Bangla Handwritten Character Recognition Using Deep Belief Network

Md. Musfiqur Rahman Sazal[1,*], Sujan Kumar Biswas[1], Md. Faijul Amin[1], and Kazuyuki Murase[2]
[1]Khulna University of Engineering & Technology, Bangladesh
[2]University of Fukui, Japan
E-mail: musfiqsazal@msdn.net.bd[*], sujan.kuet@msdn.net.bd, mdfaijulamin@yahoo.com, and murase@u-fukui.ac.jp

Abstract— **Recognition of Bangla handwritten characters is a difficult but important task for various emerging applications. For better recognition performance, good feature representation of the character images is a primary requirement. In this study, we investigate a recently proposed machine learning approach called deep learning [1] for Bangla hand written character recognition, with a focus on automatic learning of good representations. This approach differs from the traditional methods of preprocessing the characters for constructing the handcrafted features such as loops and strokes. Among different deep learning structures, we employ the deep belief network (DBN) that takes the raw character images as input and learning proceeds in two steps – an unsupervised feature learning followed by a supervised fine tuning of the network parameters. Unlike traditional neural networks, the DBN is a probabilistic generative model, i.e., we can generate samples from the model and it can fit both the semi-supervised and supervised learning settings. We demonstrate the advantages of unsupervised feature learning through the experimental studies carried on the Bangla basic characters and numerals dataset collected from the Indian Statistical Institute.**

Keywords—Deep belief network, unsupervised feature learning, Bangla handwritten character recognition, supervised learning, backpropagation.

I. INTRODUCTION

Handwritten character recognition has received a lot of attention because of its various applications such as postal mail sorting according to zip code [2], signature verification, and bank-check processing. In the Bangla language, however, most of the works have been done for the recognition of printed characters [3]. While each category of the printed characters has relatively less variation in the character images and easier to be dealt with, it is far more difficult to recognize a handwritten character because of various factors like inconsistency in the writer's handwriting, different ways of writing and noise in the data. Moreover, the characters in Bangla language have high variations such as different shapes, sizes, loops and strokes. Therefore, the recognition of Bangla handwritten characters is a challenging task.

In order to achieve better recognition performance, research on character or object recognition from images has been heavily relied on good feature representation, which involves hand-crafted feature engineering and goes through complicated preprocessing steps [4]. Such feature engineering approach is application dependent and requires human prudence or ingenuity. Consequently, when building an application of pattern recognition, most of the human effort is spent for constructing discriminating features and less is given to the classifier design [5].

Recently, hierarchical feature learning by deep architectures in an unsupervised manner has revolutionized the machine learning researches [1]. The unsupervised feature learning is a new perspective whose goal is to learn good representation of input data automatically without considering the labels [6]. In order to achieve this goal layers of neurons are arranged hierarchically to form a deep architecture. Each layer learns a new representation from its previous layer with a goal of modeling different explanatory factors of variation behind the data. It is hypothesized that high level complex features can be learned from low level simpler features [7]. For example, in vision problems, bottom layer might learn edges from the scene. The next layer could discover contours and the subsequent layers could learn even more complex features. Once good features are learnt, they can be exploited in various tasks.

In this paper, we study deep belief network (DBN) [1] for Bangla handwritten character recognition problem. The DBN is a particular type of deep architecture that can gradually learn complex structures of the data by learning its probability distribution. It takes raw image data as input with the hope that subsequent layers would learn good feature representation. Learning in the network takes two steps – an unsupervised feature learning followed by a supervised learning of discriminating function. The unsupervised learning is performed by using contrastive divergence (CD) algorithm which is an approximation of maximum likelihood estimation [8], while in the later stage the network parameters are fine-tuned with the gradient based backpropagation (BP) algorithm [9]. We conducted our experimental studies on a dataset of Bangla handwritten characters and numerals from the Indian Statistical Institute. It is demonstrated here that on the one hand, the unsupervised feature learning relieves human effort for constructing application dependent features, and on the other hand, the neural network classifier achieves much higher recognition performance than the classifier trained only by BP algorithm.

