

Handwritten Arabic Numeral Recognition using Deep Learning Neural Networks

Akm Ashiquzzaman and Abdul Kawsar Tushar

Computer Science and Engineering Department, University of Asia Pacific, Dhaka, Bangladesh
{zamanashiq3, tushar.kawsar}@gmail.com

Abstract—Handwritten character recognition is an active area of research with applications in numerous fields. Past and recent works in this field have concentrated on various languages. Arabic is one language where the scope of research is still widespread, with it being one of the most popular languages in the world and being syntactically different from other major languages. Das et al. [1] has pioneered the research for handwritten digit recognition in Arabic. In this paper, we propose a novel algorithm based on deep learning neural networks using appropriate activation function and regularization layer, which shows significantly improved accuracy compared to the existing Arabic numeral recognition methods. The proposed model gives 97.4 percent accuracy, which is the recorded highest accuracy of the dataset used in the experiment. We also propose a modification of the method described in [1], where our method scores identical accuracy as that of [1], with the value of 93.8 percent.

Keywords - Deep learning, ConvNet, Handwritten digit recognition, Arabic numeral.

I. INTRODUCTION

The optical character recognition (OCR) of digits on scanned images hold widespread commercial and pedagogical importance in fields such as automatic recognition, check reading, data collection from forms, and textbook digitization. Being still and active area of research, OCR recognition is being acquainted to different languages around the world.

Statistically speaking, Arabic is one of the top five most widely spoken language in the present world, being used by more than 267 million people [2]. Arabic is an official language in more than 16 countries and spoken widely in a number of other countries as well [3]. In recognition of this, Arabic has held the status of an official language of the United Nations since 1973 [4]. In addition to that, Arabic is the liturgical language of the religion of Islam, hence a large number of religious texts and writings are in Arabic. Numerous languages including Persian, Hindi, and Urdu have taken inspiration from Arabic. Sample handwritten numerals in Arabic language and their inverted images are shown in Fig. 1.

Although a number of languages have received significant attention in terms of OCR research [5], [6], Arabic is still mostly an unexplored field of study, with the maximum work done previously on this language being based on printed characters [7]. The important distinction between Arabic and major roman-based languages is that Arabic words as well as characters within words are written from right to left, as





















Arabic Digit	English Digit	Image	Inverted Image
١	1		
٢	2		
٣	3		
٤	4		
٥	5		
٦	6		
٧	7		
٨	8		
٩	9		
٠	0		

Fig. 1. Arabic handwritten digits and their inverted images.

opposed to left to right as with English. Nonetheless, digits of an Arabic number is written from left to right.

Das et al. [1] have devised a novel method for Arabic handwritten digit recognition with help of a multi layer perceptron (MLP) which can bring significant accuracy to the method of Arabic digit OCR. In this paper, we propose a methods which can increase the accuracy of this process by using convolutional neural network (CNN) in place of MLP. Furthermore, we also provide a modification of the method of Das et al. which achieves the same level of classification accuracy. Our methods also solve the problem of overfitting that has affected the method in [1], by applying dropout regularization.

II. BACKGROUND

A novel method for recognizing Arabic numerals has been devised by [1]. This method utilizes an MLP where a set of 88 features is used. The feature set is divided into 72 shadow features and 16 octant features. For this work, a database of 3000 samples is used which is obtained from [8]. The images are scaled to the pixel size of 32x32 and then defined with gray scale pixel values. Finally, conversion to binary images

takes place.

The MLP classifier used in this method has one input layer, one output layer, and one hidden layer in between, since a single hidden layer suffices for the computation of a uniform approximation of given dataset [9]. This setting of layers is shown in Fig. 2. Quantity of the neurons residing in each layer are chosen by trial and error. Each of the neurons has sigmoid as the activation function. Supervised training is undertaken for this method, which is performed with the help of 2000 training samples. Back propagation algorithm is used to train the MLP for recognizing the input-output relationship of the problem at hand.

