Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 171 (2020) 2294–2301

Procedia Computer Science

www.elsevier.com/locate/procedia

Third International Conference on Computing and Network Communications (CoCoNet'19)

OCR-Nets: Variants of Pre-trained CNN for Urdu Handwritten Character Recognition via Transfer Learning

Mohammed Aarif K.O[a*] Sivakumar Poruran[b]

[a] Faculty of Electronics & Communication Engineering, C. Abdul Hakeem College of Engineering & Technology, Melvisharam, 632509, India
[b]Faculty of Electronics & Communication Engineering, Dr. N.G.P.Institute of Technology, Coimbatore, 641048, India

Abstract

Deep Convolutional neural networks (CNN) have been among the utmost competitive neural network architectures and have set the state-of-the-art in various fields of computer vision. In this paper, we present OCR-Nets, variants of (AlexNet & GoogleNet) for recognition of handwritten Urdu characters through transfer learning. Our proposed networks are experimented using an integrated dataset. To compare the recognition rate with traditional character recognition methods and to confirm the fairness of the experiment an additional Urdu character dataset is manually generated with different fonts and size. The experimental result shows that OCR-AlexNet and OCR-GoogleNet produce significant performance gains of 96.3% and 94.7% averaged success rate respectively.

Keywords: Pattern recognition;Transfer learning;Optical Character recognition;AlexNet;GoogleNet; Document analysis.

1. Introduction

CNN has exhibited to perform outstandingly in various machine learning issues in computer vision. CNN has not only propelling in the classification of the image yet in further play as all intense feature extractors for various acknowledgment work. In the ongoing work on acknowledgment utilizing CNN, a significant number of the scientists have used it for classification and few of them have moved toward it for feature extraction, as in [1] has exhibited face and image representation in deep CNN features and [2] has displayed iris acknowledgment utilizing off the shell CNN features.

* Corresponding author. Tel.: +919791317928
E-mail address: aarifko.ece@ cahcet.edu.in

10.1016/j.procs.2020.04.248

With promising precision on scene arrangement and tanning leather image characterization with pre-trained CNN has been accounted in [3,4]. Urdu is one of a primitive and essential language of Indian sub-landmass. Around the globe Urdu is spoken in excess of 15 nations with an estimation more than 250 million speakers. An Enormous collection of significant Urdu literature from Islamic investigations to science is accessible, which yet has not been digitized. Urdu text recognition system based on script has been a critical field of research for decades, which can be used as a part of the learning and teaching method to the learners and how to investigate and get a handle on instructive substance of Urdu content. Over the most recent multi decade not many research work has concentrated on building up an optimize acknowledgment framework for Urdu content. Nastalique style Arabic content is utilized for composing Urdu. Because of its cursive nature, (ligature) word development by joining at least two character dynamic structure makes Urdu character recognition hard to process. Majority of the character acknowledgment strategies utilized for Urdu content are acquired from available systems used on Arabic characters; although, different methodologies are refined distinctly for handwritten or printed Urdu character recognition. In this paper we proposed a character recognition system for handwritten Urdu character by transfer learning pre-trained CNN. The major contribution of this work is (i) Exploiting pre-trained CNN by transfer learning Urdu character, (ii) Optimized fine tuning network to quickly learn the new dataset and (iii) Consistency is accuracy compare to other conventional systems.

2. Related Work

The interest for character recognition has gotten sensible consideration since a couple of decades. From the past few years, CNN has gained extreme attention in the field of character recognition. Like, Layer-wise preparing of Deep CNN for Handwritten Devanagari acknowledgment has been exhibited in [5] and accomplished great outcomes. Reference [6] attempted CNN for Isolated Bangla Handwritten Character Recognition and achieved results close to the state-of-the-art in Bangla character recognition. Script identification in multilingual document images using deep CNN is presented in [7]. CNN based Word level script identification of Latin and Indic content has additionally been shown in [8]. Arabic and Urdu script are highly cursive, and few works have been exhibited utilizing CNN, where it is used either as a sole classifier or composite classifiers [9, 10]. An Urdu ligature level classification utilizing CNN has additionally been given promising results in [11]. Indeed, even with few classes, CNN outshines some other technique by transfer learning on huge datasets [12], [13]. A semi-regulated transfer learning of CNN for recognition of Chinese characters is introduced in [14]. The transfer learning approach utilizing CNN has satisfactorily reinforced for numeral acknowledgment of Hindi, Arabic & Bangla in [15] and Oriya, Telugu, and Devanagari in [16]. The literature shows that CNN has been effectively applied to printed Devanagari [5], Chinese [13], Bangla [6], [14] Tamil [17] and Arabic [9] contents. In this paper, IFHCDB one of the largest datasets for the cursive script is used for experimentation which has a total of 52380 characters and 17740 numerals images which is trained in our newly designed OCR-Nets which are variants of AlexNet and GoogleNet. The test accuracy is matched depending on the training iterations and presents the classification experiment which is based on the additional Urdu character data which is not available in the IFHCDB. Then, the recognition rates are compared with conventional methods which are considered as state-of-the-art in character recognition. The rest of this paper is structured as follows- Section 3 provides a brief introduction to CNN and section 4 describes our OCR-Net architectures. Section 5 provides the experimentation details and compare the performance of proposed OCR-AlexNet and OCR-GoogleNet with the newly created dataset. Section 6 presents a comparison of test accuracy with conventional methods. Future work and conclusions are drawn in section 7.

