Arabic digit recognition using robust deep convolution neural network

Qasim Mahdi Haref a, Mohsin N. Srayyih AL-Maliki b, Maytham Mohammed Tuaama c, Hayder M. Albehadili d

aImam alkadhum collage, Amara, Iraq, b University Of Misan, Amara, Iraq, c Imam alkadhum collage, Amara, Iraq, d Misan Oil Company, Amara, Iraq,

Email: qasimmahdi@alkadhum-col.edu.iq, muhsen@uomisan.edu.iq, Maythammohammed@alkadhum-col.edu.iq, hadereng@yahoo.com

Abstract: Recently, digit recognition becomes one of interest problems for many researchers. However, Arabic digits have lack for such research. In this work, we used a robust deep convolution neural network (DCNN) to evaluate our collected Arabic digit dataset. We introduce substantial changes to CNN models to achieve superior results to prior work. Extensive experimental results are contacted to show the robustness of proposed models. Our detector achieves best current state-of-the-art results.

Keywords: Convolution Neural Network, Arabic Digit Recognition.

1. Introduction

Recently, digit recognition becomes one of attracting challenges in the computer vision community because of its wide applications. After the revolution of convolution neural network (CNN), recognizing digits achieves highest performance because CNN is end-to-end training model that can boost many computer visions approaches.

Convolutional Neural Networks (CNNs) plays crucial role and leverages factor in variety of recent machine learning applications. It boosts several state-of-the-arts results on different complicated image classification such as ImageNet [1, 2, 3, 4, 5,6], object detection [1,7, 8], image segmentation [9,10], and face recognition [11, 12, 13, 14]. Accordingly, we mainly depend CNNs in this work for digit recognition. It achieves superior results comparing with former models of Arabic digit recognition. It is worth to mention that digit recognition has variety of applications used for such as banks for reading checks, post offices, and others. In this work, we use our collected dataset. Like MNIST [15] dataset, our dataset was collected from different school levels. It is challenging dataset because of its variety of different sketch for each digit, as we will see next sections.

2. Related work

Wide range of extensive works has been used to push forward results on digit recognition. Especially after CNNs and parallel processing of GPUs appearing, it helps researchers to achieve more accurate results. The proposed work introduced by [16] to extract features from input images. Also authors used GPU to implement their approach for obtaining accurate results. Proposed method also exhibited in [1]. A visualization method implemented to obtain more perception for layers. More advanced works are obtained in [7] to achieve better results comparing to prior works. The best reported results on MNIST dataset was obtained by [7]. They used different technique by introducing multi-columns deep convolution neural network. However, there are few works achieved on Arabic datasets because as we said the lack for large size of dataset.
3. Contributions

We collected new dataset, which has new challenge comparing to all existing datasets. It has thousands of patterns collected from different schools levels. Different models of CNN are used for evaluation. We introduced a broad study by carefully diving into sensitive parameters of CNN that can influence model performance. Not only the influence of the parameters is studies to show effectiveness of introduced models but also the depth of the CNN and how it can degrades or boosts models accuracy. We compare the results to all former models and it showed that our proposed model achieve state-of-the-art comparing to all of them. Our models inherited and inspired by the techniques used in SPPnet proposed in [18].

Different image sizes for each patterns are suggested to allow zooming different parts of the digits.

3.1 Pipeline steps of digit recognition

Since our CNN designed and deployed in this work has ability to receive different image sizes (multi-scale input images) for each input pattern because it works identically to SPPnet as described in [18]. Thus it has the leverage of extracting features of different parts of input image resulting in better performance. In this section, we show the main pipeline steps of our digit classification models. The pipeline of image recognition presented in this paper is exhibited in fig. 1.

We can summarize and divided the model into three main parts. In the first part, the input digit pattern is passed into smart zoomed SPP-net to create multi-scale image sizes to be presented to the next level of network. After generating different sizes of input images, the last is passed to the next step of the next level of the network, which is the main part of the CNN to extract features from given images. It is worth mentioning that extracting features is the main and the critical part of our main purpose. Thus, we carefully use and explore different parameters that can crucially influence the feature extractions. Also, there is more criteria that can influence extracted features such as how deep is the network. In this work, we contacted our experiments on different levels of network to obtain the best fit for our experiments. Finally, the last level of extracted features is by passing those features into robust classification step. In this work, SVM is used on the top of the network to classify extracted features.