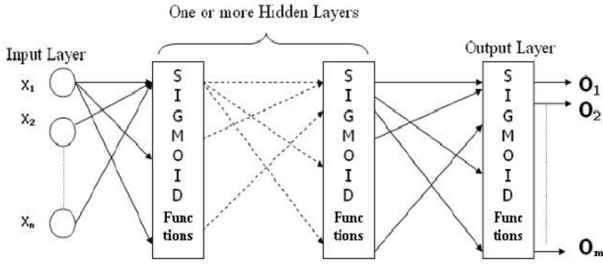


Fig. 2. Multi layer perceptron used in [1].

The results are evaluated with three-fold cross validation, where in each fold the number of neurons are adjusted for optimal performance. By increasing the number of neurons in hidden layer, performance is improved, but only up to a certain threshold point, beyond which increase in number of neurons translates as decrease in performance. This phenomenon is termed as the problem of overfitting [10]. The number of neurons in the hidden layer is finally fixed at 54, and the performance for numeral recognition is obtained at 93.8%.

III. DATASET

Deep learning is solely depended on the data and hence it needs a large amount of data to function properly. Our model is trained and tested on the CMATERDB 3.3.1 Arabic handwritten digit dataset [8]. The entire dataset consists of 3000 images, making it a source of 3000 unique samples. We divide the dataset in the same way as Das et al. suggested in their studies [1], which is 2000 training samples and 1000 test samples.

The dataset consists of 32x32 pixel RGB bitmap images. The images are preprocessed to convert them into gray scale values. Then the image are inverted in order to enhance their feature. As a result, sample image now consists of white foreground and black background, as oppose to conventional white background and black foreground. Then all the images are normalize in order to reduce computation.

For the MLP model, we reshape the images as a single one-dimensional valued tensor. But for the CNN model, the images are kept in the original dimensions. Fig. 3 shows the sample input image of a digit as opposed to the final preprocessed one.

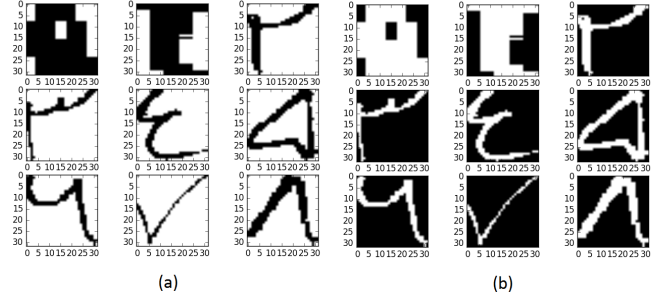


Fig. 3. (a) Raw data and (b) final preprocessed data.

IV. PROPOSED METHODS

In this section we discuss our two different methods and how they improve the recognition performance of the given dataset. The first method is the MLP used in the method in [1] with two key changes - we use Rectified Linear Unit (ReLU) as the activation of each neuron in input layer and hidden layer, and use softmax function in the outer classifying layer. In order to tackle the overfitting problem, during training process a number of neurons in each layer are randomly chosen not to get gradient update. This process is called dropout [11]. Input layer and intermediate layer have a dropout percentage of 25% to prevent overfitting. The first output layer of the MLP consists of 512 neurons and take the input image as an one-dimensional array. The intermediate layer consists of 128 neurons and the outer layer consists of 10 neurons. Fig. 4 represents our MLP method.

Our second method employs Convolutional Neural Network, or Convnets. Convnets are a special variety of Artificial Neural Network (ANN), with trainable weights and biases. Convnets are trained with backpropagation algorithm [12], nonetheless the architecture of the layers is somewhat different [13].

The first hidden layer is a convolutional layer which has 30 feature maps, each with a kernel size of 55 pixels and a ReLU activation function. This is termed as the input layer, which takes in images with 32x32 pixel values. Next we define a pooling layer that takes the maximum value. It is configured with a pool size of 22. The following hidden layer is another convolutional layer that has 15 feature maps, each with a kernel size of 3x3 pixels as well as a ReLU activation function. This layer is followed by another pooling layer which is the same as previous pooling layer.

