ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Handwritten Digit Recognition Using Convolutional Neural Networks

Haider A. Alwzwazy[1], Hayder M. Albehadili[2], Younes S. Alwan[3], Naz E. Islam[4]

M.E Student, Dept. of Electrical and Computer Eng. University of Missouri-Columbia, MO, USA[1,2,3]

Professor, Dept. of Electrical and Computer Eng. University of Missouri-Columbia, MO, USA[4]

ABSTRACT: Recently handwritten digit recognition becomes vital scope and it is appealing many researchers because of its using in variety of machine learning and computer vision applications. However, there are deficient works accomplished on Arabic pattern digits because Arabic digits are more challenging than English patterns. Hence, the lacking research of using Arabic digits endeavours us to dig deeper by creating our challenge Arabic Handwritten Digits which consists of more than 45,000 samples. As a challenging dataset is used for evaluation, a robust deep convolutional neural network is used for classification and superior results are achieved.

KEYWORDSHandwritten Digit Recognition; Arabic Handwritten Digits;

I. INTRODUCTION

Recently Deep Convolutional Neural Networks (CNNs) becomes one of the most appealing approaches and has been a crucial factor in variety of recent success and challenging machine learning applications such as challenge ImageNet [1, 2, 3, 4, 5,24], object detection [1, 6, 7], image segmentation [9,10], and face recognition [11, 12, 13]. Therefore, CNNs is considered our main model for our challenging tasks of image classification. Specifically, it is used for handwriting digits recognition which is one of high academic and business transactions [14]. Handwriting digit recognition application is used in different tasks of our real life time purposes. Precisely, it is used in banks for reading checks, post offices for sorting letter, and many other related tasks.

Apparently English Handwriting datasets are widely available and significant achievements have been made for English digit datasets such as CENPARMI [15], CEDAR[16], and MNIST[17], However, there are rare works accomplished on Arabic digit datasets for many reasons. One of critical factor that can influence working on Arabic dataset is lacking to dataset. The unavailability of dataset can be one of the essential factors that can diminish working on Arabic datasets. Hence, deficiency of large challenging Arabic dataset strives us to extensively working on creating a largest and most challenging dataset which contains more than 45,000 patterns. Furthermore, we investigate and demonstrate a powerful DCNN used for classification. Not only designing powerful DCNN is presented but also critical parameters of CNN is carefully selected and tuned to produce final concrete model which achieves superior results.

II. RELATED WORK

Handwritten digit recognition (HDR) is considered one of trivial and critical machine learning problems. It has been used widely by researchers as experiments for theories of machine learning algorithms for many years. In recent years, neural networks and conventional neural network currently provide the best solutions to many problems in handwritten digit recognition. A novel hybrid CNN–SVM model for handwritten digit recognition is designed by [18]. This hybrid model automatically extracts features from the raw images and generates the predictions. For this work, the author used non-saturating neurons and a very efficient GPU implementation of the convolution operation to reduce overfitting in the fully-connected layers. To enhance method proposed in [8], [19] tackled critical investigations to diminish limitation inherited from [8]. The author introduces a novel visualization technique that gives insight into the function of feature layers and the procedure of the classifier [20] have observed convolutional net architecture that can be used even when the amount of learning data is limited. [21] have used new network structure, called Spatial Pyramid Pooling SPP-net, can generate a fixed-length representation regardless of image. Multi-column DNN (MCDNN) used MNIST digits. The result has a very low 0.23% error rate [22]. Hayder M. Albeahdili et al. [9] have performed a new

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

CNN architecture which achieves state-of-the-art classification results on the different challenge benchmarks. The error rate for this approach is 0.39 % for MNIST dataset.

III. DATASET

Our digit dataset is composed of 46,000 digits written by 840 participants. Each participant wrote fifty patterns distributed over ten digits (0-9). To ensure including different writing samples, the database was gathered from different institutions: Colleges, high school, and middle school. After collecting sample forms as shown in fig (1), they were scanned with 300 dpi resolution then digits are manually extracted, categorized, and bounded by bounding boxes using Photoshop. The dataset is partitioned into two sets. The first set consists of 36,000 samples used for training and second set has 10,000 samples used for testing. Both drawn from the same distribution and centred in a fixed size image where the centre of gravity of the intensity lies at the centre of the image with 64 x 46 pixels. Thus, the dimensionality of each image sample vector is 64 x 64 = 4096.

