Optimization of Arabic Handwritten digits recognition using CNN

Hamdy Amin Morsy

Abstract— Arabic handwritten digits recognition (AHDR) is one of the most challenging convolutional neural networks (CNN) applications due to the difficult nature of Arabic digits and similarity of Arabic digits with other symbols in other languages. In this research paper, we will introduce a new technique for AHDR using CNN and entropy function as an object function. our technique introduced minimum error rate and maximum accuracy compared to other techniques.

Index Terms— Neural Networks, Natural Language Processing, Machine Learning, Image Processing, Object Recognition, Deep Learning, Pattern Recognition

1 Introduction

Convolutional neural network (CNN) works basically on visual image. It processes the image pixels and extract the features of the input image to have sufficient information so that it can predict the input object at the output. The CNN has many applications, one of the most important applications is the handwritten digits recognition (HDR). The inputs are images of 10 digits from 0 to 9. Each image is treated as pixels which have values changes with the changing of the digits. The convolutional neural networks consist of the input layer, output layer and the hidden layer which can be one or more layers sometimes maybe tens or hundreds of hidden layers. Each layer has many neurons or perceptrons, each one of the them is connected to the neurons of the previous and the next layers. For digit recognition, the number of neurons in the input layers is the number of pixels of the input image and the number of neurons of the output layer is the number of digits to be recognized which in this case is 10 digits.

Numerals have many different forms for different writing systems. The Chinese numerals have three different forms, the Arabic numerals, the roman are all have different writing systems numerals. There are more than 30 different numerals in different writing systems, which needs a different neural network for each different writing system [1]. The Arabic numeral writing system is used by 422 million (2020 stats.) people in the middle east, the Arabic handwritten digits require more extensive investigation to achieve perfect recognition of the Arabic numerals. Little work has been done on these Arabic numerals of digits recognition. Those digits that can be collected from scanned documents and images are considered to be offline dataset. The scanned images can be scaled up or down to have all dataset with the same dimension [2], [3]. Online datasets can be considered as a system that processes the input as digit by digit. In the case of offline dataset, a large number of scanned documents which represent the Arabic handwritten digits are processed into the neural network system to have good prediction at the output [4].

The convolutional Neural Network is a subset of the deep learning neural networks. It is a some of neurons connected together like the biological brain in the human. Each interconnected neuron carries a signal from a neuron in the present layer to the neuron in the next layer. These layers are connected through neurons of each layer from the input to the output. The concept of the neural network is based mainly on the weight of each interconnected neuron and adding some bias for each neuron to optimize the output of this neuron to good perfect match with the input image.

The handwritten digits recognition in general has many difficulties regarding the changes of writing from one individual to another. Also, the person can't keep exact way of writing for repeated digits. Feature extraction becomes hard as the variability of input dataset for the same digit is altered. There are many techniques applied to handwritten digits recognition such as, support vector machine (SVM), K nearest neighbor (KNN) and convolutional neural networks (CNN). Also, in convolutional neural networks some algorithms are applied to substantially reach perfect matches with the input dataset. Stochastic gradient descent (SGD) algorithm is utilized in CNN networks to predict the weights and biases for each perceptron [5]. More specific in CNN networks, an objective function is needed to provide an indicator for the output accuracy as a function for the weights and biases for each individual neuron.

In this paper we will present the cross-entropy function as an objective function to measure the output accuracy as a function of the weights and biases in the network and we will also present a new model utilizing the convolutional neural network (CNN). The paper will be organized as follows: Section II will discuss the convolutional neural networks and the different systems applied to handwritten digits recognition. Section III will discuss new model and the comparison with current approaches. Finally, the conclusion will be presented in section IV.

• *Hamdy Amin Morsy is an Assistant Professor at Department of Electronics and Communications Engineering, Faculty of Engineering, Helwan University, Cairo, Egypt*
Email: hamdy_morsy@h-eng.helwan.edu.eg

ISSN 2229-5518

2 CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural networks (CNN) is a class of deep learning neural networks. CNN network is mostly applied to analyzing and processing of images and videos recognition, image classification, natural language processing, medical imaging. It is also known as shift invariant or space invariant artificial neural networks based on weights and invariant characteristics [6], [7]. The CNN networks utilize the regularized multilayer perceptrons in which each perceptron in each layer is connected with all perceptrons of previous and next layer. They are fully connected networks, each connection between two perceptron has a weight multiplied by the perceptron activation value of the previous connected perceptron [8]. There is a value added to each weight which is called a bias resulting in the activation value of the perceptron. The activation of a perceptron can be calculated as follows:

$$a = wx + b \tag{1}$$

Where a is the activation, w is the weight of the input value x and b is the bias [9], [10]. Equation (1) describes the effects of the factor on the input x and the variables w and b are optimized continuously to achieve the desired output.