The organization of paper is as follows. Section II provides an overview of the related works. Section III describes deep belief network. Section IV presents details of our experiment and a performance comparison. Finally, Section V provides conclusion.

II. Related Works

There has been extensive research works on handwritten character recognition for different languages, such as Roman scripts, English, several European languages, and some Asian languages like Chinese, Korean and Japanese. However, relatively few works have been done for the Bangla language. An important research contributions relating to the handwritten Bangla characters involve a multistage method developed by Rahman et al. [10]. Another interesting work is an MLP classifier developed by Bhowmik et al. [11]. The prime features for the multistage approach include upper part of the character, vertical line, and double vertical line. For the MLP classifier [11], alphabetic features were designed by the human expert.

The DBN, an alternative approach that do not require feature engineering by the human expert has been recently developed and employed for recognizing English handwritten numerals [12]. Here the features are extracted from the pixel data of character image in an unsupervised manner, i.e., without considering the labels. Another technique that extracts local features is the convolutional neural network [13], [14]. Note that these methods have a common goal of learning distributed representation [9] which resulted in the resurgence of research in the neural networks. However, it was realized that gradient based BP algorithm is not suitable for several layered architectures. The DBN and convolutional neural network avoids such difficulties, but retain the original goal of learning distributed representation. We noted that such approaches have not been applied yet on the handwritten Bangla character recognition problem. Therefore, we study here one of aforementioned methods, the DBN, to analyze its recognition performance on the Bangla handwritten character recognition task.

III. Deep Belief Network (DBN)

The DBN is a multilayer neural network that can work as a probabilistic generative model [1]. The DBN as shown in Figure 1 is composed of several layers of stochastic hidden variables and one layer of visible units. The basic module of a

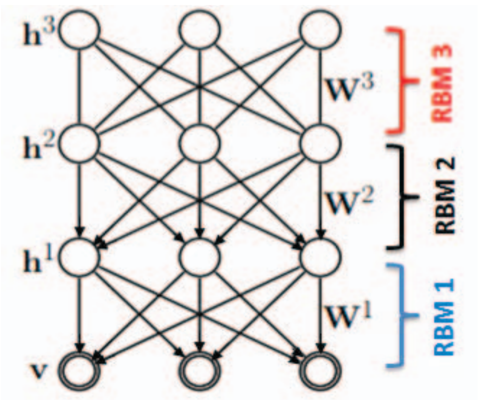

Figure 1. DBN composed with three stacking RBMs

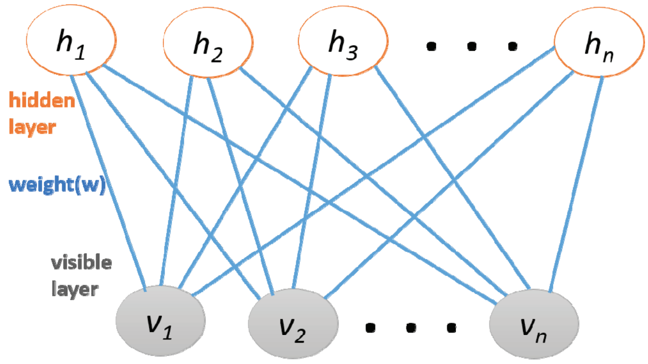

Figure 2. An RBM with no hidden-to-hidden or visible-to-visible connections.

DBN is called restricted Boltzmann machine which we explain next.

A. *Restricted Boltzmann Machine (RBM)*

The RBM is a stochastic neural network consisting of one layer of visible units and one layer of hidden units. It can be considered as a bipartite graph where all visible units are connected to all hidden units, but there are no visible-visible or hidden-hidden connections as shown in Figure 2.