The next layer is a regularization layer, also called Dropout. It is configured to randomly exclude 25% of neurons in the layer in order to reduce overfitting. After this we have a layer that converts the two-dimensional matrix data to a vector called Flatten. It allows the output to be processed by standard fully connected layers, which is our next layer. It contains 128 neurons and ReLU activation function, as well as a Dropout layer which is configured to randomly exclude 50% of neurons in the layer. Finally, the output layer has 10 neurons for the 10 classes or digits in Arabic numeral and a softmax activation

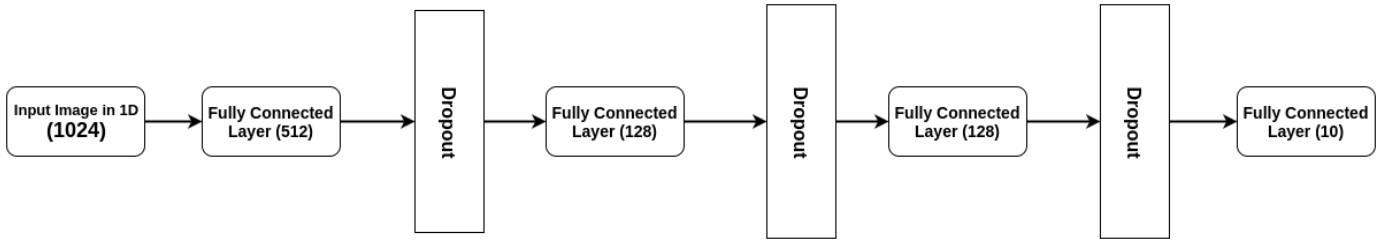


Fig. 4. Description of MLP method.

function to output probability-like predictions for each class. Fig. 5 represents our proposed CNN method.

The MLP model takes the whole input image, which has the dimension of 32x32 in a single 1 dimensional array. So the whole image of 1024 pixels are selected as features. In our convnet approach, we used the same principle. The 2 dimensional convnet takes the whole image as the feature. The MLP method proposed by Das et al. [1] used a pre selected 88 features to feed into the MLP network, however we did not pre-select any features or reduce them. Both of the models takes the whole Image as features.

V. EXPERIMENT

Both of the proposed models are trained in the same configuration to evaluate their performances in a comparative way. Both the proposed models, MLP and CNN, are trained and tested against the CMATERDB Arabic handwritten digit dataset and validated against the test data samples. The total parameter size for the proposed MLP method is 591745 whereas for the proposed CNN method the size is 75383. The experiments are done in Keras [14] along with Theano backend [15].

Both of the models are trained for 1000 epochs with the batch training size of 128. Adadelta optimizer is used as the optimizing function whereas categorical cross entropy is used to calculate the loss. It is also possible to use mean squared error (MSE) as the loss function; however, that is not conducted in this study. All the models are trained with no layer freezed. The supervised training of two proposed methods are done with 2000 training samples, followed by validation with 1000 test samples.

Experimental Models were implemented in Python programming languages with Theano and Keras Library. The experimental setup computer had an intel i5-6200U CPU @ 2.30GHz 4 with 4 Gb RAM. Nvidia Geforce GTX 625M

Graphics card is used to utilize CUDA for faster GPU operation. The whole training process is done for 1000 epoch for each models. Data batch size were 128, during both training and validation.

VI. RESULT AND DISCUSSIONS

The proposed CNN method outperforms both the proposed MLP method as well as the MLP used in [1], the latter two yielding identical results for the dataset used in this paper. Table 1 compares the final accuracy of the models.

TABLE I
PERFORMANCE COMPARISON OF PROPOSED METHODS AND METHOD DESCRIBED IN [1]

Method Name	Accuracy
Das et al. [1]	93.8
Proposed MLP	93.8
proposed CNN	97.4

After analyzing and evaluating the result of our model for the test samples and validation accuracy, we can come to the conclusion that even though the MLP model has reached the previous accuracy in [1], it cannot be improved anymore in the current configuration.

The computational disadvantage of the basic MLP against CNN here is that, CNN uses the convolution in its layer to detect the features of images in more robust ways [16]. The basic CNN model, when trained properly over a dataset, can easily detect image features more accurately, thus performing well above MLP in the fields of classic image recognition problems. In our experiment we demonstrate the performance of the classic MLP. The performance is on par with the performance in [1]. However, the added convolutional layers in the proposed CNN method enhance the accuracy in