$$a_m = \sum_{n=1}^{N} w_{nm} x_n + b_m \tag{2}$$

$$\begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_M \end{bmatrix} = \begin{bmatrix} w_{11} & w_{21} & \dots w_{N1} \\ w_{12} & w_{22} & \dots w_{N2} \\ & & \dots \\ & & \dots \\ & & \dots \\ w_{1M} & w_{2M} & \dots w_{NM} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_N \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_M \end{bmatrix} \tag{3}$$

Equation (2) shows the activation value of a perceptron m due to the effects of the input $x_1, x_2, \dots x_N$ with weights $w_{1m}, w_{2m}, \dots, w_{Nm}$ and bias b_m. The general form of (2) is presented in (3) which shows the relationship between any two layers adjacent to each other with N neurons at the first layer and M neurons at the second layer. The subscripts in the weight matrix looks inverted to show the direction from first layer to second layers as shown in **Error! Reference source not found.**. The weight matrix has M rows and N columns and the bias is Mx1 matrix, a is also an Mx1 matrix. The output of **Error! Reference source not found.** show the Arabic digits from 0 to 9.

The output layer has number of neurons corresponds to each of the input digit and there is a threshold to decide which one is the output. When the input neurons take values such as a zero or a one, a small change in the weight or bias will affect the whole neural network. Therefore, a sigmoid function, $\sigma = (1 + \exp(-a))^{-1}$ -where a is the activation- is proposed to allow for different values between 0 and 1 in the input $x_1, x_2, \dots, x_N$. Any changes in the weight and bias will make a small change in the neural network. The sigmoid function is the smoothed curve of the step function which provides values between 0 and 1. The step function produces values of only 0 or 1.

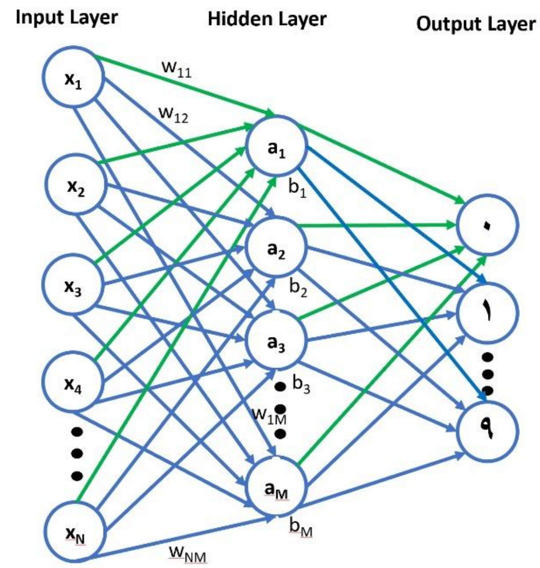

Fig. 1. The neural networks three-layer structure

The transition of data from the input to the output through the hidden layer is called feedforward neural networks. In this case, information is fed forward till reach the output. In our case, the output layer has 10 neurons numbered from 0 to 9 and one neuron is turned on one at a time to refer to the corresponding number entered to the neural network.

The neural network system needs a training dataset to train the system for all the output digits required and a testing dataset to test the system for its accuracy against new dataset inputs. The weight and bias of the network need to be changed continuously until a minimum objective function is achieved. Stochastic gradient descent (SGD) and adam algorithms are two iterative algorithms used to optimize the objective function of the neural network [11]. There are many functions to calculate the output objective such as quadratic objective function and cross entropy objective function [12].

$$O_q = \frac{1}{2n} \sum_x \| y(x) - a(x) \|^2 \tag{4}$$

Equation (4) describes the objective function O_q as an average and a function of the desired output vector y(x) and the activation vector a(x). The goal is to find minimum of w_{nm} and b_m to minimize the objective function. The derivative of w and b can bring global minima to equation (4) as a(x) is a function of w_{nm} and b_m. This will be done using stochastic gradient descent algorithm and the new values for w_{nm} and b_m will be given as follows:

$$\begin{aligned} w'_{nm} &= w_{nm} - r \frac{\partial O_q}{\partial w_{nm}} \\ b'_m &= b_m - r \frac{\partial O_q}{\partial b_m} \end{aligned} \tag{5}$$