3. Convolutional Neural Network – Character Recognition

CNN which is the reformist neural system has a massive accurate limit that learns exceptional features at each layer of the visual order. CNN have numerous layers, in particular, convolutional layers, Rectified Linear Unit (ReLU) layers, normalization layers, pooling layers, fully associated layers, dropout layers, and softmax layers. The layers are organized in such a manner that the data pours through each layer in the systems, convert the data, and permits the data on the following layer. The system will gain explicitly from the data and develops the many-sided quality and detail of whatever it is collecting from the sequence of layers. The execution of the structure may be enhanced by managing diverse layers of the system and the use of pre-trained CNN in light of the way that the framework may slow down out in a neighborhood optima specific to starting space of the training data [18].

4. Design of OCR-nets for Urdu character recognition

OCR- AlexNet: Winner of ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012, first successful CNN application for such a big dataset with top-5 test error rate of 15.3%. The network consist of 8 deep layers, relatively simple layout (compared to modern architectures) trained well with 1000 classes of ImageNet dataset. Subsequently, the system has refined unexpected element portrayals for a wide scope of pictures. Transfer learning is regularly utilized in profound learning applications utilizing CNN when we don't have an enormous arrangement of labelled data. Calibrating a system with transfer learning is generally a lot quicker and simpler than preparing a system with arbitrarily initialized weights from scratch. We can quickly transfer learned features to our character recognition task using a lesser number of training images. Our OCR-AlexNet overall architecture is similar to AlexNet. The last three layers of AlexNet are fully connected layers with the last layer conveying a classification of 1000 labels. The primary convolution layer takes a contribution of 227 x 227 size picture. Thus, all the pictures are resized to the required estimations. 4096 neurons are related to each of the full associated layers [19]. The over-fitting in AlexNet is diminished by Data Augmentation and Dropout. In dropout [20], the covered neurons are set to zero with a probability of 0.5. This maintains a strategic distance from forwarding pass and back-propagation and empowers neurons to learn ground-breaking highlights as they never again rely upon different neurons [19]. These three layers must be fine-tuned for our new classification problem. So we inherit all the layers excluding the last three from the pretrained network and then reassigned the layers to our new classification task by substituting the last three layers with a fully connected layer, a softmax layer, and a classification output layer. We manipulate the options for the new fully connected layer according to the new data. We set the completely associated layer to have a similar size as the quantity of classes (54) as in our new classes. We also increased the WeightLearnRateFactor and BiasLearnRateFactor values of the fully connected layer to learn faster. The detailed OCR-AlexNet is depicted in figure 1.