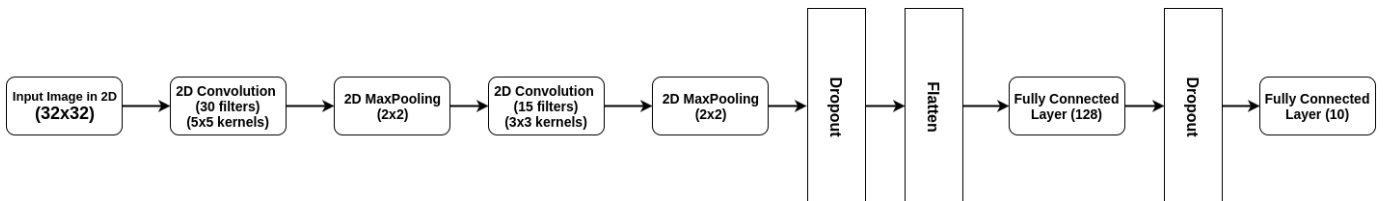


Fig. 5. Description of CNN method.

digit recognition, making the accuracy 97.4 percent. This is the recorded highest accuracy for the CMATERDB Arabic handwritten digit dataset.

VII. CONCLUSION

This paper proposes a novel algorithm for recognizing numerals in handwritten Arabic with the help of CNN. Our proposed CNN model uses several convolutional layers along with ReLU activation, with dropout used as a regularization layer. Then the output is fed into a fully connected layer (which is the same as MLP model) with softmax activation to obtain prediction for each class. Our proposed novel method achieves better accuracy than the method discussed in Section II. We also propose a modification of the method in Section II using MLP. In the MLP model, we implement dropout regularization to reduce overfitting in between fully connected layers. The output layer, consisting of 10 neurons with softmax activation, predicts the probability for 10 individual digit classes (0-9). Our proposed MLP method achieves identical accuracy score compared to the method of Das et al. [1]. The proposed methods are described in detail, and the performances are compared.

REFERENCES

- [1] N. Das, A. F. Mollah, S. Saha, and S. S. Haque, "Handwritten arabic numeral recognition using a multi layer perceptron," *CoRR*, vol. abs/1003.1891, 2010. [Online]. Available: <http://arxiv.org/abs/1003.1891>
- [2] "Summary by language size — ethnologue," <https://www.ethnologue.com/statistics/size>, accessed: 2016-12-25.
- [3] "Languages spoken in each country of the world," <http://www.infoplease.com/ipa/A0855611.html>, accessed: 2016-12-25.
- [4] "Official languages — united nations," <http://www.un.org/en/sections/about-un/official-languages/>, accessed: 2016-12-25.
- [5] R. Plamondon and S. N. Srihari, "Online and off-line handwriting recognition: a comprehensive survey," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 1, pp. 63–84, 2000.
- [6] P.-K. Wong and C. Chan, "Off-line handwritten chinese character recognition as a compound bayes decision problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 9, pp. 1016–1023, 1998.
- [7] A. Amin, "Off-line arabic character recognition: the state of the art," *Pattern recognition*, vol. 31, no. 5, pp. 517–530, 1998.
- [8] "Google coe archive - long-term storage for google code project hosting," <https://code.google.com/archive/p/cmaterdb/downloads>, accessed: 2016-12-30.
- [9] M. Kubat, "Neural networks: a comprehensive foundation by simon haykin, macmillan, 1994, ISBN 0-02-352781-7," *Knowledge Eng. Review*, vol. 13, no. 4, pp. 409–412, 1999. [Online]. Available: <http://journals.cambridge.org/action/displayAbstract?aid=71037>
- [10] P. M. Domingos, "Bayesian averaging of classifiers and the overfitting problem," in *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000)*, Stanford University, Stanford, CA, USA, June 29 - July 2, 2000, 2000, pp. 223–230.
- [11] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [12] Y. Hirose, K. Yamashita, and S. Hijiya, "Back-propagation algorithm which varies the number of hidden units," *Neural Networks*, vol. 4, no. 1, pp. 61–66, 1991.
- [13] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.
- [14] F. Chollet, "keras," <https://github.com/fchollet/keras>, 2015.
- [15] Theano Development Team, "Theano: A Python framework for fast computation of mathematical expressions," *arXiv e-prints*, vol. abs/1605.02688, May 2016. [Online]. Available: <http://arxiv.org/abs/1605.02688>
- [16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.