Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol.6, No.1, June 2018, pp. 285~295
Available online at: http://pen.ius.edu.ba

An Efficient Convolutional Neural Network Based Classifier to Predict Tamil Writer

Thendral Tharmalingam[1]*, Vijaya Vijayakumar [2]

[1] 1Research Scholar, Department of Computer Science, PSGR Krishnammal College for Women, Coimbatore, India.
2Associate Professor, Department of Computer Science, PSGR Krishnammal College for Women, Coimbatore, India

Article Info

Article history:

Received Jan 12th, 2018
Revised Apr 20th, 2018
Accepted Jun 26th, 2018

Keyword:

Convolutional Neural Networks (CNNs),
Artificial Neural Network (ANN),
Architecture,
Classification,
Deep Learning, Neurons,
Tamil handwriting, Writer Identification,

ABSTRACT

Identification of Tamil handwritten calligraphies at different levels such as character, word and paragraph is complicated when compared to other western language scripts. None of the existing methods provides efficient Tamil handwriting writer identification (THWI). Also offline Tamil handwritten identification at different levels still offers many motivating challenges to researchers. This paper employs a deep learning algorithm for handwriting image classification. Deep learning has its own dimensions to generate new features from a limited set of training dataset. Convolutional Neural Networks (CNNs) is one of deep, feed-forward artificial neural network is applied to THWI. The dataset collection and classification phase of CNN enables data access and automatic feature generation. Since the number of parameters is significantly reduced, training time to THWI is proportionally reduced. Understandably, the CNNs produced much higher identification rate compared with traditional ANN at different levels of handwriting.

Corresponding Author:

Thendral Tharmalingam,
Departement of Computer Engineering,
PSGR Krishnammal College for Women, Coimbatore, India.
Email: thendral@psgrkc.ac.in

1. Introduction

Writer identification can be defined to be the capability of a computer to accept and understand intelligible handwritten input received from sources like paper documents, photographs, touch-screens and other devices. Writer identification has become a huge challenge in the present day, as the character, word structure and orientation is dependent on different factors concerned with the persons who are writing it. Moreover, the field of Writer identification is classified into off-line and on-line identification. Owing to the absence of temporal data, off-line writer identification task is regarded to be more tedious compared to on-line. It is also evident that the off-line scenario is the one, which is associated with the traditional reading task carried out by individuals [1].

The challenges in Writer identification systems are the similarities observed between few characters, words with one another, innumerable character shapes, distorted and unintelligible characters. Indian scripts are distinct from the other scripts. Many of the Tamil letters [2] exhibit circular shapes partly because of the fact that they were actually engraved with needles over palm leaves, which is a technology that preferred round shapes. Tamil is one among the most traditionally spoken languages in the entire world [3]. This work predominantly focuses on offline handwritten detection of South Indian languages, especially Tamil.

Intelligent system that can perform the identification of Tamil writers is still an open challenge for the research personnel. One among the most significant research fields in the present day world includes pattern analysis and machine intelligence. Even though a huge number of new system and classification mechanisms have been evolved in this area like [4-6], correct recognition rate with regard to the prediction of the pattern is still debatable.

Deep learning can be observed to be a form of hierarchical learning, in which algorithms exploit different layers of representations in order to slowly modify information into high level concepts. A "deep" architecture is basically one that comprises of more than one layer of modifications from the input end to output end. Traversing from the input, through every layer of modifications, features at the higher level are obtained from that of the lower level, resulting in a hierarchical representation. Deep Learning has developed as a new potential area of research in the field statistical machine learning. Learning algorithms used for deep architectures revolve around the learning of resourceful representations of data, which better suit the task at present and are structured in a hierarchy having different levels.

Deep Neural Networks (DNN) [7-9] is an extension of classical Neural Networks (NNs) with deep layers having high dimension of parameters (millions to billions). DNN models [10] have moved to possess multiple structures for diverse applications, inclusive of Multiple-Layer Perceptron's (MLP), Convolutional Neural Networks (CNN), Deep Belief Networks (DBN) employed in image classification, recognition, and deep auto encoders utilized in writer classification.

On the converse, Deep Neural Networks (DNN) has been getting increased attention in the fields of machine learning and pattern identification. As the DNNs model can have multiple layers, much more complex functions exist compared to the common Neural Networks (NNs) classifier [11]. The availability of immense scale training data and advancements in computing technologies, the training of such sort of deep networks has been made possible, resulting in a broad adaptation of DNNs in several problem domains. For instance, Deep Convolutional Neural Networks (DCNNs) have exhibited remarkable performance in several image processing applications, defeating benchmark performances by huge margins [12-14].