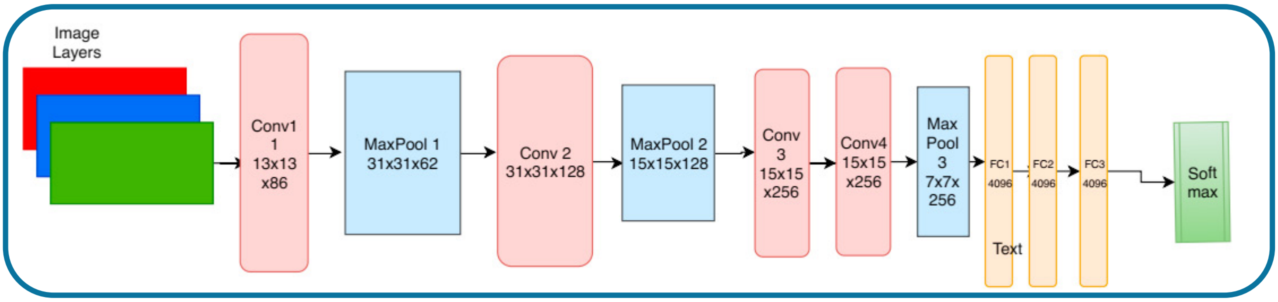

Fig. 1. OCR-AlexNet Architecture

OCR-GoogleNet: GoogleNet is a convolutional neural framework that is set up on more than a million pictures from the ImageNet database [19] and won ILSVRC-2014 which has a lot further design than existing CNN structures. The framework is 22 layers deep and can describe pictures into 1000 distinct classes. The network has a picture input size of 224-by-224. This network is constructed using 9 inception modules which are different from conventional CNN and it has only a one dimensional series configuration. This inception well prompts the local space features by sectioning the local characteristic of the kernel space of different sizes to compute the convolutional value and all the convolutional results are concatenated in the final layers of inception. The convolutional layers of the system extricate image features and the last two layers are used for classification. These two layers, loss3-classifier, and yield in GoogleNet, contains data on the most proficient method to join the features that the network extricates into class probabilities, predicated labels and a loss function. We retrain the pretrained network with only 4 inception module and replace the last two layers with new layers revised to our new dataset. We then freeze the weight of the earlier layer in the network by setting the learning rates in those as zero. Freezing the earlier layers reduces the probability of overfitting to the new dataset. We set the initial learn rate to a little, an incentive to back off learning in the transferred layers that are not effectively frozen. Then we expanded the learning rate factors for the last learnable layer to accelerate learning in the new last layers. This procedure of settings, results in quick learning in the new layers, slower learning in the center layers, and no learning in the frozen layers.

5. Experiment Analysis

Our Integrated dataset is constructed using IFHCDB database by extracting character images which are similar in both Urdu and Farsi script and the other character images are manually developed which totally has 54 classes of Urdu characters [21], in a total of 52380 characters images and 17740 numerals images. These images are subdivided into multiple ratios from each class for training, testing, and validation. The input image size to the OCR-AlexNet is 227 x 227 by 3 and for OCR-GoogleNet it requires 224 x 224 by 3. So we utilized an augmented image datastore to dynamically resize the input images. For both the OCR-Nets we set the preliminary learning rate to 0.9 and the momentum constant to 0.01 for all training iteration. By 0.96 fold we decrease the learning rate for every iteration. All through the investigation, an NVIDIA -1060 graphics Zotak (CUDAv10.0) was utilized for a fast response. Table 1 shows that OCR-AlexNet has the highest accuracy rate for all the ratio of IFHCDB dataset than OCR-GoogleNet. Again for the manually prepared dataset OCR-AlexNet shows better performance than OCR-GoogleNet.

Table 1. Comparison of OCR-AlexNet and OCR-GoogleNet.

Datasets Ratio	OCR-AlexNet	OCR-GoogleNet
(6:4)	95.4	93.7
(7:3)	95.8	93.9
(7.5:2.5)	96.2	94.8
(8:2)	96.8	95.2
(8.5:1.5)	97.3	95.7
Self-Prepared Dataset	93.8	91.4

Figure 2. Shows that the OCR-AlexNet took 29.9 Minutes and OCR-GoogleNet took 235.34 Minutes on average to complete the training and execution. This is because OCR-GoogleNet spends more time in the inception module in finding features than OCR-AlexNet. For time constraint recognition like online Urdu OCR, OCR-AlexNet is more efficient for training Urdu characters.

6. Performance comparison of OCR-nets with conventional methods

To compare the performance of our proposed network features we took the following most potential feature vectors from the literature which have proven to be the actual state-of-the-art in character recognition system. 1. Histogram of Oriented Gradients (HOG), 2. Gabor features, 3.Moment Invariant, 4. Geometric features. For the conventional handcraft features, the images are resized to 80 x 80 and the individual feature vector is constructed by the following. (i). Gabor features [22] Selective mean features in different orientation and frequency- Feature Vector. (ii) Histogram of oriented gradient [23] for single-cell size 144 descriptor, concatenated (4x4) and (6x6) = 288 features. (iii) Invariant moment [24] 7 invariant moment for each character image + 7 for each equal size 4 blocks (28)+ 7 for each 3 vertical zone (21) + 7 for each 3 horizontal zone (21)= 77 features. (iv) Geometric features [25] is based on a number of vertical, horizontal and diagonal line in different zoning and direction. For the Urdu character image, it found averaged to be 366. Both the handcrafted features and features extracted using OCR-Nets are trained and validated using SVM, K-NN, Random forest LDA classifiers. Table 2 summarizes the recognition rate of our proposed feature compared with all 4 individual conventional methods. The feature extraction rate of conventional methods is greater than our OCR-AlexNet even though conventional methods are utilized a minimum number of features.