Equation (5) describes the new values of w_{nm} and b_m to find the global minima and find the minimum object function O_q. r is the learning rate which determines the step by which the w_{nm} and b_m can be changed. The cross entropy object function is

ISSN 2229-5518

given by [13], [14]:

$$O_e = -\frac{1}{n}\sum_x \left[\begin{array}{c} y(x) \ln a(x) + \\ (1 - y(x)) \ln(1 - a(x)) \end{array} \right] \qquad (6)$$

Equation (6) is used as a formula for calculating the minimum values of w_{nm} and b_m to achieve minimum errors. The partial derivative of w_{nm} and b_m performed for all inputs x to find the minimum w_{nm} and b_m of the neural network for this particular input of the 10 digits as shown in (7).

$$\begin{gathered} \frac{\partial O_e}{\partial w_{nm}} = \frac{1}{n}\sum_x a_m(x)(a_m(x) - y(x)) \\ \frac{\partial O_e}{\partial b_m} = \frac{1}{n}\sum_x (a_m(x) - y(x)) \end{gathered} \qquad (7)$$

4 Proposed Algorithm and Results

In this research paper, we will adapt the cross entropy objective function to reach the minimum output error with minimum w_{nm} and b_m. The proposed algorithm is given below:

- ***Input x as pixels values.***
- ***Calculate the weight and bias of each link (feedforward).***
- ***Calculate the output error.***
- ***Recalculate the weigh and bias (backpropagation).***
- ***Output the corresponding digit.***

This is a general technique for finding the minimum output of the input dataset. The proposed algorithm starts by diving the Arabic digits into 28 by 28 grayscale images with 784 pixels. Each pixel corresponds to one input x to the neural network as the Arabic handwritten digits shown in **Error! Reference source not found.**.

Fig. 2. The Arabic digits.

The handwritten digits dataset is collected from some students and manipulated with data augmentation to increase the dataset size. The dataset is divided into two groups the training dataset which has 60,000 digits and the testing dataset which has 10,000 digits.

Error! Reference source not found. shows the accuracy of the system versus the number of epochs for different values of learning rates. As can be seen from the figure, the learning rate r=0.01 shows the maximum accuracy compared to other values as illustrated in **Error! Reference source not found.**.

The proposed system is trained for 300 epochs with learning r =0.01. The system shows very high accuracy for different values of epochs as shown in Fig. 5. Also, the error rate of the test data decreases with increasing the number of epochs as illustrated in Fig. 6. The error rate of the training data is lower than the error rate of the test data as shown in Fig. 7.

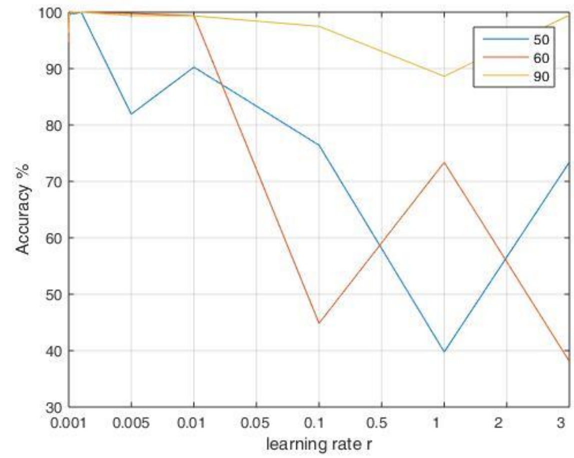

Fig. 3. Accuracy in percentage vs learning rate r.

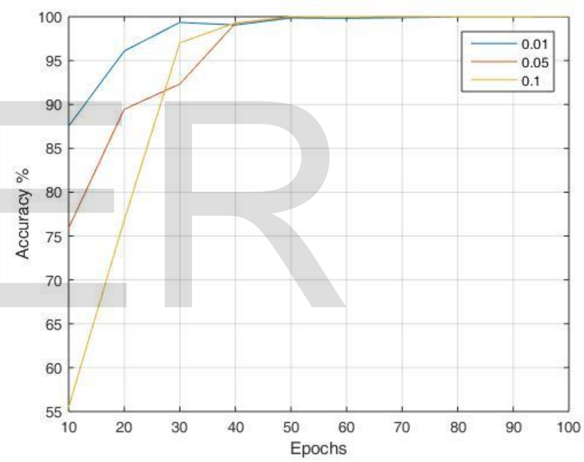

Fig. 4. Accuracy in percentage vs number of epochs.

