Development of English Handwritten Recognition Using Deep Neural Network

Teddy Surya Gunawan*[1], Ahmad Fakhrur Razi Mohd Noor[2], Mira Kartiwi[3]
[1,2] Department of Electrical and Computer Engineering, Kulliyyah of Engineering, Malaysia
[3] Department of Information Systems, Kulliyyah of ICT, International Islamic University Malaysia, Malaysia

Article Info	ABSTRACT
	Due to the advanced in GPU and CPU, in recent years, Deep Neural Network (DNN) becomes popular to be utilized both as feature extraction and classifier. This paper aims to develop offline handwritten recognition system using DNN. First, two popular English digits and letters database, i.e. MNIST and EMNIST, were selected to provide dataset for training and testing phase of DNN. Altogether, there are 10 digits [0-9] and 52 letters [a-z, A-Z]. The proposed DNN used stacked two autoencoder layers and one softmax layer. Recognition accuracy for English digits and letters is 97.7% and 88.8%, respectively. Performance comparison with other structure of neural networks revealed that the weighted average recognition rate for patternnet, feedforwardnet, and proposed DNN were 80.3%, 68.3%, and 90.4%, respectively. It shows that our proposed system is able to recognize handwritten English digits and letters with high accuracy.

Corresponding Author:

Teddy Surya Gunawan,
Department of Electrical and Computer Engineering,
Kulliyyah of Engineering, Malaysia..
Email: tsgunawan@iium.edu.my

1. INTRODUCTION

Handwritings are the most common standard and regular medium that human beings use in communicating. It is also an effective and efficient way to record information even with the introduction of new technologies, like swiftkey keyboard, sound command, etc. Generally, handwriting recognition system is a mechanism that used to recognize human handwritten in any languages either from scanned handwritten image or real time handwriting using stylus pen on electronic device which also can be called as offline and online handwriting respectively [1]. Besides, the application of this system can be categorized into three which are numeral, character and cursive word. It is widely used in numerous applications such as language translation, bank cheques and keyword spotting [2, 3].
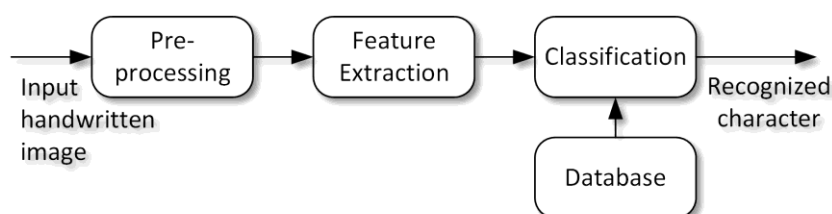

Figure 1. A Typical Handwritten Recognition System

As shown in Figure 1, the common processes of handwriting recognition system are image acquisition, preprocessing, segmentation, feature extraction and classification [2]. Image acquisition is the first step to get an image form of handwritten that will act as input to preprocessing by using scanner. The preprocessing step is to remove noise or distortions of the scanned image which usable for further process [3, 4]. One of the process in preprocessing is thresholding which convert scanned image into binary image. Next, segmentation is used to divide each word into sub-images where this stage is very important step particularly for continuous handwriting in order to extract features from each image of character which will be perform in feature extraction process [1].

Then proceeds with feature extraction where it will extract every characteristic of the features in each image. This feature is very useful for classification in the last step. There are many classification techniques such as K-Nearest Neighbour (KNN), Neural Network and Support Vector Machine (SVM) where these classifiers have different approach to recognize the image [4]. Generally, most researcher evaluate the performance of the system based on classification accuracy [5, 6].

Although many paper has been conducted on offline handwritten recognition, but the use of Deep Neural Network (DNN) is still in the early stage. Therefore, the objective of this paper is to develop offline handwritten recognition using DNN. We will use two popular database, including MNIST [7] and EMNIST [8] due to the clean data provided. Thus, the algorithms that will be used in this project start with feature extraction where we will use image pixel as feature input to classifier [9] as DNN will handle both feature extraction and classifier. Finally, we will evaluate the performance of this proposed project based on confusion matrix to find the classification accuracy.