Table 2. Recognition Rate Comparison

Classifiers	OCR-A	OCR-G	Gabor[21]	HOG[22]	M.Invr[24]	Stru[25]
SVM	**96.6**	**82.3**	91.8	89.4	86.1	80.2
K-NN	**96.2**	**82.5**	91.9	90.5	86.6	82.3
Random Forest	**95.9**	**80.2**	89.7	87.0	82.4	82.5
LDA	**95.3**	**93.8**	86.4	84.3	81.9	80.2

Fig. 2. Execution Time Comparison.

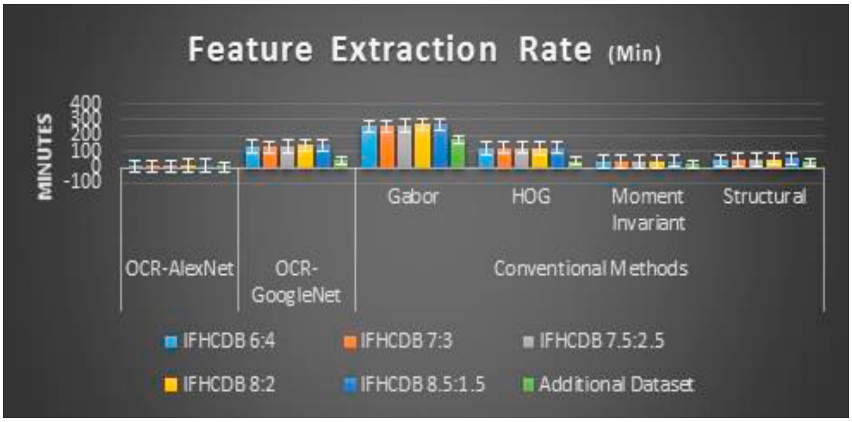

Fig.3. Comparison of feature extraction rate

Figure 3. Demonstrates the feature extraction time correlation of our proposed work. In the conventional methods input image needs to encounter course of action of endeavors before figuring the features while in the prepared component extraction strategy, the features are particularly isolated from the training and testing image with enactment work which is the weights from the specific completely associated layer. The remarkable execution of the

proposed method basically caused by utilizing OCR-Nets which is really prepared for harder portrayal issues with 1000 classes of objects.

Our OCR-Nets are capable of recognizing both offline and online characters. For Online Real time Visual Inspection system mostly touchscreen, light pen, etc. are utilized [26] [27]. In our experiment for online recognition, we have interfaced smart phone camera with Matlab to capture the handwritten character written on screen or on a sheet. Our network are made to get the captured image and fed it to the input layer after resizing it in to the required size dynamically. Figure 4 shows the demonstration set-up of visual inspection of our Urdu character recognition system.