CNN could be particularly advantageous for Tamil Handwriting Writer Identification (THWI) for several reasons. First, CNN carries out features learning and classification inside a unified framework. As the features are automatically learned from the data itself, it may be feasible to capture subtle features to isolate the puzzling characters seen in Tamil handwritings. Secondly, the performance of CNN in extraction is high-level features is very good. Specifically, the convolution and subsampling Layers utilized by CNN have been indicated to be very efficient in dealing with shape changes that will probably be the key challenge in coping up with the too much of cursiveness in Tamil writings. Hence, CNN is regarded to be well-prepared in overcoming the two important hurdles in THWI.

2. Deep learning and CNN

The recent evolution in machine learning has resulted in a manifold increase of attention towards Artificial Neural Networks (ANN) [15-16]. ANNs consists of a set of neurons, gathered together by means of synapses. Neurons carry out ordinary computational task, usually an elementary yes/no decision [17]. Synapses connect neurons together by connecting their inputs and outputs. In terms of programming, a synapse is typically an object that connects one neuron that is linked to its input to another that is linked to its output. A neuron is a little more complicated object that can be linked to one or more number of input synapses and one or more number of output synapses. Therefore, the structure of any neural network is defined by the means in which different neurons and synapses are connected together. The fundamental structure of ANN is illustrated in Fig. 1. It is composed of three important layers like input layer, hidden layer and output layer. The input dataset is fed in the input layer and then hidden layer stage classification or identification is carried out by employing the objective function. Finally the categorized results are shown in the output layer phase.

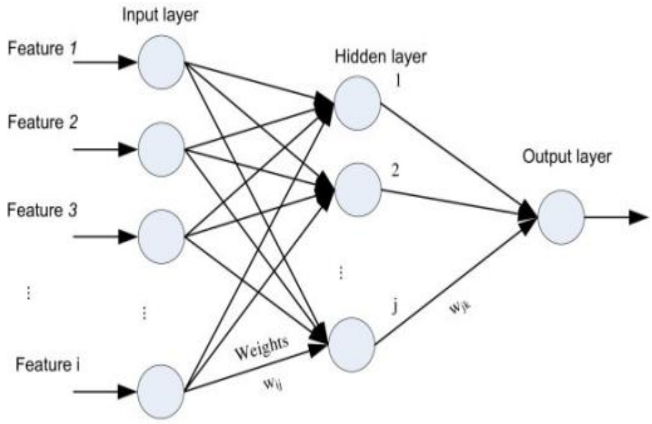

Fig.1. Artificial Neural Network (ANN) architecture

Generally, the organization of neurons is done in layers. Multiple layers may indulge in various types of modifications on their inputs. Signals move from the input, then to the output layer, probably after travelling through the layers several number of times. The actual objective of the neural network technique was to resolve challenges in the same manner as it would be done by a human brain. Over time, the focus on matching particular mental capabilities, led to divergence from biology like Backpropagation (BP), or traversing information in the opposite direction and adapting the network to exhibit that information.

DNN is motivated by the observations made in the mammalian visual cortex that comprises of a group of processing elements, each one of which is related to a diverse representation of the raw evident input. Actually, it was recently discovered that the features that are learnt in deep architectures are similar to those that are seen in the first two visual cortex phases of V1 and V2 [18], and also their invariance to factors of changes in higher layers [19] is increasing more and more. Learning a hierarchy of features helps in increasing the convenience and feasibility of evolving representations, were specially made to certain tasks, and still capable of borrowing the statistical strength from other relevant tasks. At last, learning of the feature representation can result in higher-level feature that are more reliable to unexpected sources of variance extant in actual data.

In the last few years, Deep Learning under Neural Network (DNNs) has generated remarkable results in the fields of machine learning and pattern identification [20]. One more reason for the DNNs performing better is that DNNs facilitate the combined training of feature extractors and classifiers. Contrary to classical classifiers, most of the DNNs permit raw images in the form of input, and do not need individual feature extraction or pre-processing, except for the case of size normalization. The low- and middle-level DNN layers perform the extraction and abstraction of the feature from the input image, whereas high-level layers carry out classification. In this manner, a DNN can be considered to be a unified framework, which provides the integration of all the modules within a single network, which can be optimized systematically with regard to a single objective function. Such kind of unified training can frequently result in a better performance compared to those that are on the basis of independent training of every module.

Along with the general benefits of DNNs, DCNN has few additional characteristics: DCNN can efficiently learn to extract and abstract the features. Specifically, the sub-sampling layer of DCNN shows great efficiency in the absorption of shape variations. Also, being comprised of sparse connection having tied weights, DCNN has considerably lesser parameters compared to an entirely connected network of similar size. On top of all, DCNN is trainable with the help of gradient-based learning algorithm, and is less affected by the problem of diminishing gradient. But the gradient-based algorithm trains the entire network to reduce an error rate, therefore DCNN can generate every accurate and optimized weights.