3.2 Dataset

The collected Arabic digit dataset has of 46,000 images. Written by 840 writers (with minor difference for some categories), each writers wrote fifty patterns distributed over ten digits (0-9). To ensure including different writing samples, the database was gathered from different institutions: Colleges, high schools, middle schools and employs. The English digits are written from left to right, but the Arabic digits are written from right to left. After collecting sample forms as shown in fig. (2), they were scanned with 300 dpi resolution in 24 bit color format with a high speed scanner then digits in our dataset we have three files, the first file is an elementary school and it has about (371) image, the second file is a high school and it has about (269) image. The last file contains bachelor students and employees and it has about (200)
image. For each image we have 60 digits which is manually extracted, categorized, and bounded by bounding boxes using Photoshop as shown in fig.(2). In the first file we have about 22,260 boxes, second file has 3,228 boxes and last file has 12,000 boxes. After we extract images in the files, we remove some damaged boxes because it is not clear. Thus we collect all images that we

![Figure 2: samples of collecting forms of hand written digits.](image)

Already extracted and clean it, next step was our challenge in which we have three files and each file has a random image from (0-9) digit. In our work, each file has subfolder between (0-9) since we work on supervised learning and in this way we need to label our data set. When we choose set of numbers to train our program we have to label it at the beginning to recognize it from program, therefore we design new code that works to extract all images and categorize it from (0-9). For purpose of training, the dataset is divided into two parts. The training part has 36,000 samples and 10,000 samples for testing. The following table shows the some statistic information about the dataset.

| Writer                  | Number of writer | Total wrong Samples | Total Digit | Total correct Digit |
|-------------------------|------------------|---------------------|-------------|--------------------|
| Elementary school       | 371              | 53                  | 3,180       | 318                |
| High School             | 269              | 5                   | 300         | 264                |
| Bachelor degree and employs | 200              | 6                   | 360         | 194                |

4. Conclusion
Recently, digit recognition becomes one of most appealing task in machine learning. In this work, we present a proficient deep learning architecture for a challenging Arabic Digit dataset classification. The dataset is collected and maintained from different school levels. It is one of the most challenging dataset which consists of more than 40,000,000 digits for training and 10,000,000 digits for testing.

Then a robust model of convolution neural network (CNN) was used for recognition. Extensive experiments are performed on proposed model and a state-of-the-art result is achieved on our collected dataset after.
References

[1] Hayder M. Albeahdili, Haider A. Alwzwazy, Naz E. Islam. “Robust Convolutional Neural Networks for Image Recognition”. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 11, 2015.

[2] Fabien Lauer, Ching Y. Suen, and G’erard Bloch “A trainable feature extractor for handwritten digit recognition”, Journal Pattern Recognition 2007.

[3] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, Zhuowen Tu, “Deeply-Supervised Nets “ NIPS 2014.

[4] M. Fischler and R. Elschlager, “The representation and matching of pictorial structures”, l IEEE Transactions on Computer, vol. 22, no. 1, 1973.

[5] Kevin Jarrett, Koray Kavukcuoglu, Marc’Aurelio Ranzato and Yann LeCun “What is the Best Multi-Stage Architecture for Object Recognition” CCV’09, IEEE, 2009.

[6] Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey.” ImageNet classification with deep convolutional neural networks”. In Advances in Neural Information Processing Systems 25 (NIPS’2012), 2012.

[7] Kaiming, He and Xiangyu, Zhang and Shaoqing, Ren and Jian Sun “Spatial pyramid pooling in deep convolutional networks for visual recognition!” European Conference on Computer Vision, 2014.

[8] X. Wang, M. Yang, S. Zhu, and Y. Lin. “Regionlets for generic object detection”. In ICCV, 2013.

[9] C. Couprie, C. Farabet, L. Najman, and Y. LeCun. “Indoor semantic segmentation using depth information”. International Conference on Learning Representation, 2013.

[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik. “Rich feature hierarchies for accurate object detection and semantic segmentation”. CoRR, abs/1311.2524, 2013.

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, S. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, and F.F. Li. “Imagenet large scale visual recognition challenge”. IJCV, 2015.

[12] Raia Hadsell, Sumit Chopra, Yann LeCun, “Dimensionality Reduction by Learning an Invariant Mapping” CVPR 2006.

[13] Raia Hadsell, Sumit Chopra, and Yann LeCun “Dimensionality Reduction by Learning an Invariant Mapping” CVPR 2006.

[14] Omkar M. Parkhi, Andrea Vedaldi,and Andrew Zisserman “Deep Face Recognition” British Machine Vision Conference, 2015.

[15] LeCun, Y. et al.,” Comparison of learning algorithms for handwritten digit recognition”. In: International conference on Artificial Neural networks, France, pp. 53–60. 1995

[16] Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey. “ImageNet classification with deep convolutional neural networks”. In Advances in Neural Information Processing Systems 25 (NIPS’2012), 2012.

[17] Ciresan, Dan; Meier, Ueli; Schmidhuber, Jürgen (June 2012). "Multi-column deep neural networks for image classification". IEEE Conference on Computer Vision and Pattern Recognition (New York, NY: Institute of Electrical and Electronics Engineers (IEEE)). 2012.

[18] Kaiming, He and Xiangyu, Zhang and Shaoqing, Ren and Jian Sun “Spatial pyramid pooling in deep convolutional networks for visual recognition” European Conference on Computer Vision, 2014.