The weights on the connecting edges and the bias units define a probability distribution over a binary vector of visible units, $\mathbf{v} = [v_1, v_2, \ldots, v_V]$ and a binary vector of hidden units, $\mathbf{h} = [h_1, h_2, \ldots, h_H]$. The energy function for the joint configuration is given by

$$E(\mathbf{v}, \mathbf{h}; \theta) = -\sum_{i=1}^{V}\sum_{j=1}^{H} w_{ij} v_i h_j - \sum_{i=1}^{V} b_i v_i - \sum_{j=1}^{H} a_j h_j \quad (1)$$

where $\theta = (\mathbf{W}, \mathbf{b}, \mathbf{a})$ and w_{ij} represents the weights between visible units i and hidden units j and b_i and a_j are their biases. V and H are the number of visible and hidden units. The probability of the visible vector $\mathbf{v}$ is defined by

$$p(\mathbf{v}; \theta) = \frac{\sum_{h} e^{-E(v,n)}}{\sum_{n}\sum_{h} e^{-E(u,h)}} \quad (2)$$

As there is no connection among the units of same layer, conditional distributions (see [15] for details) are given by

$$p(h_j = 1|\mathbf{v}; \theta) = \sigma\left(\sum_{i=1}^{V} w_{ij} v_i + a_j\right) \quad (3)$$

$$p(v_i = 1|\mathbf{h}; \theta) = \sigma\left(\sum_{j=1}^{H} w_{ij} h_j + b_i\right) \quad (4)$$

where $\sigma(x) = (1 + e^{-x})^{-1}$ is an activation function. Thus an RBM models a joint distribution of visible and hidden units.

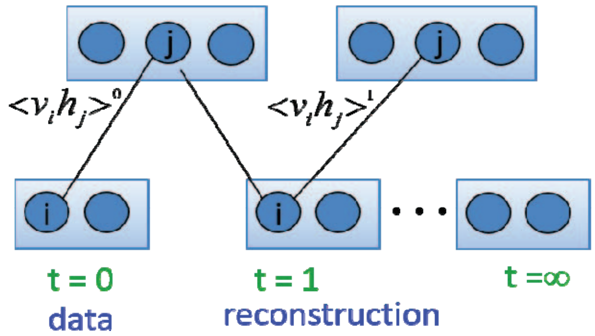

Figure 3. Training a single layer RBM with CD

For training an RBM to model joint distribution of data and class labels [16], an additional visible vector is added with a binary vector of class labels, $\mathbf{l} = [l_1, l_2, \ldots, l_L]$. The energy function with class labels becomes:

$$E(\mathbf{v}, \mathbf{l}, \mathbf{h}; \theta) = -\sum_{i=1}^{V}\sum_{j=1}^{H} w_{ij} h_j v_i - \sum_{y=1}^{L}\sum_{j=1}^{H} w_{yj} h_j l_y - \sum_{j=1}^{H} a_j h_j - \sum_{y=1}^{L} c_y l_y - \sum_{i=1}^{V} b_i v_i \quad (5)$$

and here,

$$p(l_y = 1|\mathbf{h}; \theta) = \text{softmax}\left(\sum_{j=1}^{H} w_{yj} h_j + c_y\right) \quad (6)$$

Then $p(\mathbf{l} \mid \mathbf{v})$ is calculated using

$$p(\mathbf{l} \mid \mathbf{v}) = \frac{\sum_{h} e^{-E(v,l,n)}}{\sum_{l}\sum_{h} e^{-E(u,l,h)}} \quad (7)$$

B. Training RBM

An RBM is trained to learn probability distribution of the data with the help of hidden stochastic variables. Since maximizing the likelihood requires a Markov chain Monte Carlo simulation taking a long time, an approximation called CD is generally employed [17]. The training process of a single layer RBM using CD is shown in Figure 3. In the CD procedure, the weights of the RBMs are updated to learn the features from the sample data. The weight update rule is defined as:

$$w_{ij}^{new} = w_{ij}^{old} + \Delta w_{ij} \quad (8)$$

Where $\Delta w_{ij} = \varepsilon\left(\left\langle v_i h_j \right\rangle_{data} - \left\langle v_i h_j \right\rangle_{model}\right)$ (9)

Here, $\left\langle v_i h_j \right\rangle_{data}$ is the expectation over training data and $\left\langle v_i h_j \right\rangle_{model}$ is the expectation over reconstructed data and ϵ is the learning rate parameter.