2.1. Convolutional Neural Networks (CNNs)

The aim of Convolutional Neural Networks (CNNs) to make use of spatial information between the pixels present in Tamil script images having three levels like characters, word and paragraph. This CNNs classifier functions on the basis of discrete convolution function. The layers of CNNs are explained as below [21] and [22].

a) Convolution

For the sake of simplicity, let it be assumed that a grayscale Tamil script images is specified by a function

$$THI: \{1, .. n_1\} \times \{1, \dots n_2\} \rightarrow W \subseteq \mathbb{R}, (i, j) \mapsto THI_{i,j} \quad (1)$$

So that the Tamil handwriting images (THI) can be indicated by an array of size $n_1 \times n_2$. Provided the filterK $\in \mathbb{R}^{(2h_1+1\times 2h_2+1)}$, the discrete convolution of the Tamil handwriting image I with filter K is expressed by

$$(THI * K)_{r,s} := \sum_{u=-h_1}^{h_1} \sum_{v=-h_2}^{h_2} K_{u,v} THI_{r+u,s+v} \quad (2)$$

where the filter K is expressed by $K = \begin{pmatrix} K_{-h_1,-h_2} & \cdots & K_{-h_1,h_2} \\ \vdots & K_{0,0} & \vdots \\ K_{h_1,-h_2} & \cdots & K_{h_1,h_2} \end{pmatrix}$ (3)

b) Layers

Fig.2. indicates the various layers of the CNNs employed in the proposed research work. Five diverse kinds of layers exist in CNN are: Input layer, Convolution layer, Sub-sampling layer, Pooling layer and output layer. The dataset of the Tamil writer is provided as input to the CNN input layer, and at the Convolution layer, the input samples are mapped in order to feature the maps from the layer before. After this, in the sub-sampling layer, the elementary operation of the CNNs is carried out in order to determine the results of identification. At the pooling layer, back propagation of feature activations is carried out for measuring the results. At last, the results are shown in the output layer.