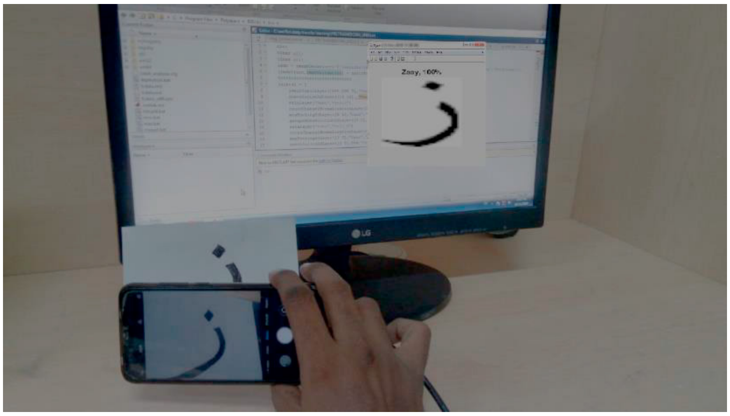

Fig.4 Online Urdu OCR set-up using Smart phone camera

7. Conclusion and future work

The deep convolutional neural network is continuously proving to be the state-of-the-art in various fields of computer vision and practically considered in significant areas of document analysis and character recognition. In this paper, two CNN architectures OCR-AlexNet and OCR-GoogleNet are presented which are variants of AlexNet and GoogleNet pretrained convolutional neural network. In AlexNet the initial layer weights are freeze and the last three-layers are fine-tuned to train and classify Urdu characters. In GoogleNet only four inception module is used and the last two layers are modified to learn and classify the Urdu characters. Our experimental analysis shows that OCR-AlexNet and OCR-GoogleNet produce significant performance gains of 96.3% and 94.7% averaged recognition rate respectively. We also manually developed a dataset that was not in the IFHCDB database for objective evaluation. The experimental results showed that OCR-AlexNet can extract effective and exceptional features form Urdu character images compare to conventional methods. The recognition rate accuracy of AlexNet based features validates that the fine-tuned AlexNet-SVM offers the state-of-the-art significant results. In addition to the recognition rate, we also measured the temporal factor of the experiment. The execution time for OCR-AlexNet is short compared to OCR-GoogleNet and feature extraction time for OCR-AlexNet is the least compare to OCR-GoogleNet and all other conventional methods. It is additionally found in our experiment that the recognition rate is impacted by the number of training images, the size of the info pictures and the weight learners in the network. So, investigating the strategy of the input image for enhancing classification accuracy is our future heading research.

Acknowledgment

We are very much thankful to Dr. Karim Faez, Machine Vision Research Lab (MVRL), Amirkabir University of Technology, Teheran, for providing us the IFHCDB [28] database.

References

[1] Connor J. Parde, Carlos Castillo, Matthew Q. Hill, Y. Ivette Colon, Swami Sankaranarayanan, Jun-Cheng, Chen and Alice J. OToole. (2017) "Face and Image Representation in Deep CNN Features" *In: Proceedings of IEEE 12th International Conference on Automatic Face & Gesture Recognition, Washington, DC.*

[2] Kien Nguyen, Clinton Fookes, Arun Ross, Sridha Sridharan (2017)."Iris Recognition with Off-the-Shelf CNN Feature A Deep Learning Perspective", *IEEE ACCESS.*

[3] Lisha Xiao, Qin Yan, and Shuyu Deng, Scene Classification with Improved AlexNet Model (2017*). In: proceeding of 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)* Nanjing, China.

[4] Sri Winiarti, Adhi Prahara, Murinto, Dewi Pramudi Ismi.(2018). "Pre-Trained Convolutional Neural Network for Classification of Tanning Leather Image". *International Journal of Advanced Computer Science and Applications*, Vol. 9,

[5] Mahesh Jangid and Sumit Srivastava. (2018) "Handwritten Devanagari Character Recognition Using Layer-Wise Training of Deep Convolutional Neural Networks and Adaptive Gradient Methods". *Journal of Imaging*, 4, 41. https://doi.org/10.3390/jimaging4020041.

[6] Mujadded Al Rabbani Alif, Sabbir Ahmed, Muhammad Abul Hasan. (2017) "Isolated Bangla Handwritten Character Recognition with Convolutional Neural Network". *In: Proceedings of 20th International Conference of Computer and Information Technology (ICCIT),* Dhaka, Bangladesh.

[7] Jieru Meiy, Luo Dai, Baoguang Shi, Xiang Bai (2016). "Scene Text Script Identification with Convolutional Recurrent Neural Networks." *In: proceeding of 2016 23rd International Conference on Pattern Recognition (ICPR)* Cancun Centre, Cancun, Mexico.

[8] Soumya Ukil, Sk Swarnendu Ghosh Md Obaidullah, K. C. Santosh, Kaushik Roy and Nibaran Das. (2018) "Deep learning for word-level handwritten Indic script identification", *Published 2019 in ArXiv*.

[9] Mohamed Elleuch, RaniaMaalejand Monji Kherallah.(2016) "A New Design Based-SVM of the CNN Classifier Architecture with Dropout for Offline Arabic Handwritten Recognition." *In: proceeding of ICCS International Conference on Computational Science*, Volume 80 .San Diego, California, U.S.A, Elsevier.

[10] Max Jaderberg, Karen Simonyan, Andrea Vedaldi & Andrew Zisserman Deep S. (2015). "Structured Output Learning For Unconstrained Text Recognition".*In: proceeding of International Conference on Learning Representation,ICLR*. San Diego, CA.

[11] Nizwa Javed, Safia Shabbir, Imran Siddiqi, Khurram Khurshid. (2017). "Classification of Urdu Ligatures Using Convolutional Neural Networks." *In: proceeding of IEEE International Conference on A Novel Approach, Frontiers of Information Technology (FIT),*Islamabad, Pakistan.