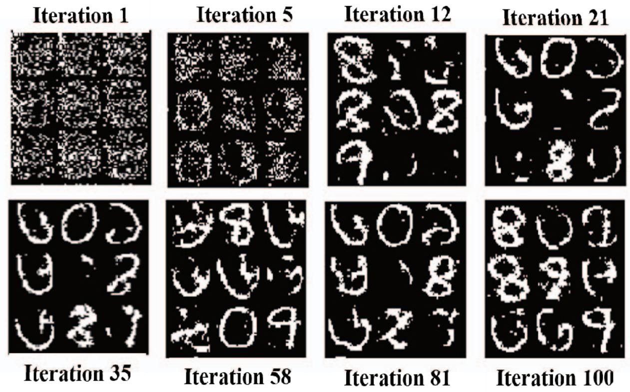

Figure 4. Reconstructed digit images with the corresponding iteration number

C. Training DBN

Basically, learning in the DBN involves a greedy layer-wise unsupervised learning algorithm. When the weights of an RBM have been learned, the vector of hidden feature activations can be used as visible data for training the next RBM that learns a higher layer of features. In this way, a greedy layer-wise unsupervised training is done in the DBN with RBMs as the building blocks for each layer. This is also called pertaining [16], after which a softmax layer (Eq. (6)) is added for classification problem and the entire weights are optimized using BP algorithm to minimize the error rate.

IV. EXPERIMENT AND PERFORMANCE ANALYSIS

In order to evaluate the effectiveness of unsupervised feature learning approach using DBN, we investigated the performance on the dataset of Bangla numerals and other basic characters collected from Indian Statistical Institute [18]. There are 10 numerals and 50 basic characters in the data set.

We used 27900 images for training and 8600 images for testing. All the images were rescaled to 28×28 pixels and binarized. The architecture of DBN was 784-500-1000-2000-60, i.e., it consisted of three RBMs. The DBN was trained with the algorithms described in Section III. In order to see the usefulness of unsupervised pre-training (feature learning), we also the trained another DBN without unsupervised training part, but only conjugate gradient BP was used. The overall classification accuracy for DBN with and without unsupervised pre-training was 90.27% and 75.30%, respectively, as can be seen from TABLE I. It can be clearly observed that unsupervised feature learning significantly improved the recognition performance.

To show a glimpse of recognition performance, a confusion matrix for only numerals is given in TABLE II. The largest confusion was observed between one ("১") and nine ("৯"). Indeed, the digits look almost similar in the handwritten form. The average accuracy over the numerals was 91.30% which is slightly higher than the overall accuracy.

One of the prime features of the DBN approach is the ability of image reconstruction as it can work as a generative model. The better the reconstruction, the better the learning,

TABLE I: PERFORMANCE COMPARISONS OF DIFFERENT APPROACHES

Method	**Average classification accuracy**
Only Supervised Learning (Conjugate gradient Backpropagation)	75.30%
Unsupervised Learning + Supervised Fine Tuning	90.27%

and it helps attaining better classification accuracy. To illustrate, a few samples of reconstructed images, during the unsupervised feature learning stage, are shown in Figure 4. We can see that the model gradually learns a probability distribution of the visible data such that the sampled (reconstructed) images become meaningful (recognizable) and clearer. The reconstruction error rate curve in the training phase is shown in Figure 5.

We now compare the recognition performance of DBN with other methods, hierarchical learning architecture (HLA) schemes proposed by Bhowmik et al. [19] who reported their results on the same dataset we used in our experiments. The comparison is given in TABLE III. The comparison result reveals that the DBN approach can achieve higher recognition rate even though no handcrafted feature was used. Useful features in the DBN are rather learned during the training process.