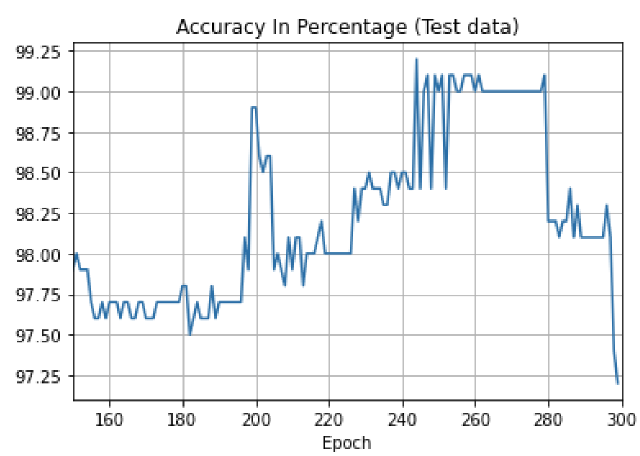

Fig. 5. Accuracy in percentage vs number of epochs (Test data).

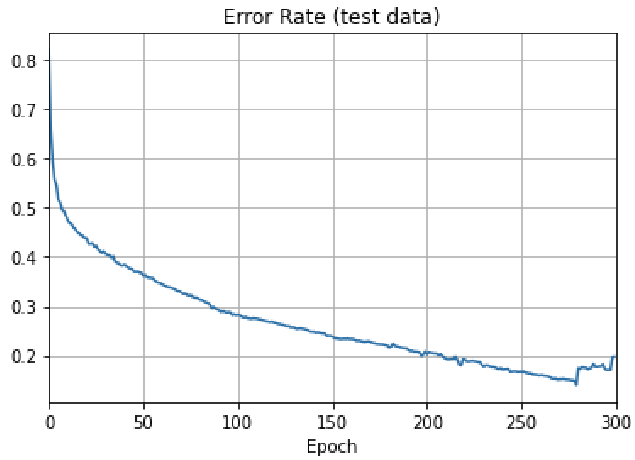

Fig. 6. Error rate vs number of epochs (Test data).

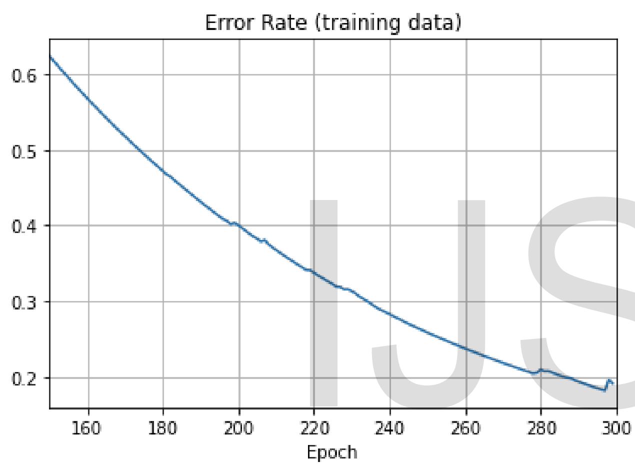

Fig. 7. Error rate vs number of epochs (Training data).

A comparison of error rates of various techniques is shown in Table 1. All these techniques introduced solutions for Arabic handwritten recognition with different error rates as shown in the table. The first one used fuzzy C mean with SVM machine to introduce three level classifier [15]. They used minimum dataset (2106 training data, 1404 test data) which results in 12% error rate. The second algorithm introduced a techniques based on dynamic Bayesian network (DBN) [16]. They utilize DCT coefficients to classify Arabic handwritten digits with ADBase dataset. They got less improvement of 15% error rate. The third technique used two systems for Arabic handwritten digits recognition (AHDR) [17]. The error rate is greatly improved with 5%. The rest techniques introduced improvements in error rates [18]–[21]. Our proposed shows enhancement of 1.75% error rate. In fact, this is the average value of error rate. The minimum error rate is 0.8 % which introduce enhancements compared to other systems.

TABLE 1
COMPARISON OF THE ERROR RATE OF VARIOUS TECHNIQUES.

Authors	Database	Train, Test	Error rate %
Takruri [21]	Public	2106, 1404	12
AlKhateeb [22]	ADBase	60,000, 10,000	15
Salameh [23]	Fonts	1000, 1000	5
Melhaoui [24]	Private	600, 400	1
Selvi [25]	Private	Unknown	4
Mahmoud [26]	Private	14784, 6336	0.15, 2.16
El-Sawy [27]	MADBase	60,000, 10,000	1, 12
Proposed	Private	50,000, 10,000	1.75

5 CONCLUSIONS

There are many approaches to Arabic handwritten recognition based on convolutional neural network. Our approach based on entropy function as an objective function introduced minimum error rate compared to other existing techniques. The technique is based on achieving optimum values for number of hidden layer neurons and learning rate.