[12] Joris Gurin, Olivier Gibaru, Eric Nyiri, and Stphane Thiery. (2017). "CNN features are also great at unsupervised classification." *In Proceeding of 30th Conference on Neural Information Processing Systems*. Long Beach, CA, USA.

[13] Yejun Tang, Bing Wu, Liangrui Peng, and Changsong Liu, Semisupervised. (2017). "Transfer Learning for Convolutional Neural Network-based Chinese Character Recognition". *In: proceeding of IEEE- 14th IAPR International Conference on Document Analysis and Recognition.* Kyoto, Japan.

[14] Tushar A.K., Ashiquzzaman A., Afrin A., Islam M.R. (2018). "A Novel Transfer Learning Approach upon Hindi, Arabic, and Bangla Numerals Using Convolutional Neural Networks". *In: Hemanth D., Smys S. (eds) Computational Vision and Bio-Inspired Computing. Lecture Notes in Computational Vision and Biomechanics*, vol 28. Springer, Cham

[15] Abdullah and Mohammad S. Hasan. (2017) "An Application of pre-Trained CNN for Image Classification". *In: proceeding of 20th International Conference of Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh.

[16] Durjoy Sen Maitra, Ujjwal Bhattacharya, and Swapan K. Parui. (2015). "CNN Based Common Approach to Handwritten Character Recognition of Multiple Scripts. *In: proceeding of 13th International Conference on Document Analysis and Recognition*, Tunisia.

[17] M. A. Pragathi, K. Priyadarshini, S. Saveetha, A. Shavar Banu and Mohammed Aarif. "Handwritten Tamil Character Recognition UsingDeep Learning". In: proceeding of IEEE -2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), India. DOI: 10.1109/ViTECoN.2019.8899614.

[18] S. Sudholt and G. A. Fink. (2016). "PHOCNet: A deep convolutional neural network for word spotting in handwritten 44 documents." *In: Proceeding of International Conference on Frontiers in Handwriting Recognition ICFHR, S*henzhen, China.

[19] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012) "ImageNet classification with deep convolutional neural networks". *In: proceeding of advance in Neural Information Processing System,* Lake Tahoe.

[20] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. (2012). "Improving neural networks by preventing co-adaptation of feature detectors" *arXiv: 1207. 0580vl*.

[21] Zaheer Ahmed, Urdu Optical Character Recognition Using Neural Networks. (2009) "A thesis for the degree of Master of Science in Information Technology". *(MS-IT), Institute of Management Sciences* Peshawar, Pakistan January.

[22] Peeta Basa Pati and A. G. Ramakrishnan.(2008) "Word-Level Multi-script Identification". *Pattern Recognition Letter*, Volume 29 Issue 9, July 2008 Pages 1218-1229. Elsevier Science Inc.

[23] Dalal, N. and Triggs, B. (2005). "Histograms of Oriented Gradients for Human Detection". *In: proceeding of IEEE Computer Society Conference*, vol. 1, San Diego, CA, USA, USA.

[24] Imran Khan Pathan, Abdulbari Ahmed Ali and Ramteke R.J. (2012). "Recognition Of Offline Handwritten Isolated Urdu Character". *Advance in Computer Research* ISSN: 0975- 3273 pp.-117-121.

[25] Muhammad, Arif Mohamad, Dewi Nasien, Haswadi Hassan, Habibollah Haron. (2015). "A Review on Feature Extraction and Feature Selection for Handwritten Character Recognition". *International Journal of Advanced Computer Science and Application*, Vol. 6, No. 2, 2015.

[26] Sen S., Shaoo D., Paul S., Sarkar R., Roy K. (2018) Online Handwritten Bangla Character Recognition Using CNN: A Deep Learning Approach. In: Bhateja V., Coello Coello C., Satapathy S., Pattnaik P. (eds) Intelligent Engineering Informatics. Advances in Intelligent Systems and Computing, vol 695. Springer, Singapore.

[27] Suresh Sundaram, A.G. Ramakrishnan (2009). An Improved Online Tamil Character Recognition Engine using Post-Processing Methods" in the Proceedings of the 10th International Conference on Document Analysis and Recognition July 2009, pp. 1216-1220.

[28] S. Mozaffari, K. Faez, F. Faradji, M. Ziaratban and S. M. Golzan, "A. Comprehensive Isolated Farsi/Arabic Character Database for Handwritten OCR Research," International Workshop on Frontiers in Handwriting Recognition, 2006.