American Journal of Science, Engineering and Technology
2022; 7(3): 57-61
http://www.sciencepublishinggroup.com/j/ajset
doi: 10.11648/j.ajset.20220703.11
ISSN: 2578-8345 (Print); ISSN: 2578-8353 (Online)

Performance Analyses of the Eastern Arabic Hand Written Digits Recognition Using Deep Learning

Hamdy Amin Morsy

Electronics and Communications Engineering Department, Faculty of Engineering, Helwan University, Cairo, Egypt

Email address:
hamdy_morsy@h-eng.helwan.edu.eg

To cite this article:
Hamdy Amin Morsy. Performance Analyses of the Eastern Arabic Hand Written Digits Recognition Using Deep Learning. *American Journal of Science, Engineering and Technology*. Vol. 7, No. 3, 2022, pp. 57-61. doi: 10.11648/j.ajset.20220703.11

Received: July 6, 2022; **Accepted**: July 20, 2022; **Published**: July 28, 2022

Abstract: Deep convolutional neural networks are one of the most promising techniques to recognize objects, letters and digits. The recognition of handwritten digits is of great importance in many applications such as digits as password on cell phones, simplifying teaching children numbering systems. The Arabic handwritten digits recognition (AHDR) is an example of using deep learning in recognizing Arabic handwritten digits. The Arabic digits systems or Hindu digits are difficult to process using object recognition due to the similarities of digits to each other, which is totally different from the most popular English language digits as the English numbers are more variant than Arabic numbering systems. The Deep convolutional neural networks techniques have many different layers properties which depends on number of neurons, filter dimensions and strides which are called hyperparameters to achieve higher performance in recognizing the Eastern Arabic digits more than other techniques. The Eastern Arabic digits system has some varieties than other number systems which makes the recognition of handwritten digits are more challenging than other numbering systems. In this paper, multi hidden layers using deep convolutional networks will be applied to Arabic handwritten digits recognition. This technique outperforms other techniques with respect to minimum cost function and maximum accuracy compared to other techniques.

Keywords: Deep Learning, Convolutional Networks, Machine Learning, Object

1. Introduction

A convolutional neural network (CNN) is simply a network which receives inputs and processes them to recognize image objects, classify the image or analyse the image [1, 2]. The input is the image with pixels representing by the input layer and the output are classes represented by the output layer, the in between layers of the input layer and output layer are represented by some of hidden layers. Each layer in the input and output layers has number of neurons which equals to number of pixels in the input and number of classes at the output. The number of neurons in each layer are not equal to each other. Each subsequent layer may have neurons less than or equal to the previous one.

The CNN in its simplest form has three layers namely: input layer, output layer and a single hidden layer in between. The input layer is mostly a visual image which has many features that varies from one class to another. The images for the same class can vary a little bit from each other. This brought up the need to network learning through thousands of images of the same class. The layer's neurons are connected to the neurons of the previous and the subsequent layer's neurons. If each neuron in the layer connected to each neuron in the next layer, this is called fully connected layers. This type of connected layers consumes much time for processing. In the deep convolutional networks, the neurons in the hidden layer will receive inputs from a region in the input layer. This is called a local receptive field. Assume the input is an image 28x28 pixels, which will have 784 neurons in the input layer. So instead of connecting the 784 neurons to each neuron in the hidden layer, a small region, 5x5, will be connected to each neuron in the hidden layer [3].

The input image may be any object from animals to characters and numbers. The output layer depends on the number of objects to be classified. For example, if the input objects just cats and dogs, the output will contain only two neurons which represent a cat or a dog. For numerals, for most numbering systems the number of objects will be 10 which represent the numbering system from 0 to 9. These numerals have different forms for different languages such as

Chinese numerals, Arabic numerals and Roman numerals [4].