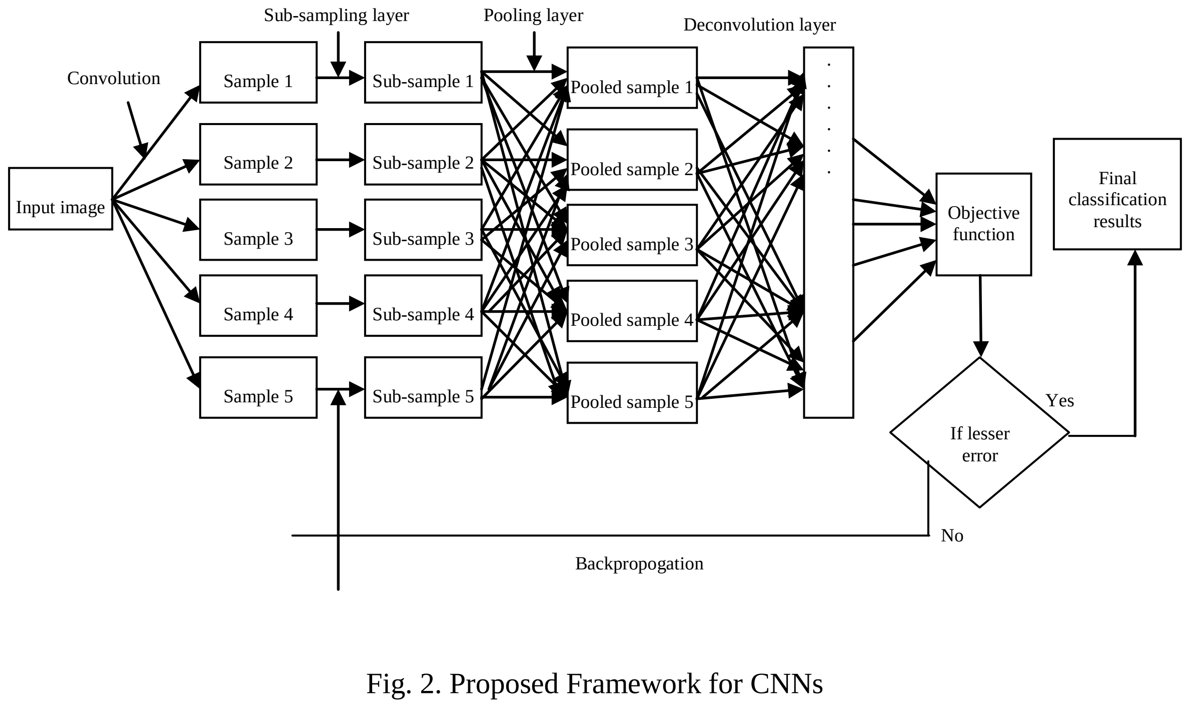

Fig. 2. Proposed Framework for CNNs

Convolutional Layer

Consider layer l refer to a convolutional layer. Then, the input of layer l consists of $m_1^{(l-1)}$ Tamil Handwritten Image feature maps from the previous layer, each with size $m_2^{(l-1)} \times m_3^{(l-1)}$. In the scenario where $l = 1$, the input is basically a single image I comprising of one or more number of channels. By this manner, a CNN directly gets the raw images in the form of input. The output of layer l comprises of $m_1^{(l)}$ feature maps of size $m_2^{(l)} \times m_3^{(l)}$. The i[th] feature map in layer l, represented $Y_i^{(l)}$, is calculated as

$$Y_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{m_1^{(l-1)}} K_{i,j}^{(l)} * Y_j^{(l-1)} \qquad (4)$$

Where$B_i^{(l)}$refers to a bias matrix and $K_{i,j}^{(l)}$ indicates the filter of size $2h_1^{(l)} + 1 \times 2h_2^{(l)} + 1$ linking the j[th] feature map in layer (l −1) with the i[th] feature map in layer l [23]. As stated above, $m_2^{(l)}$ and $m_3^{(l)}$ are affected by border effects. When only discrete

Convolution is applied in the valid region of the input THWI dependent feature maps that are only for pixels where the sum of equation (2) is specified, the output feature maps obtain size.

$$m_2^{(l)} = m_2^{(l-1)} - 2h_1^{(l)} \text{ and} m_3^{(l)} = m_3^{(l-1)} - 2h_2^{(l)} \qquad (5)$$

Frequently the filters utilized for the computation of a fixed feature map $Y_i^{(l)}$ are similar, that is $K_{i,j}^{(l)} = K_{i,k}^{(l)}$for $j \neq k$. Moreover, and the sum in equation (2) may also spread over a subset of the input feature maps. In order to associate the convolutional layer and its operation as expressed by equation (2) to the multilayer perceptron, the above equation is rewritten. Every feature map $Y_i^{(l)}$in layer l comprises of $m_2^{(l)}.m_3^{(l)}$ units sorted in a two-dimensional array. The unit present at position (r, s) calculates the output

$$\left(Y_i^{(l)}\right) = \left(B_i^{(l)}\right)_{r,s} + \sum_{j=1}^{m_1^{(l-1)}} (K_{i,j}^{(l)} * Y_j^{(l-1)})_{r,s} = \left(B_i^{(l)}\right)_{r,s} + \sum_{j=1}^{m_1^{(l-1)}} \sum_{u=-h_1^{(l)}}^{h_1^{(l)}} \sum_{v=-h_2^{(l)}}^{h_2^{(l)}} \left(K_{ij}^{(l)}\right)_{u,v} \left(Y_j^{(l-1)}\right)_{r+u,s+v} \qquad (6)$$

Feature Pooling and Subsampling Layer

The inspiration behind sub sampling the feature maps got from earlier layers is reliability towards noise and disturbances [20]. Let l refer to a pooling layer. Its output consists of $m_1^{(l)} = m_1^{(l-1)}$ feature maps of minimized size. Generally, pooling functions by keeping windows at non-overlapping positions in every feature map and placing one value for each window so that the feature maps get subsampled.

Max pooling: For the purpose of max pooling, the maximum value of every window is considered. The layer is indicated by P_M. Max pooling is utilized for getting rapid convergence during the time of training.

Max pooling can also be brought into used by employing overlapping windows of size 2p×2p that are kept at a distance of q units apart. So the windows overlap when q < p. This is revealed to minimize the possibility of over fitting the training set. Then the objective function will get the error rate. In case the error is greater, then the backpropagation layer is invoked in order to carry out the new convolution operation so as to minimize the error rate.

Identical to CNNs, Deconvolutional Neural Networks are in accordance with the concept of generation of feature hierarchies through the convolution of the input image by means of a set of filters present at every layer [24]. Nonetheless, by definition, deconvolutional neural networks are unsupervised. Moreover, deconvolutional neural networks are dependent on a top-down approach [25]. This shows that the aim is the reconstruction of the network input from its activations and filters. This inspires to realize CNN for the current research.

