P.; Guyon, I.; Henderson, D.; Howard, R.E. & Hubbard, W. “Handwritten digit recognition: Applications of neural network chips and automatic learning.” IEEE Communications Magazine, 1989, **27**(11), 41-46.
7. Liu, Cheng-Lin.; Nakashima, K.; Sako, H. & Fujisawa, H. “Handwritten digit recognition: Investigation of normalisation and feature extraction techniques.”, *Pattern Recognition,* 2004, **37**(2), 265-279.
8. Susan, S. & Singh, V. “On the discriminative power of different feature subsets for handwritten numeral recognition using the box-partitioning method.” *In* 2011 Annual IEEE India Conference, pp. 1-5. IEEE, 2011.
9. Narang, S.R.; Jindal, M.K. & Sharma, P. “Devanagari ancient character recognition using HOG and DCT features.” *In* 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), pp. 215-220. IEEE, 2018.
10. Guruprasad, P. “Handwritten Devanagari word recognition using robust invariant feature transforms.” *In* 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), pp. 327-330. IEEE, 2015.
11. LeCun, Y.; Bengio, Y. & Hinton, G. “Deep learning.” nature, 2015, **521**(7553), 436-444.
12. Alex, K.; Sutskever, I. & Hinton, G.E. “Imagenet classification with deep convolutional neural networks.” *In* Advances in neural information processing systems, pp. 1097-1105. 2012.
13. Ian, G.; Bengio, Y. & Courville, A. Deep learning. MIT press, 2016.
14. Saha, P. & Jaiswal, A. “Handwriting Recognition Using Active Contour.” *In* Artificial Intelligence and Evolutionary Computations in Engineering Systems, pp. 505-514. Springer, Singapore, 2020.
15. Aneja, N. & Aneja, S. “Transfer learning using CNN for handwritten Devanagari character recognition.”

In 2019 1st International Conference on Advances in Information Technology (ICAIT), pp. 293-296. IEEE, 2019.

16. Jeel, S.; Suba, Y. & D'silva, M. "Devanagari character classification using capsule network." *In* International Conference on Intelligent Systems Design and Applications, pp. 1040-1049. Springer, Cham, 2018.
17. Sazal, Md Musfiqur Rahman; Biswas, S.; Amin, Md Faijul & Murase, K. "Bangla handwritten character recognition using deep belief network." *In* 2013 International Conference on Electrical Information and Communication Technology (EICT), pp. 1-5. IEEE, 2014.
18. Acharya, S.; Pant, A.K. & Gyawali, P.K. "Deep learning based large scale handwritten Devanagari character recognition." *In* 2015 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), pp. 1-6. IEEE, 2015.
19. Susan, S. & Malhotra, J. "Learning interpretable hidden state structures for handwritten numeral recognition." *In* 2020 4th International Conference on Computational Intelligence and Networks (CINE), pp. 1-6. IEEE, 2020.
20. Tim, L.; Frank, E.; Pfahringer, B. & Holmes, G. "On calibration of nested dichotomies." *In* Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 69-80. Springer, Cham, 2019.
21. Zhang, Q. & Zhu, S. "Visual interpretability for deep learning: a survey." *Front. Inf. Technol. Electron. Eng.,* 2018, **19**(1), 27-39.
22. Hou, Bo-Jian & Zhou, Zhi-Hua. "Learning with interpretable structure from rnn." arXiv preprint arXiv:1810.10708 (2018).
23. Quanshi, Z.; Cao, R.; Wu, Ying Nian & Zhu, Song-Chun. "Growing interpretable part graphs on convnets via multi-shot learning." *In* Thirty-First AAAI Conference on Artificial Intelligence. 2017.

CONTRIBUTORS

Dr Seba Susan is an Associate Professor in the Department of Information Technology at Delhi Technological University (DTU). She completed her Ph.D from the Electrical Engineering Department of Indian Institute of Technology (IIT) Delhi in 2014. Her research areas are Machine Learning, Computer Vision, Speech and Natural Language Processing.
Contriution in the current study is that she designed the idea and concept, and wrote the paper.

Mr Jatin Malhotra is a postgraduate student in Information Systems in the Department of Information Technology at Delhi Technological University (DTU). His areas of interest are Machine learning and Image processing. His current research involves developing deep learning modules for handwriting recognition.
Contribution in the current study is that he contributed to the design of the CNN module and did the software simulation.