2.   DESIGN OF PROPOSED HANDWRITTEN RECOGNITION SYSTEM

Figure 2 shows the proposed handwritten recognition system. The processing step includes image thresholding, character thinning using morphological operation, slant correction, and finally image segmentation. The pixel value from image segmentation will be treated as the incoming input to the DNN, in which both feature extraction and classification will be conducted.
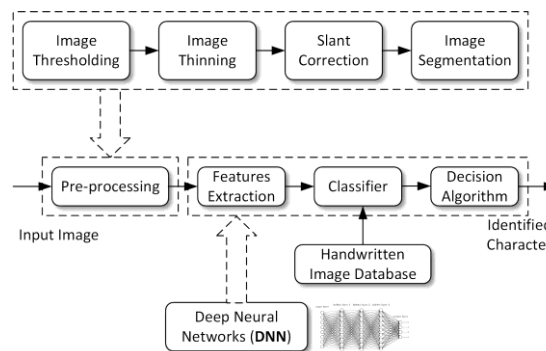
Figure 2. Proposed Handwritten Recognition System

2.1   Feature Extraction and Classifier Using Deep Neural Network (DNN)

In this project, we only used the pixel value of each segmented images as feature input to Deep Neural Network (DNN). The image size from dataset are 28-by-28 pixels in grayscale which means one digit in one image. When we are working with images, the features we use is the raw pixel values. Since the image has 28x28 pixels as illustrated in Figure 3 thus it equal to 784 pixels, so we have 784 input features for each image.

In 2006, deep belief network was introduced by [10] which forms a DNN model composed of Restricted Boltzmann Machines (RBM). It is trained in unsupervised fashion, one layer at a time from the lowest to the highest layer. Another variation, i.e. deep feed-forward networks, were effectively trained using the same idea by first pre-training each layer as a RBM, then fine-tuning by backpropagation [11]. Nowadays, ANNs with deep structure are trained on powerful GPU machines, overcoming both resources and training time limits.

Other types of DNN includes recurrent neural networks (RNN) and conventional neural networks (CNN). An RNN is similar to an ANN except that it allows a self-connected hidden layer that associates with a time delay. A CNN is similar to an ordinary feedforward neural network as the output of each layer is a

combination of the input, weight matrix, and the bias vector followed by a non-linear transformation. However, CNN has local connectivity, in which a CNN uses some small filters, slides it across all sub-regions of the input matrix and aggregates all results. Means that, it takes advantages of the convolution operation between the filters and the input. A CNN is normally consisted of several types of layers, including convolutional layer, pooling layer, and fully-connected layer.
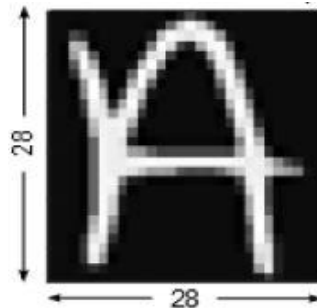

Figure 3. Example of Pixel Value of an Image as Feature Input to DNN

2.2 Selected Handwritten Image Databases

The database used for this research have two categories which are digit and alphabet. Digit database is from the MNIST database of handwritten digits where this database contains 60000 examples of training set and 10000 examples of test set [4]. The digits are all a uniform size and are centred in the image. Each image is centred in a 28x28 image by computing the centre of mass of the pixels, and translating the image so that the centre of mass is at the centre of the image. Each image is a binary image of a handwritten digit and consists of 30,000 patterns from two databases, one from Census Bureau employees and the other from high-school students, from about 250 different writers in total. The test set consists of 5,000 patterns from each database. The datasets are labelled to enable easy verification. The sample of training image for MNIST as shown in Figure 4.