In convolutional neural networks, data sets are needed to train the neural networks for recognizing objects. The data sets range from thousands to tens of thousands of images to reach the maximum accuracy in recognizing the input numerals in our case. The training stops when there is no improving in accuracy. The data sets are divided into training data and testing data. The neural network is initialized first by inputting a training data to train the network till reaching the maximum accuracy and then tested with testing data. The number of training images is divides into small sets called batches. The number of images in the batch is a small number compared to the data sets to train the network for both forward and backword propagation.

There are many different numerals systems, almost 30 different systems for different writing systems. Some of these writing systems have different forms such as Chinese numeral system that has three different forms. The Arabic language has two forms, the Arabic numerals which mostly used by western countries and the second is the eastern Arabic numeral systems which used by most Arabic countries. There are 405 million people use the eastern Arabic numeral system in everyday [5]. The eastern Arabic hand written digits have different shapes which is different from one person to another. This requires the need to deep learning to train the network to recognize the different shapes for the same number and other numbers. Some numerals are similar to some characters and punctuations in the English and Arabic language such the zero ( ) in eastern Arabic is the full stop in English, the numerals ( ) and ( ) represents the same number two [6-8].

The paper will be organized as follows: Section II will represent literature review on the deep learning networks with max pooling and feature maps. Section III will represents method and results namely, the new technique for recognizing the eastern Arabic handwritten digits with comparison to the current techniques. Finally, the conclusion will be presented in section IV.

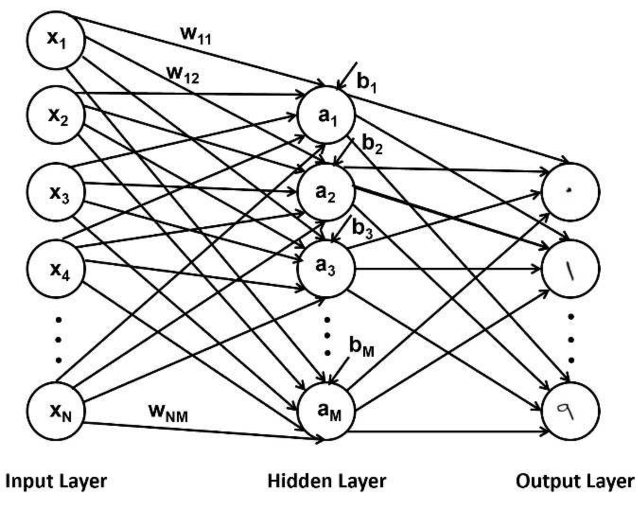

Figure 1. *The neural network.*

2. Literature Review

Deep learning is a subfield of machine learning and it is also known as deep neural networks. Deep learning is applied to raw data to train the network to recognize the objects and is based on artificial neural networks. The deep neural networks have different layers such as input layer, output layer and one or more hidden layers as shown in Figure 1. Each layer has certain numbers of nodes called neurons, these neurons in the input layer represent each pixel in the image. In fully connected layers, neurons of each layer are connected to all neurons in the previous or next layer. The whole network doesn't have to be fully connected layers. In this type of neural networks, the system consumes much time to process the inputs and to learn the network. Each neuron has weights from the previous layer's neurons and a bias which will be changed during network learning to achieve maximum accuracy at the output [9, 10]. When initializing the network, the weights from the previous neurons take random values around zero until training the neural network [11].

The path from the input to the output is called forward propagation. Each neuron gathers weights from the previous layer's neurons and a bias added to the sum. This sum occupies a large range of numbers which will produce some errors when optimizing the network for maximum accuracy. An activation function is needed to squeeze those numbers and results in faster calculations and accuracy. The sum of weight and bias in a neuron is a = wx+b which represents the weights from all neurons in the previous layer and a bias add to this neuron. The equation can be represented in matrix form as shown in equation (1).

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

Where wij represents the weight and bj is the bias.