TABLE III: PERFORMANCE COMPARISONS BETWEEN DBNS AND HLA FOR BANGLA CHARACTERS

Scheme	**Average test accuracy (%)**
Using DBNs	90.27
HLA with disjoint groups (HLA_DG)	84.78 ± 2.02
HLA with overlapped groups (HLA_OG)	88.02 ± 1.55
HLA with groups using neural gas (NG)	83.58 ± 1.98

TABLE II: CONFUSION MATRIX FOR BANGLA NUMERALS

Predicted Digits

Target Digits

	০	১	২	৩	৪	৫	৬	৭	৮	৯
০	95.7	0	0	1.25	0	1.30	0.05	1.65	0	0.05
১	0.20	94.3	1.36	0.5	0	0.05	0	0.25	0	3.34
২	0	1.72	89.65	1.63	0	0	0	0.15	0	0.20
৩	1.2	1.25	0	90.48	0	0.02	6.85	0	0	0.2
৪	0	0	0	0	93.75	0	0	4.2	1.52	0.53
৫	6.53	0	0.2	0	0	89.01	3.37	0.05	0.75	0
৬	0	0	0	7.5	0	1.75	89.99	0	0.5	0.26
৭	3.62	0	0	0	0	0.78	0	95.25	0	0.35
৮	0	0	0.05	0	6.5	0	2.2	0.05	90.35	0.85
৯	0	8.2	3.1	0	0.95	0	0.12	0	0.11	87.52

V. CONCLUSION

In this paper, we investigated a feature learning based approach by the DBN for handwritten Bangla character recognition problem. We also used traditional supervised learning approach to show the effectiveness of deep learning. The main focus of this paper is to demonstrate the power of unsupervised feature extraction and learning. In this approach, there is no need to create handcrafted features like loops, stroke, and curves, yet the recognition result is satisfactory. The training and test dataset we used is not sufficient with respect to other researches in English language (for example, the MNIST dataset [20]). If it was possible to use larger dataset for training then performance would be better. As a future work, this work can be applied on Bangla compound characters and Bangla speech processing.

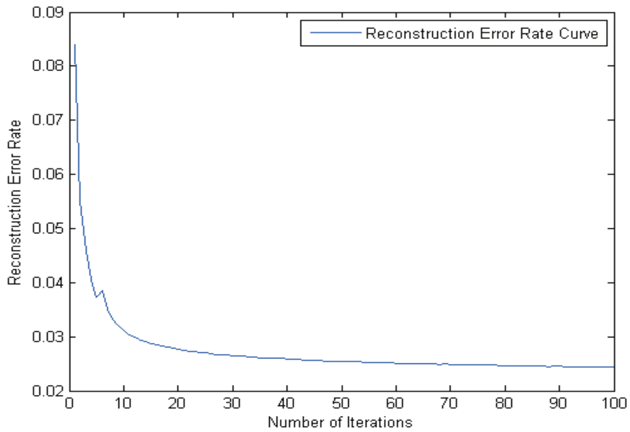

Figure 5. Reconstruction error rate curve in the training phase

ACKNOWLEDGMENT

We would like to thank Ujjwal Bhattacharya from Indian Statistical Institute for providing us the *ISI handwritten character dataset.*

REFERENCES

[1] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527-1554, Jul. 2006.

[2] Ofer Matan, H.S. Baird, J. Bromley, Christopher J.C. Burges, J.S. Denker, L.D. Jackel, Y. Le Cun, E.P.D. Stenard, W.D. Pednault, C.E. Satterfield,T.J. Thompson, "Reading handwritten digits: A ZIP code recognition system," *Computer,* vol. 25, no.7, pp. 59-63, Jul. 1992.