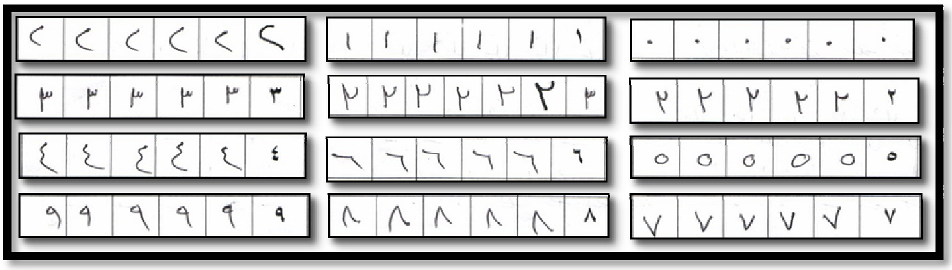

Fig.1. A sample of the dataset

IV. CLASSIFICATION STEPS

Image classification is not trivial task which can be achieved using various approaches. However, recently deep learning has been successfully applied to a wide range of machine learning applications. Accordingly, in this work we proposed a subtle Convolutional Neural Networks (CNNs) which is used to train and test our handwritten digits. Constructing CNNs plays an essential role in justifying both performance and time consumption. Thus, in our implementation, we designed an elegant CNN after carefully investigating its parameters. In general, using CNNs for handwritten digits recognition consists of a certain number of steps described below:

- Preparing patterns before feeding to the CNN. All images are pre-processed before passing into the network. In our experiments, CNN is designed to receive an image size of 64x64 pixels. Hence, all images have been cropped to the same size to be fed to the model.
- After preparing images, they are fed to the deep model to extract features. As demonstrated earlier a robust CNN is used in this experiment to extract robust features used in the final decision to justify the class to which they belong to.
- Finally, the last layer named *softmax* layer is used at the top of CNN to minimize the error.

In this work, we carefully explored the architecture of CNN consisting of five layers as shown in fig. (2) .It is clear that the network has three main stages for classification and as described below:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

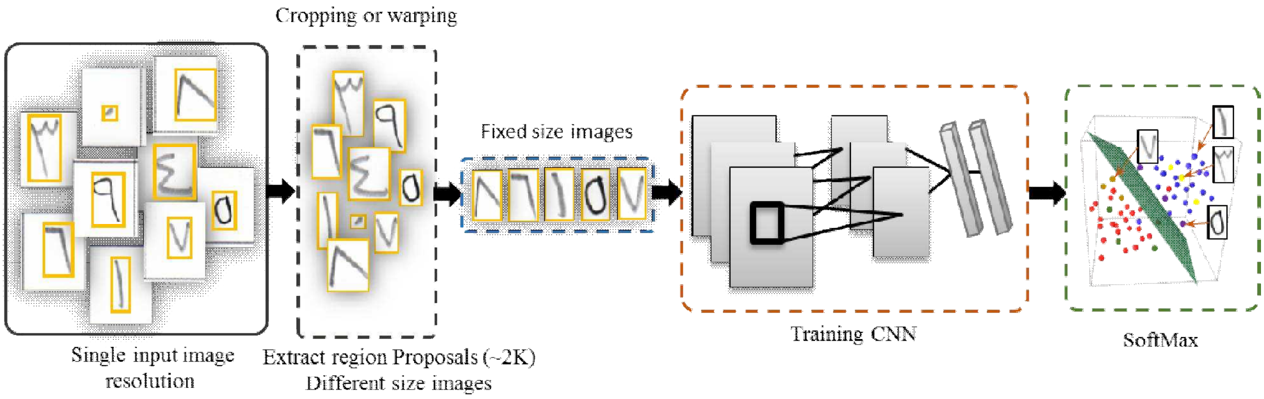

Fig.2. classification steps of our Arabic dataset

V. EXPERIMENTAL SETUP

In this work, the dataset is collected from primary, secondary, and university's students. The dataset consists of 46,612 samples. All patterns are resized to be (64x64 RGBpixels). A sample of the dataset is shown in fig.1. It is worth mentioning that the number of samples collected form eliminatory school is (371) asdepicted in fig.3 and for High School is (269) asshown in fig.4. However, the number of samples collected from students who are studying as Bachelor degree is (200) as exhibited in fig.5. The data is divided into two parts as training and testing part. Each student was given a form having ten squares to write digit numbers (0-9). There are 36,612 samples used for training and the remaining is used for testing.

Testing samples are collected from the whole dataset. 4000 samples are randomly chosen from eliminatory school's student, 4000 patterns from secondary school's student, and 2000 samples are randomly chosen from higher education school.

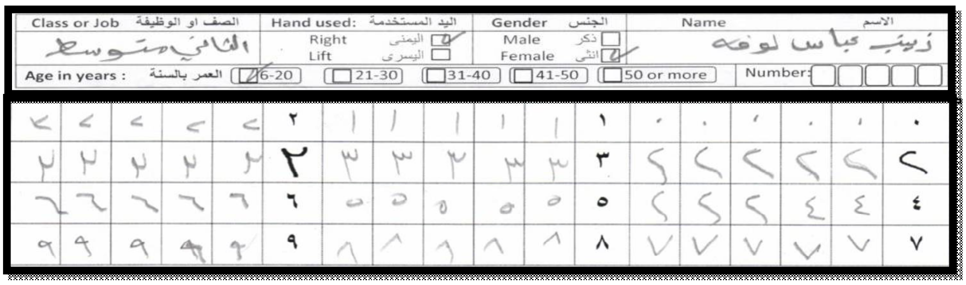

Fig. 4. A sample of the forms written by elementary school.