However, we only used one type of datasets from the list which is EMINST Letters for this proposed project. The EMNIST Letters dataset merges a balanced set of the uppercase and lowercase of handwritten letters into a single 26-class task. The training sample for our proposed project is 124800 and test sample is 20800. Some of training images sample are shown in Figure 5.
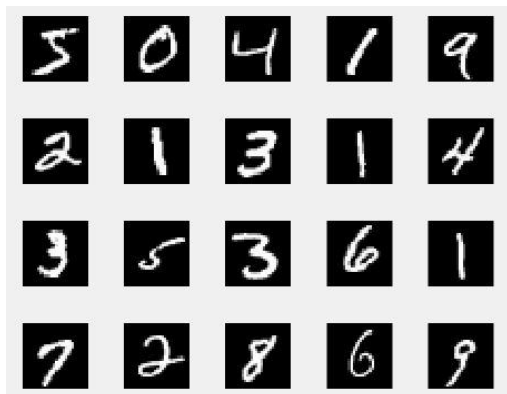

Figure 4. MNIST Training Image Samples

Figure 5. EMNIST Training Image Samples

In addition, letter database used for this project is from Extended Modified NIST (EMNIST). Both EMNIST and MNIST was derived from NIST Special Database 19 [8]. The conversion of images from the original used the same methodology as in [4]. There are six different splits provided in this dataset. A short summary of the dataset is provided as follows:
- EMNIST ByClass: 814,255 characters. 62 unbalanced classes.

- EMNIST ByMerge: 814,255 characters. 47 unbalanced classes.
- EMNIST Balanced: 131,600 characters. 47 balanced classes.
- EMNIST Letters: 145,600 characters. 26 balanced classes.
- EMNIST Digits: 280,000 characters. 10 balanced classes.
- EMNIST MNIST: 70,000 characters. 10 balanced classes.

3. IMPLEMENTATION OF DEEP NEURAL NETWORK

In this proposed project, Deep Neural Network (DNN) is used as feature extraction and classifier of the handwritten system. We conducted training and testing phase of DNN on the different samples.

3.1 Training Phase of the DNN

Deep Neural Network (DNN) is a network that consist of many hidden layers with different number of neuron in each layer. For this research, a stacked autoencoders for handwriting recognition is used to train multiple layers. The number hidden layer used is three, including two hidden layers and one softmax layer in which these three layers will be stacked together in order to form deep network. The first and second layer will be trained without using label from training data which means unsupervised fashion [12]. Softmax layer differ from hidden layer in which we trained this layer with supervised fashion using labels from training dataset.

This autoencoder uses regularizers to learn a sparse representation in the first layer. L2WeightRegularization controls the impact of an L2 regularizer for the weights of the network (and not the biases). SparsityRegularization controls the impact of a sparsity regularizer, which attempts to enforce a constraint on the sparsity of the output from the hidden layer. Note that, this is different from applying a sparsity regularizer to the weights. Sparsity Proportion is a parameter of the sparsity regularizer. It controls the sparsity of the output from the hidden layer. A low value for SparsityProportion usually leads to each neuron in the hidden layer specializing by only giving a high output for a small number of training examples. For example, if SparsityProportion is set to 0.1, this is equivalent to saying that each neuron in the hidden layer should have an average output of 0.1 over the training examples. This value must be between 0 and 1. The ideal value varies depending on the nature of the problem. The configuration used for digits and letters that affect the performance of our test are illustrated in Table 1.

Table 1. Configuration of DNN

Class	Number of character	Number of hidden layer	Number of nodes in input layer	Number of neurons in 1st hidden layer	Number of neurons in 2nd hidden layer	Number of neurons in 3rd hidden layer
Digits	10	3	784	300	50	10
Letters	26	3	784	300	50	27