[3] S. H.B Shaikh, Nabendu Chaki, Marek Tabedzki, and Khalid Saeed,"A view-based approach for recognition of Bengali printed characters," in Proc. *World Congress on Nature & Biologically Inspired Computing,* Dec. 2009, pp. 1005-1009.

[4] Fahim Irfan Alam and Bithi Banik, "Offline isolated bangla handwritten character recognition using spatial relationships," in Proc. *IEEE International Conference on Informatics, Electronics & Vision,* May 2013, pp. 1-6.

[5] G. G. Rajput, and H. B. Anita,"Handwritten Script Recognition Using DCT, Gabor Filter, and Wavelet Features at Word Level." in *Proc. International Conference on VLSI, Communication, Advanced Devices, Signals & Systems and Networking,* 2013, pp. 363-372.

[6] Rajat Raina, Alexis Battle, Honglak Lee, Benjamin Packer, and Andrew Y. Ng, "Self-taught learning: transfer learning from unlabeled data," in Proc. *ACM 24th international conference on Machine learning*, Jun. 2007, pp. 759-766.

[7] Adam Coates, Blake Carpenter, Carl Case, Sanjeev Satheesh, Bipin Suresh, Tao Wang, David J. Wu, and Andrew Y. Ng, "Text detection and character recognition in scene images with unsupervised feature learning," in Proc. *IEEE International Conference on Document Analysis and Recognition*, Sep. 2011, pp. 440-450.

[8] Tijmen Tieleman and Geoffrey Hinton,"Using fast weights to improve persistent contrastive divergence," in Proc. *ACM 26th Annual International Conference on Machine Learning*, Jun. 2009, pp. 1033-1040.

[9] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams, "Learning representations by back-propagating errors," *Nature,* vol.323, no. 6088, pp. 533-536, 1986.

[10] A.F.R. Rahman,R. Rahman,M.C. Fairhurst, "Recognition of Handwritten Bengali Characters: a Novel Multistage Approach," *Pattern Recognition*, vol. 35, no. 5, pp. 997-1006, May 2002.

[11] T. K. Bhowmik, U.Bhattacharya and Swapan K. Parui, "Recognition of Bangla Handwritten Characters Using an MLP Classifier Based on Stroke Features," *Neural Information Processing*, vol. 3316, p.p. 814-819, 2004.

[12] Geoffrey E. Hinton, and Ruslan R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science,* vol. 313, no. 5786, pp. 504-507, Jul. 2006.

[13] Aiquan Yuan, Gang Bai, Po Yang, Yanni Guo, and Xinting Zhao, "Handwritten English Word Recognition based on Convolutional Neural Networks," in Proc. *IEEE International Conference on Frontiers in Handwriting Recognition,* Sep. 2012, pp. 207-212.

[14] Yann LeCun,, Koray Kavukcuoglu, and Clément Farabet, "Convolutional networks and applications in vision," in Proc. *IEEE International Symposium on Circuits and Systems,* Jun. 2010, pp. 253-256.

[15] M. Welling, M. Rosen-Zvi, and G. E. Hinton, "Exponential family harmoniums with an application to information retrieval," in *Proc. Neural Information Processing Systems*, Dec. 2005, pp. 1481-1488.

[16] A-R.Mohamed, Tara N. Sainath, George Dahl, Bhuvana Ramabhadran, Geoffrey E. Hinton, and Michael A. Picheny, "Deep belief networks using discriminative features for phone recognition," in Proc. *IEEE International Conference on*. *Acoustics, Speech and Signal Processing*, May 2011, pp. 5060-5063.

[17] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002.

[18] URL: http://www.isical.ac.in/~ujjwal/download/database.html

[19] Tapan Kumar Bhowmik, Pradip Ghanty, Anandarup Roy, and Swapan Kumar Parui, "SVM-based hierarchical architectures for handwritten Bangla character recognition." *International Journal on Document Analysis and Recognition (IJDAR),* vol. 12, no. 2, pp. 97-108, Jul. 2009.

[20] URL: http://yann.lecun.com/exdb/mnist/