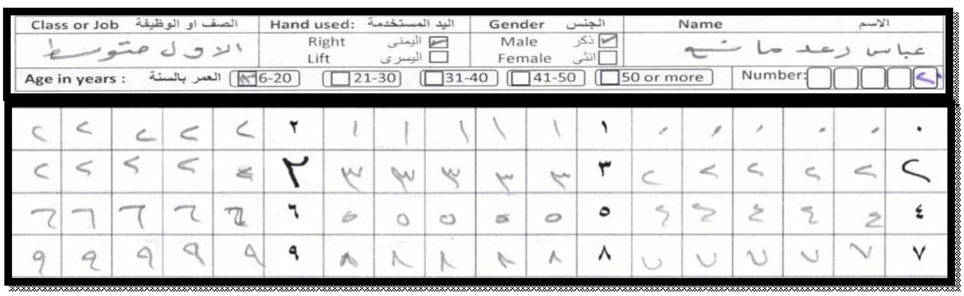

Fig. 5. A sample of the forms written by high school.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

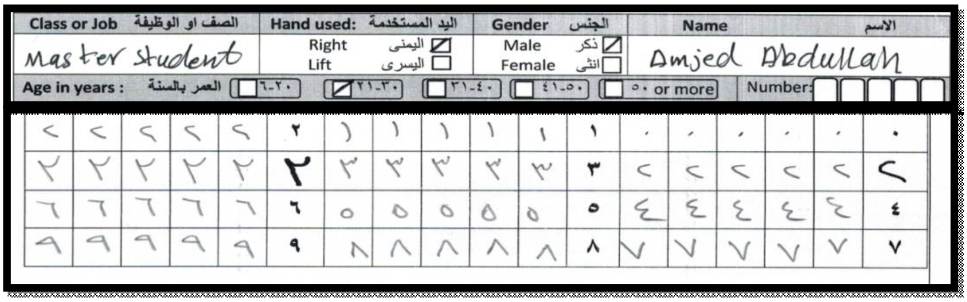
Class or Job الصف او الوظيفة | Hand used: اليد المستخدمة | Gender الجنس | Name الاسم
Master student | Right اليمنى ☑ Lift اليسرى ☐ | Male ذكر ☑ Female انثى ☐ | Amjed Abdullah
Age in years : العمر بالسنة ☐ ٦-٢٠ ☑ ٢١-٣٠ ☐ ٣١-٤٠ ☐ ٤١-٥٠ ☐ ٥٠ or more Number:

Fig.6. A sample of the forms written by high school.

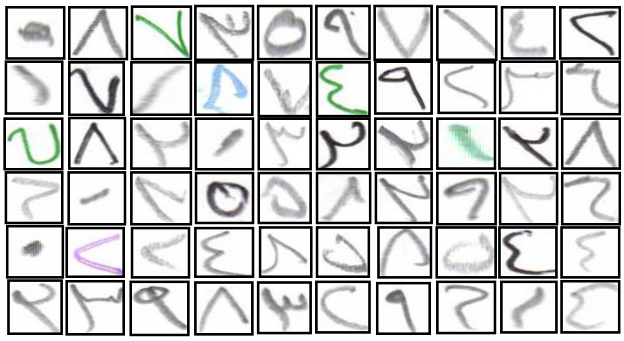

Fig. 7. A sample of samples after they cropped from the original form

It is worth to mention the steps of collecting and processing the dataset. Originally the dataset is collected as forms distributed over hundreds of students. A sample of these forms is shown in fig.7. Then we used Photoshop to cut each form a certain number of digits according to each form how many digits it has. Thus at the end, different sizes of patterns are created because they are manually processed and there is not guaranteed to have same size. Therefore, the patterns are resized to have the same size which is (64x64 pixels). Our final deep CNN model is depicted in fig. 8. It is noticeable that our model matches same model proposed by [23] because of comparison purpose to justify how the same model can behave for different datasets. The model consists of the following layers.

- Convolutional layer: the first layer of the model depicted in fig. 8, has 6 feature maps of size 28x28 pixels after cropping images as shown in fig. 8.
- Subsampling layer: consists of 6 feature maps also but each one has size of 14x14.
- Convolutional layer: the third layer marked as C3 is also convolutional layer which consists of 16 feature maps of 10x10 pixels.
- The fourth layer is max-pooling layer which has same number of feature maps as in prior layer.
- Then, the model has two fully connected layers sitting over last max-pooling layer.
- Finally, CNN ends up with soft-max layer used for final classification results.