3. Experiments and results

Initially, documents are collected from 300 Tamil writers with 100 handwritten images leading to total of (300*100=30000) images. It is segmented into three levels of samples such as characters, words and paragraphs. The samples at different levels are given as input to ANN and CNN for THWI in training mode and testing mode in the ratio of 80% and 20%. The batch size denotes the number of images taken for THWI. Epoch denotes the total number of pass through the full training set required to complete task.

The performance comparison results of three levels of handwriting such as character, word and paragraph with different batch sizes (5, 5, and 10) and their epochs are (1, 5, and 10), and input and subsampling layer are presented in Table I. The recognition rate, precision, recall and F-measure results of various networks under different settings of hidden layers for writer identification are also presented in Table I. The error rate results of CNN classifier with three different levels for writer identification is discussed in Table II. The results of recent CNN model are compared with traditional ANN against various performance metrics to validate the performance CNN based writer identification model. ANN implementation was done by extracting distinct features such as Gabor Filter, Gray Level Co-Occurrence Matrix (GLCM), Generalized Gaussian Density (GGD), Contourlet GGD and directional features from pre-processed images at paragraph level. For character and word levels of handwriting images, the local and global features were considered as per our previous work. On the other hand, CNN extract features from the images on its own by subsampling layer. The comparative results of both CNN and ANN based THWI models are presented in Table III.

CNN classifier performs better when compared to ANN classifier for all three datasets. The proposed CNN classifier achieves 96.3%, 97.5% and 98.2% of recognition rate for character, word and paragraph level respectively in batch size 10. It is concluded that the proposed CNN classifier with three different character, word and paragraph level produces better results and lesser error value of 3.7%, 2.5% and 1.8% respectively in batch size 10.

Table I: Recognition rate for different THWI in CNN

Batch Size	**Epoch**	**Input Layer**	**Subsampling Layer**	**Recognition rate (%)**	**Precision**	**Recall**	**F-Measure**
Character level Text							
5	1	5	10	82	0.867	0.905	0.886
6	5	6	12	85.7	0.894	0.910	0.902
10	10	10	20	96.3	0.960	0.994	0.977
Word level Text							
5	1	5	10	90	0.919	0.963	0.941
6	5	6	12	95.2	0.958	0.986	0.972
10	10	10	20	97.5	0.977	0.983	0.980
Paragraph level Text							
5	1	5	10	90.6	0.916	0.958	0.937
6	5	6	12	97.7	0.979	0.987	0.983
10	10	10	20	98.2	0.985	0.989	0.987

Table II: Error rate for different THWI in CNN

Batch Size	Epoch	Input Layer	Subsampling Layer	Error rate
Character level Text				
5	1	5	10	18
6	5	6	12	14.3
10	10	10	20	3.7
Word level Text				
5	1	5	10	10
6	5	6	12	4.8
10	10	10	20	2.5
Paragraph level Text				
5	1	5	10	9.4
6	5	6	12	2.3
10	10	10	20	1.8

Table III: Recognition rate comparison in ANN and CNN

	Recognition rate (%)	Precision	Recall	F-Measure
ANN				
Character	79%	0.871	0.867	0.869
Word	82%	0.836	0.937	0.883
Paragraph	86.5%	0.926	0.761	2.001
CNN				
Character	96.3	0.960	0.994	0.977
Word	97.5	0.977	0.983	0.980
Paragraph	98.2	0.985	0.989	0.987

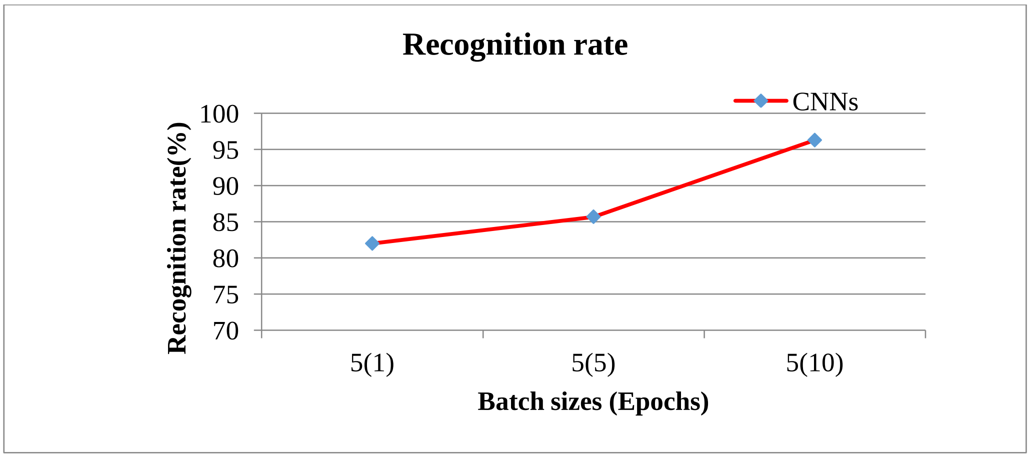

Fig.3. Recognition rate of character level identification

The recognition rates and error rates of CNN models at different handwriting levels measured with respect to batch size (epochs) along x axis and recognition rate (%) along y axis are depicted in Fig. 3. to Fig. 8. CNN achieved high performance of 98.2% compared with ANN of 86.5%. The required features are self-extracted by CNN from the input images and hence the recognition rate is simultaneously increased.