We used smaller number of neurons in first hidden layer compared to input of the DNN. Since our digits image sample will have 784-by-60000 while letters image sample consist of 784-by-124800 as input to DNN, we set number of neurons for first hidden layer is 300. Second layer will have 100 while 10 or 27 for softmax layer. Less number of neurons will make the autoencoder learns smaller and compressed representation of the input of every layer. In this training process, the input of each layer will use extract feature from previous encoder as training data to train the layer.
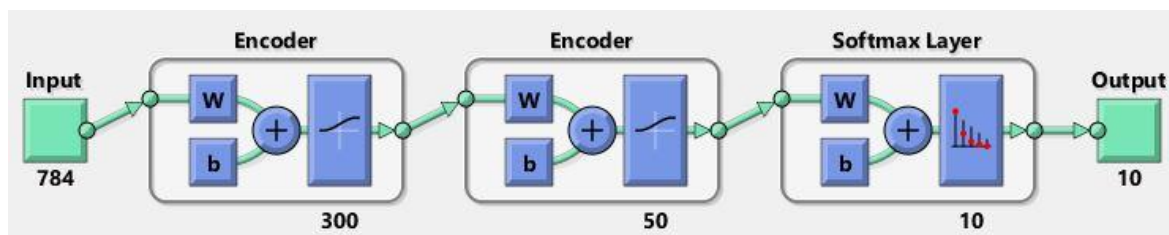

Figure 6. Stacked Layers of DNN

After trained three separate layers of a DNN, we will stack those three layers together to form a deep network, as shown in Figure 6. Furthermore, we compute the results with testing dataset using the full deep network formed. In order to increase the performance of deep network, we tune the deep network by retraining it using training dataset in supervised fashion which means including training label data.

3.2 Testing Phase of the DNN

Testing is the last part of handwriting recognition in order to evaluate the DNN. To test the network, we need to have test dataset along with test label of images. With the full deep network formed and trained, we test the network using test dataset. Based on the test, we compute the results of the system such as recognition accuracy, performance and percentage error. Results of handwriting recognition system using DNN divided into two since different in database. One for digit recognition and the other for letter recognition. Our test performance is based on the confusion matrix.

4.   RESULTS AND DISCUSSION

In this section, the experimental setup along with the testing of the trained DNN on the selected dataset will be elaborated.

4.1   Experimental Setup

All experiment will be conducted using a Lenovo G500s laptop with the following specifications, Intel Pentium 2020M, 8 GB RAM, Windows 10 Pro operating systems. The software used for this experiment is MATLAB R2017a. This version was supported with Neural Network Toolbox 10.0 where the toolbox is very important and useful for this project. Besides, R2017a also equipped with Image Processing 10.0 and Computer Vision Toolbox 7.3. The goal of both toolboxes is to simplify all processes for preprocessing, feature extraction, as well as DNN.

4.2   DNN Testing Performance on Digits and Letters Classes

Using confusion matrix, we obtained the performance metrics of handwriting recognition using DNN as shown in Table 2. It was found that the recognition rate for digits is higher than letters. This could be due to smaller number of digits (10) compared to letters (52). The weighted average of the recognition rate will be around 93.4%.

Table 2. Performance Evaluation of the Trained DNN on Testing Dataset

Class	Accuracy	Performance	Percentage of error	Execution Time
Digits	98.5%	0.0109	1.5%	3.395769 (s)
Letters	88.8%	0.0313	11.2%	7.740534 (s)

4.3   Comparison with other ANN structures

In this section, we presented the performance evaluation of different ANN structures, including patternet, feedforwardnet, and proposed DNN. The patternet and feedforwardnet were configured with 1 hidden layer and 10 number of neurons in the hidden layer. The weight average of the recognition rate is calculated to take into account different number of Digits (10) and Letters (52). Table 3 shows the performance comparison of three ANN structures, in which it can be seen that our proposed algorithm performs better compared to the other ANN structure.

Table 3. Recognition Rate of Three ANN Structures

ANN structure	Class	Accuracy	Weighted Average
patternet	Digits	92.2%	80.3%
	Letters	77.9%	
feedforwardnet	Digits	84.7%	68.3%
	Letters	65.2%	
Proposed DNN	Digits	98.5%	90.4%
	Letters	88.8%	

5.   CONCLUSIONS AND FUTURE WORKS

In this project, the handwriting recognition system is proposed by using image pixels as feature input and DNN as feature extraction and classifier. The segmented image size is 28 by 28 pixels producing 784 feature input. The DNN structure is stacked autoencoders with 300 and 50 neurons and softmax layer with 10 neurons. Both MNIST and EMNIST handwritten image database were used in the performance evaluation. The weighted recognition rate for both digits and letters classes for our proposed DNN, patternet, and feedforward were 90.4%, 80.3%, and 68.3%. Future works include different feature extraction methods, different deep neural network configuration, and different database.