There are many activation functions such as binary step, linear, sigmoid, tanh, ReLU, SoftMax functions. The gradient of binary step and linear functions became zero due to linearity. Mostly used activation functions in deep neural networks are the ones that can be derivative many times for backpropagation such as sigmoid activation function, tanh, SoftMax and ReLU activation function [12, 13]. Sigmoid function is a non-linear activation function which has the value f(x)=1/(1+e^-x) and it is continuously differentiable. The sigmoid function is suitable for ranges between -3 and 3 and otherwise the network is hard to learn. The tanh activation function tanh(x)=2sigmoid(2x)-1, has the same limitation as sigmoid function with smaller range of values (-1, 1). The tanh function provides better results than sigmoid function because it is zero centred and biased in any direction. The ReLU activation function f(x)= max (0, x), is gained popularity in deep learning because it activates all neurons at asynchronously. The ReLU has limitation for negative values which results in some neurons may not be activated. The leaky ReLU is an improved version of ReLU function, equation (2), which solves the problem of neurons with negative values which can't be activated [14, 15]. Softmax is a combination of many sigmoids which is used at the output layer because it distribute the values as a probability for each output neuron or class, equation (3).

$$f(x) = \begin{cases} 0.01x, x < 0 \\ x, x \geq 0 \end{cases} \quad (2)$$

$$\sigma(x_i) = \frac{e^{x_i}}{\sum_{k=1}^{K} e^{x_k}} \text{ for } i = 1, \dots, K \quad (3)$$

In convolutional neural network, each hidden neuron is connected to some of the input neurons [16]. With this architecture, the network is fast and easy to train as shown in Figure 2. The input layer is 28x28 neurons and the hidden layer is 24x24 neurons, the neurons of the first hidden layer are connected to a filter size of 5x5 neurons which is called a filter in the input layer. This filter is called local receptive field for the hidden neuron and it passes over the total number of input neurons. Each hidden neuron has 25 weights from the input layer and a bias added to the sum. The hidden layer in this case is called convolutional layers.

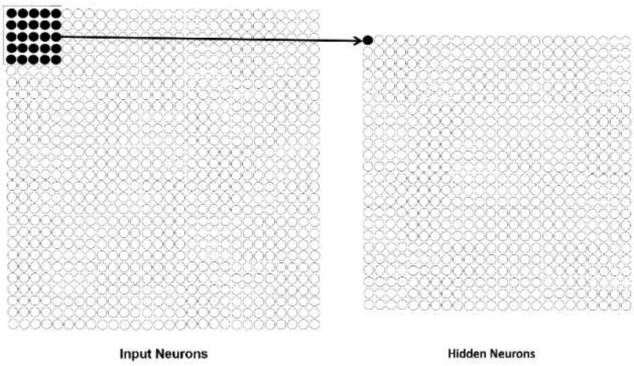

Figure 2. *Convolutional neural network.*

In this case the sum at each neuron is given as in equation (4).

$$\sigma\left(b_{i,j} + \sum_{n=0}^{4} \sum_{m=0}^{4} w_{n,m} a_{i+n,j+m}\right) \quad (4)$$

Where $b_{i,j}$ is the bias at the hidden neuron $a_{i,j}$ with weights of the 25 neurons filter assigned for each neuron.

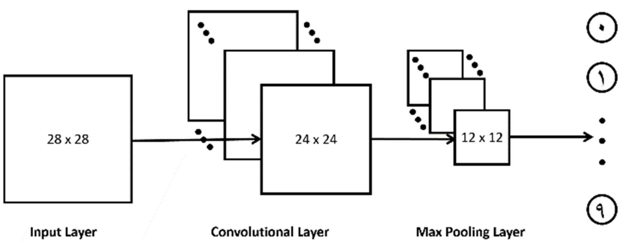

Figure 3. *The hidden layer with multiple feature maps.*

After convolutional layer, there is a pooling layer which minimize the total number of neurons in each feature map. For example, for each 2x2 neurons in the current layer, there is a one neuron in the pooling layer the pooling layer as shown in Figure 3. In max-pooling, which is mostly used in deep learning, the value of the neuron in the pooling layers is the maximum of the 4 neurons in the convolutional layer.