Fig.4. Error rate of character level identification

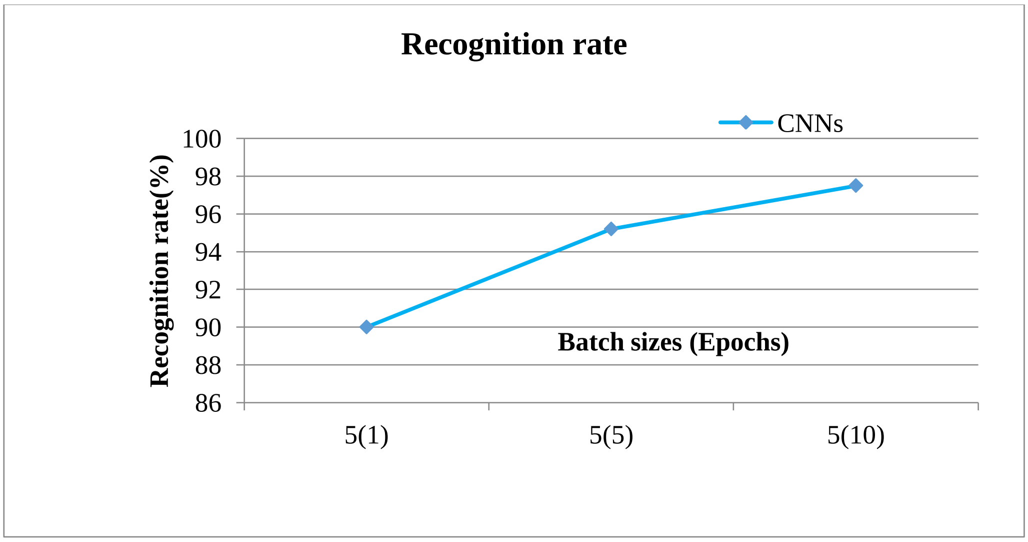

Fig.5. Recognition rate of word level identification

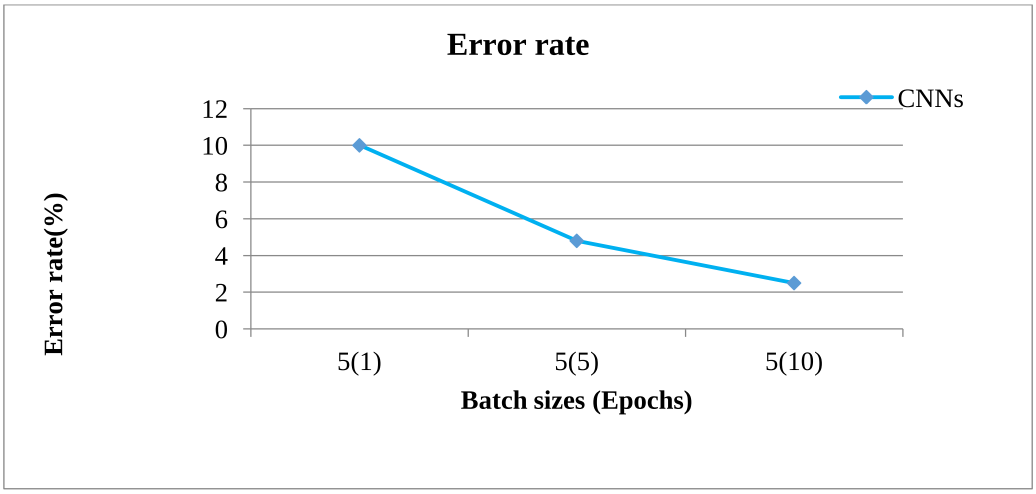

Fig. 6. Error rate of word level identification

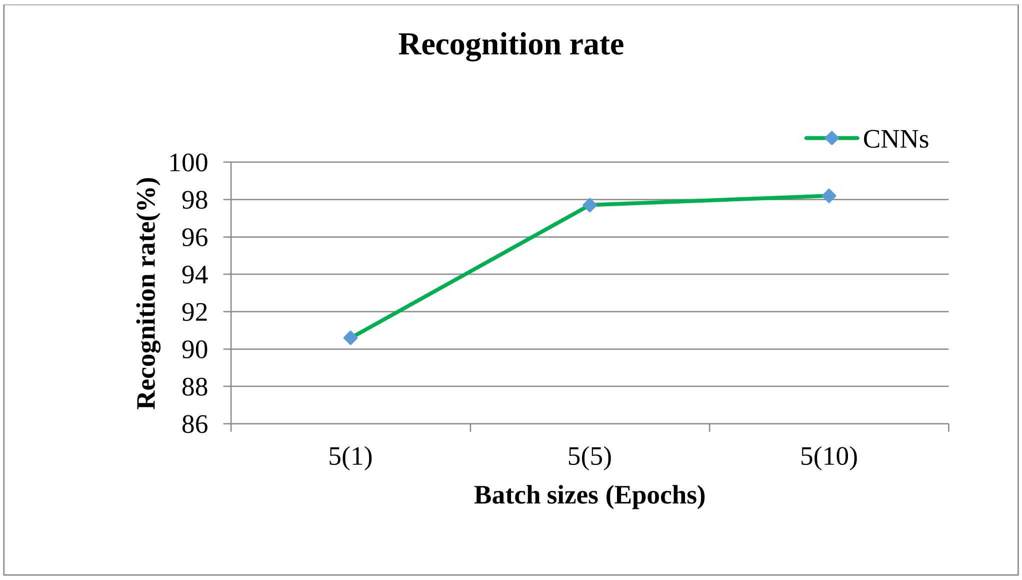

Fig. 7. Recognition rate of paragraph level identification

Fig. 8. Error rate of paragraph level identification

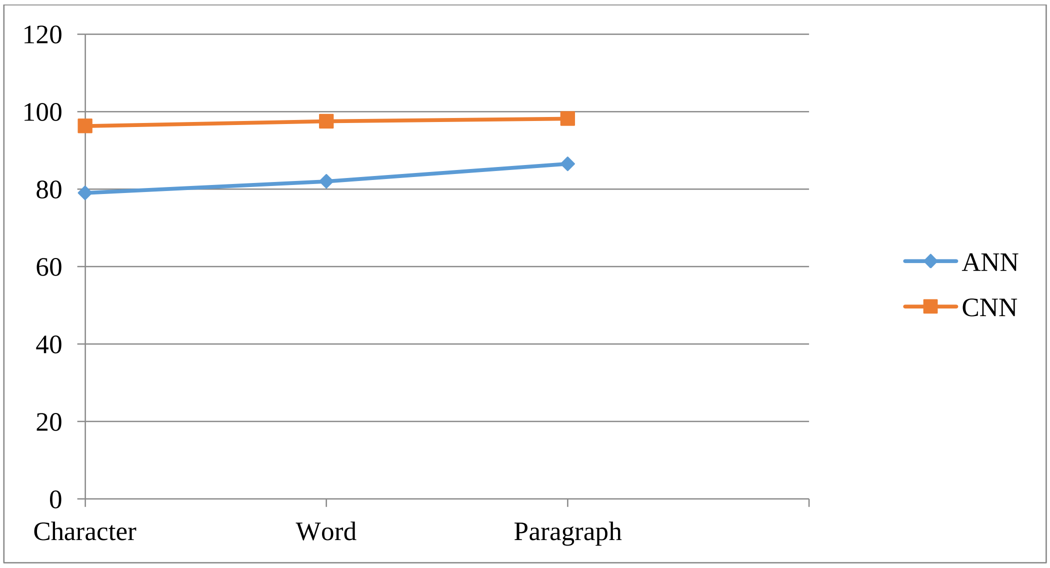

Fig.9. Comparative analysis of ANN and CNN

Thus it has been proved that Convolutional neural network can handle a difficult problem like writer identification of Tamil handwriting with good results. CNN plays a major role to extract features from the input image and the hidden layered architecture facilitates capturing of still more discriminating features. CNN is very effective in absorbing shape variations of handwritten images. It suffers less from the high dimensional problem which minimizes the error rate of the identification. CNN carries out feature learning and classification inside a unified framework which helps to predict writer more accurately with minimum time period.

4. Conclusion and future work

Present research work demonstrates application of Convolutional Neural Network based classifier for Tamil Writers identification at different levels of handwriting. From the experimental results it is concluded that the CNNs classifier produces higher recognition rate with reduced error rates in the paragraph level image than in ANN classifier. CNN is very effective in self-extraction of features and modelling Tamil handwriting writing identification. It is also possible to easily upgrade the existing CNN based classifier for other Indian languages like Kannada, Malayalam, Telugu and so on.