REFERENCES

[1]   A. Priya, S. Mishra, S. Raj, S. Mandal, S. Datta, *"Online and offline character recognition: A survey,"* in Communication and Signal Processing (ICCSP), 2016 International Conference on, pp. 0967-0970, 2016.

[2]   N. Sharma, T. Patnaik, B. Kumar, "Recognition for Handwritten English Letters: A," *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 2, 2013.

[3]   U.-V. Marti, H. Bunke, "The IAM-database: an English sentence database for offline handwriting recognition," *International Journal on Document Analysis and Recognition*, vol. 5, pp. 39-46, 2002.

[4]   Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, *"Gradient-based learning applied to document recognition,"* Proceedings of the IEEE, vol. 86, pp. 2278-2324, 1998.

[5]   D. Cireşan, U. Meier, *"Multi-column deep neural networks for offline handwritten Chinese character classification,"* in Neural Networks (IJCNN), 2015 International Joint Conference on, pp. 1-6, 2015.

[6]   A. Yuan, G. Bai, L. Jiao, Y. Liu, *"Offline handwritten English character recognition based on convolutional neural network,"* in Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on, pp. 125-129, 2012.

[7]   Y. LeCun, C. Cortes, C. J. C. Burges, "The MNIST Database of Handwritten Digits," Retrieved on: 10 December 2017.

[8]   P. J. Grother, K. K. Hanaoka, "NIST Special Database 19: Handprinted Forms and Characters Database," 2016.

[9]   C. Kaensar, *"A comparative study on handwriting digit recognition classifier using neural network, support vector machine and k-nearest neighbor,"* in The 9th International Conference on Computing and InformationTechnology (IC2IT2013), pp. 155-163, 2013.

[10]   G. E. Hinton, S. Osindero, Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, pp. 1527-1554, 2006.

[11]   G. E. Hinton, R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, pp. 504-507, 2006.

[12]   S. Narejo, E. Pasero, F. Kulsoom, "EEG Based Eye State Classification using Deep Belief Network and Stacked AutoEncoder," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 6, pp. 3131, 2016.

BIOGRAPHIES OF AUTHORS

Teddy Surya Gunawan received his BEng degree in Electrical Engineering with cum laude award from Institut Teknologi Bandung (ITB), Indonesia in 1998. He obtained his M.Eng degree in 2001 from the School of Computer Engineering at Nanyang Technological University, Singapore, and PhD degree in 2007 from the School of Electrical Engineering and Telecommunications, The University of New South Wales, Australia. His research interests are in speech and audio processing, biomedical signal processing and instrumentation, image and video processing, and parallel computing. He is currently an IEEE Senior Member (since 2012), was chairman of IEEE Instrumentation and Measurement Society – Malaysia Section (2013 and 2014), Associate Professor (since 2012), Head of Department (2015-2016) at Department of Electrical and Computer Engineering, and Head of Programme Accreditation and Quality Assurance for Faculty of Engineering (since 2017), International Islamic University Malaysia. He is Chartered Engineer (IET, UK) and Insinyur Profesional Madya (PII, Indonesia) since 2016.

Ahmad Fakhrur Razi Mohd Noor has completed his B.Eng. (Hons) degree in Communication Engineering from International Islamic University Malaysia (IIUM) in 2018. His research interests are in signal processing and artificial intelligence and affective computing. Currently, he is working as engineer at Intel, Penang.

Mira Kartiwi completed her studies at the University of Wollongong, Australia resulting in the following degrees being conferred: Bachelor of Commerce in Business Information Systems, Master in Information Systems in 2001 and her Doctor of Philosophy in 2009. She is currently an Associate Professor in Department of Information Systems, Kuliyyah of Information and Communication Technology, International Islamic University Malaysia. Her research interests include electronic commerce, data mining, e-health and mobile applications development.