3. Methods and Results

In eastern Arabic numeral systems, there are 10 digits from zero to nine which are different from Arabic numeral system that are used in western countries as shown in Figure 4. These numbers have more curves than the other numeral systems [17, 18]. This brought up the need for character recognition using deep learning [19].

Figure 4. *The Eastern Arabic Numeral System.*

Figure 5. shows random Arabic handwritten digits with some errors in recognizing the digits due to the different hand-written styles which is different from person to person. The errors in this deep learning network for example some digits such as 2, 6 and 8 are recognized as digit 1, digit 9 is recognized as digit 6 and finally digit 3 is recognized as digit 2. These are some examples of training errors in deep convolutional networks.

The proposed deep learning network has 28x28 neurons at the input and the next is a hidden convolutional layer with 20x24x24 with 20 feature maps and 5x5 local receptive field. Another hidden layer is a max-pooling layer with 20x12x12 with a stride length of 1 and 2x2 pooling window and the output with 10 classes. The activation function of the hidden layers is ReLU activation function and the output layer uses SoftMax activation function. L2 regularization is used to reduce the overfitting as shown in equation (5), where λ represents the regularization parameter (around 0.1) which is optimized for better results and the total number of training dataset is n. The mini batch size is 64 and the number of epochs is 60 with learning rate η=0.01 and λ=0.1. The cost function used is log likelihood cost function see equation (6). Where C is the cost function with n training data and a is the output function for all the network.

$$\text{cost function} = \text{loss} + \frac{\lambda}{2n} * \sum \|w\|^2 \quad (5)$$

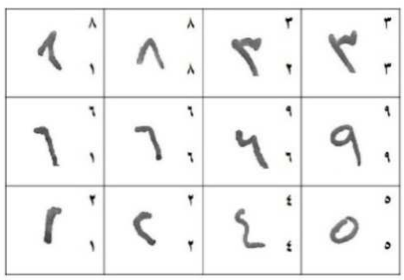

Figure 5. *Random digits.*

$$C = \frac{-1}{n}\sum_j\left(\ln a_j^L\right) \quad (6)$$

The deep convolutional network is designed with different hidden layers and hyperparameters to achieve the highest performance for the eastern Arabic handwritten recognition. There six different configurations for the deep convolutional network.

1) First: a fully connected layer with 100 neurons is added after the max pooling layer and it has sigmoid activation function.
2) Second: an extra convolutional pooling layer is added between max pooling layer and the fully connected hidden layer.
3) Third: in this deep learning network, ReLU activation function is used for all layers except the output layer with 60 epochs and learning rate of η=0.03 and λ=0.1.
4) Fourth: the training data are expanded using an algorithm to displace the image pixels using the program ex-pand_mnist. py. This program is used with eastern Arabic handwritten digits which results in 250,000 training images. The activation function in this network is the ReLU activation function.
5) Fifth: Adding another an extra fully connected layer with 100 neurons.
6) Sixth: drop out the two fully connected layers with 40 epochs.

Figure 6, Figure 7, Figure 8, and Figure 9 show the output of the network for different network settings. Figure 6, and Figure 7 show the networks with two fully connected layers with 100 neurons and the network with those two hidden layers are removed respectively. As can be seen in Figure 8 and Figure 9 the networks with convolutional layers and pooling layers have performance higher than the networks with fully connected layers, 97.96% to 97.82% validation dat and 97.89% to 97.82 % testing data. From Figure 9, the settings with fully connected layers removed have higher accuracies than other layers as the training sets increased even with lower number of epochs.