References

1. Kannan, R.J., Prabhakar, R. and Suresh, R.M., 2008, Off-line cursive handwritten Tamil character Identification. International Conference on Security Technology, 2008(SECTECH'08), pp. 159-164, 2008.
2. Selvakumar Raja, S. and John, M., 2010. A Novel Method of Tamil Character Recognition Using Slope Method. IUP Journal of Science & Technology, 6(2), pp. 7-14.
3. Bharath A. and Madhvanath S., 2012. HMM-based lexicon-driven and lexicon-free word recognition for online handwritten Indic scripts. IEEE transactions on pattern analysis and machine intelligence, 34(4), pp.670-682.
4. Khalighi S., Tirdad P. and Rabiee H.R., A novel OCR system for calculating handwritten persian arithmetic expressions. IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2009, pp. 277-282.
5. Liu C.L., Nakashima K., SakoH. and Fujisawa H., 2002. Handwritten digit recognition using state-of-the-art techniques. Eighth International Workshop on Frontiers in Handwriting Recognition, pp. 320-325, 2002.
6. Deng, L.: The MNIST Database of Handwritten Digit Images for Machine Learning Research. IEEE Signal Processing Magazine, 141–142 (2012).
7. Gu J., Zhu M., Zhou Z., Zhang F., Lin Z., Zhang Q. and Breternitz M., 2014, Implementation and evaluation of deep neural networks (DNN) on mainstream heterogeneous systems. In Proceedings of 5thAsia-Pacific Workshop on Systems, pp.1-7.
8. Zhang, J. and Zong, C., 2015. Deep neural networks in machine translation: An overview, pp.16-25.
9. Liu, Y.T., Tsao, Y. and Chang, R.Y., 2015, June. A deep neural network based approach to Mandarin consonant/vowel separation. IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), pp. 324-325.
10. Collobert R. and Weston J., 2008, “A unified architecture for natural language processing: Deep neural networks with multitask learning”, Proceedings of the 25th international conference on Machine learning, pp. 160-167.
11. Y. Bengio, Learning deep architectures for AI, Foundations and Trends in Machine Learning, vol. 2, no. 1, pp. 1-127, 2009.

12. Liu, C.L., Yin, F., Wang, D.H. and Wang, Q.F., 2011, CASIA online and offline Chinese handwriting databases. International Conference on Document Analysis and Recognition (ICDAR), 37-41.
13. D. C. Cireşan, U. Meier, J. Schmidhuber. Multi-column Deep Neural Networks for Image Classification. IEEE Conference on Computer Vision and Pattern Recognition 2012.
14. D. C. Cireşan and J. Schmidhuber. Multi-column Deep Neural Networks for Offline Handwritten Chinese Character Classification, IDSIA Technical Report No. IDSIA-05-13, 2013.
15. Abraham, A., 2005. Artificial neural networks. Handbook of measuring system design, pp.1-8.
16. Yegnanarayana B. Artificial neural networks. PHI Learning Pvt. Ltd.; 2009, pp.1-476.
17. Karayiannis, N. and Venetsanopoulos, A.N., 2013. Artificial neural networks: learning algorithms, performance evaluation, and applications, Springer Science & Business Media, Vol. 209, pp.1- 440.
18. Honglak Lee, Chaitanya Ekanadham, and Andrew Ng. Sparse deep belief net model for visual area V2. In NIPS’07, pp. 873–880. MIT Press, Cambridge, MA, 2008.
19. Ian Goodfellow, Quoc Le, Andrew Saxe, and Andrew Ng. Measuring in variances in deep networks. In NIPS’09, pp. 646–654. 2009.
20. Liao, Q. and Poggio, T. (2016). Bridging the gap between residual learning, recurrent neural networks and visual cortex. Center for Brains, Minds and Machines (CBMM) Memo No. 47, also in arXiv
21. K. Jarrett, K. Kavukcuogl, M. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition?, International Conference on Computer Vision, pp. 2146–2153, 2009.
22. Y. LeCun, K. Kavukvuoglu, and C. Farabet. Convolutional networks and applications in vision. International Symposium on Circuits and Systems, pp. 253–256, 2010.
23. Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431-3440.
24. M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus. Deconvolutional networks. In Computer Vision and Pattern Recognition, pp. 2528–2535, 2010.
25. Le Callet, Patrick; Christian Viard-Gaudin; Dominique Barba (2006). "A Convolutional Neural Network Approach for Objective Video Quality Assessment" (PDF). IEEE Transactions on Neural Networks. 17 (5): 1316–1327.