REFERENCES

[1] G. Ifrah, *The universal history of computing*. John Wiley & Sons, Inc., 2000.

[2] H. A. Morsy, “Performance analysis of the effects of non-adaptive image scaling on image edges,” *Int. J. Recent Technol. Eng.*, vol. 7, no. 6, pp. 1692–1696, 2019.

[3] H. A. Morsy, “Comparison of Commonly Used Non-Adaptive Image Scaling Techniques,” *Ciit Digit. Image Process.*, vol. 10, no. 9, pp. 177–180, 2018, [Online]. Available: https://www.researchgate.net/publication/344324331_Comparison_of_commonly_used_non-adaptive_image_scaling_techniques.

[4] “Numerals in many different writing systems.” https://omniglot.com/language/numerals.htm.

[5] Y. Lecun, L. Bottou, Y. Bengio, and P. Ha, “Gradient-Based Learning Applied to Document,” *Proc. IEEE*, no. November, pp. 1–46, 1998, doi: 10.1109/5.726791.

[6] W. Zhang, “Shift-invariant pattern recognition neural network and its optical architecture,” 1988.

[7] W. Zhang, “Parallel distributed processing model with local space-invariant interconnections and its optical architecture,” *Appl. Opt.*, vol. 29, no. 32, pp. 4790–7, 1990, doi: 1990ApOpt..29.4790Z.

[8] M. Nielsen, *Neural Networks and Deep Learning*. 2018.

[9] M. Ramzan *et al.*, “A survey on using neural network based algorithms for hand written digit recognition,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 9, pp. 519–528, 2018, doi: 10.14569/ijacsa.2018.090965.

[10] O. I. Abiodun *et al.*, “Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition,” *IEEE Access*, vol. 7, no. February 2017, pp. 158820–158846, 2019, doi: 10.1109/ACCESS.2019.2945545.

[11] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.

[12] H. A. Morsy, “Developing a New CCN Technique for Arabic Handwritten Digits Recognition,” *Int. J. Recent Technol. Eng.*, vol. 9, no. 3, pp. 520–524, 2020, doi: 10.35940/ijrte.C4588.099320.

[13] C. Arndt, *Information Measures Information and its Description in Science and Engineering*. 2001.

[14] R. M. Gray, *Entropy and information theory*. 2011.

[15] A. A.-H. M. Takruri, R. Al-Hmouz, "A three-level classifier: fuzzy C means, support vector machine and unique pixels for Arabic handwritten digits," in *World Symposium on Proceedings of Computer Applications & Research (WSCAR)*, 2014, p. 1—5, doi: 10.1109/WSCAR.2014.6916798.

[16] J. H. Alkhateeb and M. Alseid, "DBN - Based learning for Arabic handwritten digit recognition using DCT features," *2014 6th Int. Conf. Comput. Sci. Inf. Technol. CSIT 2014 - Proc.*, no. September, pp. 222–226, 2014, doi: 10.1109/CSIT.2014.6806004.

[17] M. Salameh, "Arabic digits recognition using statistical analysis for end/conjunction points and fuzzy logic for pattern recognition techniques," *World Comput. Sci. Inf. Technol. J*, vol. 4, no. 4, pp. 50–56, 2014.

[18] O. EL Melhaoui, M. El Hitmy, and F. Lekhal, "Arabic Numerals Recognition based on an Improved Version of the Loci Characteristic," *Int. J. Comput. Appl.*, vol. 24, no. 1, pp. 36–41, 2011, doi: 10.5120/2912-3830.

[19] P. P. Selvi, T. Meyyappan, "Recognition of Arabic numerals with grouping and ungrouping using back propagation neural network," in *Pattern Recognition, Informatics and Mobile Engineering (PRIME)*, 2013, p. 322—327, doi: 10.1109/ICPRIME.2013.6496494.

[20] S. A. Mahmoud, "Arabic (Indian) handwritten digits recognition using Gabor-based features," in *Innovations in Information Technology, 2008. IIT 2008.*, 2008, p. 683—687, doi: 10.1109/INNOVATIONS.2008.4781779.

[21] M. L. Ahmed El-Sawy, Hazem EL-Bakry, "CNN for Handwritten Arabic Digits Recognition Based on LeNet-5," 2017, doi: 10.1007/978-3-319-48308-5.