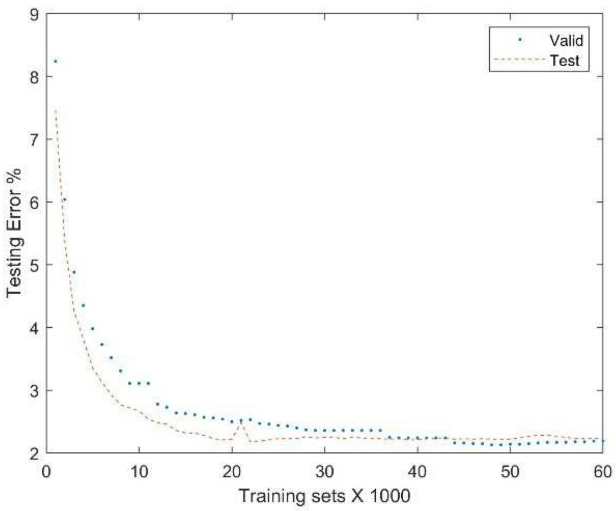

Figure 6. Validation and testing for settings 5.

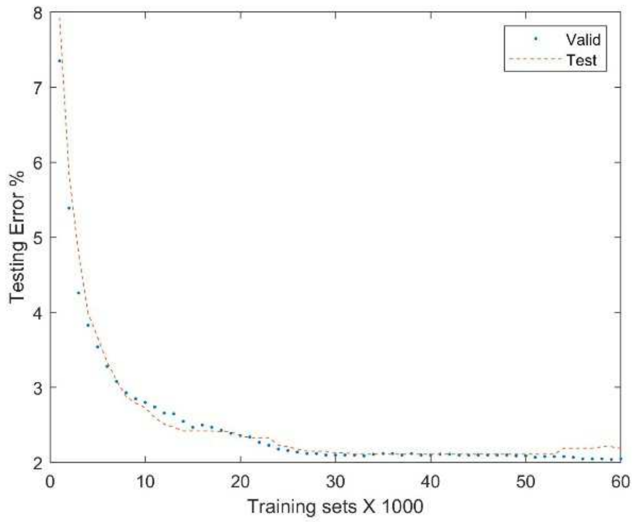

Figure 7. Validation and testing for settings 6.

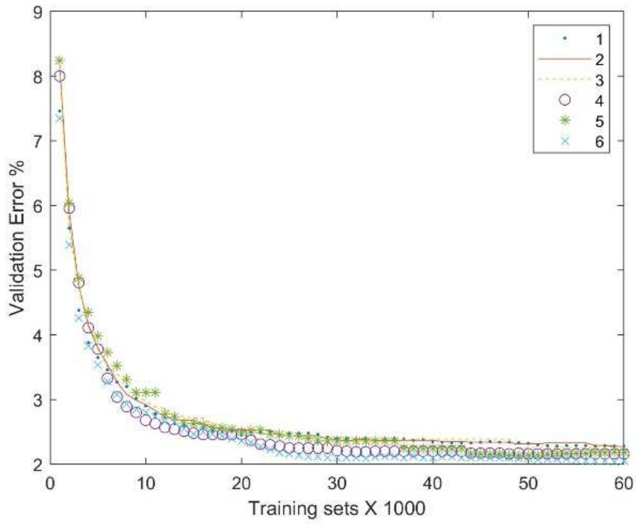

Figure 8. Validation accuracies for different settings.

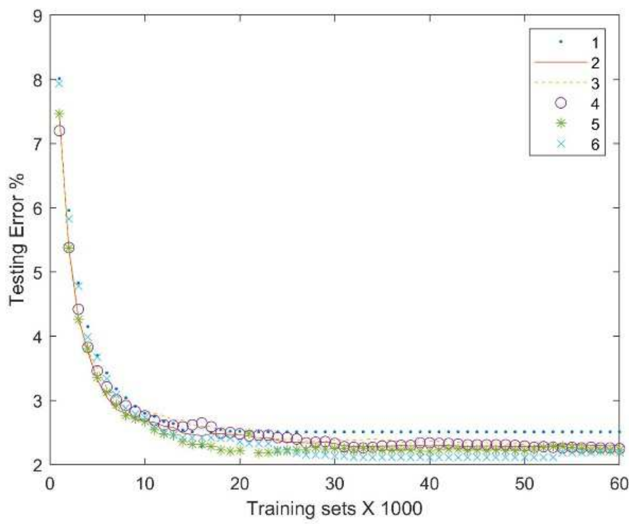

Figure 9. Testing Accuracies for Different Settings.