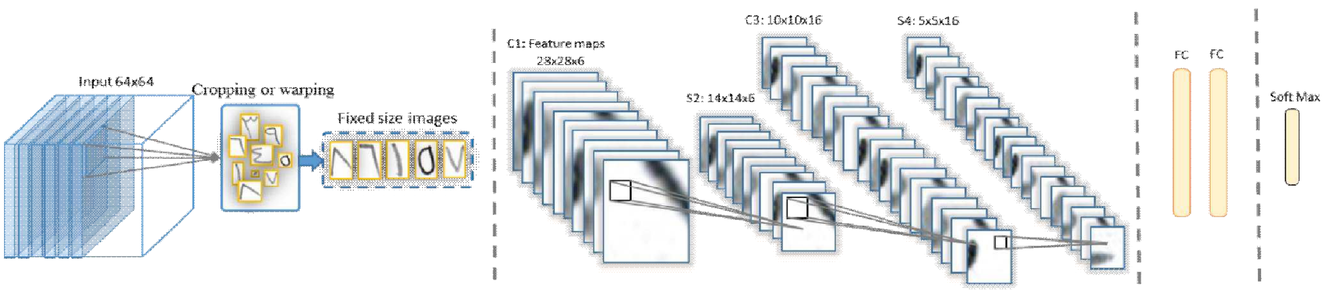

Figure.8. Final CNN model use for evaluation.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

In this experiment, we used the best open source of deep learning called Caffe [25]. In addition, the experiment is conducted using high performance Graphical Processing Units (GPUs) which has 1200 cores. Thus, the training and testing are fast which satisfy real time applications.

All parameters of CNN are randomly initialized using Gaussian distribution and the learning rate parameter is smoothly decreasing after each epoch. The number of images used in each batch is 128 samples. The network was trained with (10,000) iterations and a fast convergence was gotten after trivial number of iterations. Superior accuracy is achieved in this work which is 95.7%.

VI. CONCLUSION

In this work, a new challenging digit Arabic dataset is collected from different study levels of schools. A large dataset is collected after paying vast effort for distributing and collecting digit forms over hundreds of primary, high, college students. After we find that there are few and not challenging Arabic digit dataset, we paid vast effort for preparing such a challenging dataset.

Also the collected dataset is trained using an efficient model of CNN which represents the current state-of-the-art for variety of applications. Thus we extensively analyzed the model by carefully selecting their parameters and showing its robustness for handling our dataset.

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,Elsevier, 40 (6), pp.1816-1824, 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", 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. 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, arXiv:1406.4729v4 [cs.CV] 23 Apr 2015.
7. X. Wang, M. Yang, S. Zhu, and Y. Lin. " Regionlets for generic object detection". In ICCV, 2013.
8. 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.
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". International Journal of Computer Vision" December 2015, Volume 115, Issue 3, pp 211-252.
12. Raia Hadsell, Sumit Chopra, Yann LeCun, "Dimensionality Reduction by Learning an Invariant Mapping" CVPR , vol. 2, pp. 1735-1742, 2006.
13. Omkar M. Parkhi, Andrea Vedaldi,and Andrew Zisserman "Deep Face Recognition" British Machine Vision Conference, 2015
14. Dan Claudiu Ciresan, Ueli Meier, Luca Maria Gambardella, Juergen Schmidhuber, "Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition" Neural Computation, Volume 22, Number 12, December 2010
15. Sherif Abdel Azeem, Maha El Meseery, Hany Ahmed ,"Online Arabic Handwritten Digits Recognition ",Frontiers in Handwriting Recognition (ICFHR), 2012
16. Liu, C.L. et al.,." Handwritten digit recognition: Benchmarking of state-of-the-art techniques". Pattern Recognition 36, 2271–2285. 2003
17. LeCun, Y. et al.,." Comparison of learning algorithms for handwritten digit recognition". In: International conference on Artificial Neural networks, France, pp. 53–60. 1995
18. Xiao-Xiao Niu n , Ching Y. Suen "A novel hybrid CNN–SVM classifier for recognizing handwritten digits" Pattern Recognition 45 (2012) 1318–1325, 2011
19. M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional neural networks," arXiv:1311.2901, 2013
20. Gil Levi and Tal Hassner," Age and Gender Classification using Convolutional Neural Networks", Computer Vision and Pattern Recognition Workshops (CVPRW) IEEE, 2015
21. Aiming, 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.
22. 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
23. Yann LeCun, L´eon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.

24. 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.
25. Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor."Caffe: Convolutional Architecture for Fast Feature Embedding",arXiv preprint arXiv:1408.5093,2014