4. Conclusion

Eastern Arabic handwritten digits recognition is the one of the most challenging objects to recognize. There are many techniques applied to object recognition from computer vision with simple mathematical form to neural networks with very complex and more accurate object recognition results. Deep learning networks or convolutional neural networks, which makes the neural networks more efficient with less consuming time and less errors. Deep learning proved to have better performance using convolutional layers and max pooling layers than using fully connected layers. The ReLU activation function and soft-max layer with log likely-hood cost function have improved the performance of the network. We proved that the convolutional layers have higher accuracy than the fully connected layers can provide.

Conflicts of Interest

The author has no conflicts of interest to declare.

References

[1] M. A. Nielsen, Neural networks and deep learning, vol. 2018. Determination press San Francisco, CA, 2015.

[2] 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.

[3] M. Ramzan et al., "A survey on using neural network based algorithms for hand written digit recognition," International Journal of Advanced Computer Science and Applications, vol. 9, no. 9, pp. 519–528, 2018, doi: 10.14569/ijacsa.2018.090965.

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

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

[6] O. ELMelhaoui, M. ELHitmy, and F. Lekhal, "Arabic Numerals Recognition based on an Improved Version of the Loci Characteristic," International Journal of Computer Applications, vol. 24, no. 1, pp. 36–41, 2011, doi: 10.5120/2912-3830.

[7] H. A. Morsy, "Optimization of Arabic Handwritten digits recognition using CNN," International Journal of Scientific & Engineering Research V, vol. 11, no. 11, pp. 372–376, 2020.

[8] H. A. Morsy, "Developing a New CCN Technique for Arabic Handwritten Digits Recognition," International Journal of Recent Technology and Engineering (IJRTE), vol. 9, no. 3, pp. 520–524, 2020, doi: 10.35940/ijrte.C4588.099320.

[9] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, pp. 1–15, 2015.

[10] R. M. Gray, Entropy and information theory. 2011. doi: 10.1007/978-1-4419-7970-4.

[11] Y. Lecun, L. Bottou, Y. Bengio, and P. Ha, "Gradient-Based Learning Applied to Document," Proceedings of the IEEE, no. November, pp. 1–46, 1998, doi: 10.1109/5.726791.

[12] P. Selvi and 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.

[13] S. 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.

[14] M. Takruri, R. ALHmouz, and A. ALHmouz, "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.

[15] J. H. Alkhateeb and M. Alseid, "DBN - Based learning for Arabic handwritten digit recognition using DCT features," 2014 6th International Conference on Computer Science and Information Technology, CSIT 2014 - Proceedings, no. September, pp. 222–226, 2014, doi: 10.1109/CSIT.2014.6806004.

[16] J.-C. Vialatte, V. Gripon, and G. Coppin, "Learning Local Receptive Fields and their Weight Sharing Scheme on Graphs," Jun. 2017.

[17] A. ELSawy, H. ELBakry, and M. Loey, "CNN for Handwritten Arabic Digits Recognition Based on LeNet-5," in Proceedings of the International Conference on Advanced Intelligent Systems and Informatics, 2017, pp. 566–575. doi: 10.1007/978-3-319-48308-5_54.

[18] 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.

[19] A. ELSawy, H. ELBakry, and M. Loey, "CNN for Handwritten Arabic Digits Recognition Based on LeNet-5," 2017. doi: 10.1007/978-3-319-48308-5.