2020 International Conference on Machine Vision and Image Processing (MVIP)
Faculty of Engineering, College of Farabi, University of Tehran, Iran, 19 & 20 Feb. 2020

Improving Persian Digit Recognition by Combining Deep Neural Networks and SVM and Using PCA

Amir.M Mousavi.H

Department of Computer Engineering
Shahid Rajee University

Tehran, Iran

amir.harris@sru.ac.ir

Alireza Bossaghzadeh

Department of Computer Engineering
Shahid Rajee University

Tehran, Iran

a.bosaghzadeh@sru.ac.ir

***Abstract*—One of the machine vision tasks is optical character recognition (OCR) that researchers in this field are trying to achieve a high performance and accuracy in the classification task. In this paper, we have used a fine tuned deep Neural networks for Hoda dataset, which is the largest dataset for Persian handwritten digit classification, to extract valuable discriminative features. then, these features are fed to a linear support vector machine (SVM) for classification part. In the next experiment, In order to improve the accuracy and computational load, we applied the Principal component analysis (PCA) to reduce the extracted features dimensions then we fed it to SVM. To the best of our knowledge the proposed method was better than other methods in terms of accuracy measure**

Keywords—deep neural networks, Computer Vision, VGG16, data augmentation, SVM, PCA, feature extraction.)

I. INTRODUCTION

Deep neural networks (DNNs) are widely used in many artificial intelligence (AI) applications [1]. From speech recognition [2] and image recognition [3], to autonomous driving cars [4], from detecting cancer [5] to playing complex games. The results have shown that DNNs are able to reach human-level accuracy is recognition tasks. This is mainly due to the fact that DNNs are able to extract high-level features from a large amount of data. Furthermore, they can obtain better results compares to the classic algorithms that use hand-crafted features or rules designed by experts.

Optical character recognition (OCR) is one of the widely used tasks in machine vision. One of the active subfields of OCR is handwritten digit recognition that has a lot of applications in banking, postal services and etc. [6]. Every language has its own pattern in writing digits and Persian language is not an exception. In recent years, Persian handwritten character recognition has absorbed attention by a lot of researchers. Hoda database is the first large dataset of Persian handwritten digits which is gathered by kabir et al [7].
The rest of the article is organized as follows. The review of some of the previous works on Hoda database is described in section II. A brief explanation of deep neural networks and convolution neural networks are explained in section III. Section IV represents the proposed method. The experimental results of the proposed method are presented in section V and section VI concludes this article.

II. BACKGROUND

In this section, we review some of the recent works on Persian handwritten digits recognition. While some of the methods adopted hand-crafted feature extraction techniques, recent works have focused on deep learning methods to extract features.

Kiani et al. [8] adopted restricted Boltzmann machine for feature learning and deep belief network and spiking Neural networks (SNN) for classification part. Their proposed method was trained on 60000 samples and tested on 20000 samples and achieved the accuracy of 95%.

Shayegan et al. [9] reduced the size of dataset using modified frequency diagram and then applied principal component analysis (PCA) to reduce the feature vector's dimension and extracted 79 discriminative features. In the classification part they used a KNN classifier and took 60000 samples as training set and tested on 20000 samples that achieved the accuracy of 97.11%.

Salimi and Givaki [10] proposed singular value decomposition (SVD) based ensemble classifier. They combined the decisions of SVD with a new evolutionary rule which is called reliable multi stage particle swarm optimization. They compared their model with other classifiers such as Radial Basis Function (RBF), multilayer perceptron (MLP) and Adaptive Neuro-Fuzzy Inference System (ANFIS). Their method is robust against training images size which makes their method. They took 1000 samples as training set and tested on 5000 samples. The best accuracy that they achieved was 97.30%.

Ebrahimpour et al. [11] used a technique of location description to extract the features and for classification used a

978-1-7281-6832-6/20/$31.00 ©2020 IEEE

mixture of multilayer perceptron including 4 feed-forward neural network with 25 neurons as hidden layer and 5 neurons as input layer. Their method gains the accuracy 97.52%. Zamani et al. [12] adopted a convolutional neural network and random forest classifier and achieved the accuracy of 99.03%. Safdari and Moein [13] used a two layer sparse auto-encoder to learn hierarchical features and for the classification part used soft-max regression function. They achieved the accuracy of 98.22% on 60000 training and 20000 testing samples. Kharashazadeh and Latif [14] combined two feature extractor of histogram of oriented gradients (HoG) and chain code histogram (CCH) in four directions and extracted 264 features. For classification, they used a nonlinear support vector machine (SVM) with Gaussian kernel. Their proposed method reached 99.31% accuracy. They also used a five folds cross validation on the combination and achieved 99.58% accuracy. Alaee et al. [15] used modified boundary of digits to extract 196 power full features and used SVM for classification and achieved the accuracy of 99.02%. To increase the accuracy, they applied 5-fold cross validation and the accuracy was increased to 99.58%.

III. Convolutional Neural Networks (CNNs)

Convolutional neural network (CNN) is a common form of Deep Neural Networks which use back-propagation algorithm to train the layers [17]. Convolutional neural network (CNN) because of its ability to extract discriminative features from images is very popular in different applications of computer vision [17,3]. Convolutional neural network's training phases conclude two parts: feed-forward and back-propagation. In the feed-forward, all of the computation is performed hierarchically in order the outputs of a previous layer is the input of next layer. In the second part, the difference of the network' output and the target output which is known as error, is back propagated in the network to update the weight of the network. To update the weights, the gradient of each parameter is calculated and multiplies in a learning rate will be summed with the previous weight so the new weight is available. Then in the next step, feed-forward begins again and this loop continues. The training process is terminated after an accepted error rate or a constant number of iterations. The scheme of a convolutional neural network is illustrated in Fig. 1 and as we can see, there are three layers in CNNs including convolutional layers, pooling layers and fully connected layers. In the following, we shortly describe each layers of a CNN network.

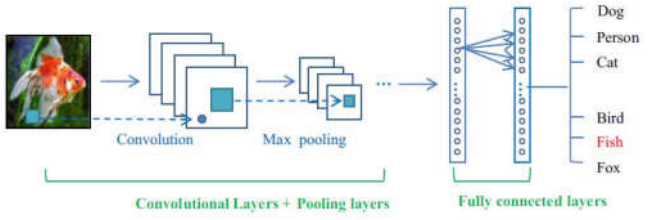

Fig. 1. general CNN architecture [22]

A. Convolutional Layers

Convolution layers are a set of filters that during the learning phase learn to respond to different kind of features. These layers are feature extractor and called CONV layers. The number, type and order of layers are determined by the complexity of the problem space and the number of input data. If the problem space is complex and a large amount of data is available, more CONV layers will be used.

As mentioned above, the core of the convolutional networks is the CONV layers. These layers include learnable weights that learn how to respond to inputs during the training process. The feature extraction by this layer is performed by convolutional operator which is known as kernel..

B. Pooling Layers

In the second step, any linear activity enters to a nonlinear activation function, such as a linear rectifier function (Relu).

In the third step, a pooling function is used to make more changes to the layer output. The pooling operator creates a degree of stability in relation to low mobility in the input and robustness to spatial translation. This means that if the input slightly translates then the maximum amount of pooling outputs will not change. The most common types of pooling layers are average pooling and max pooling. Fig. 2 demonstrates how these max and average poling perform.

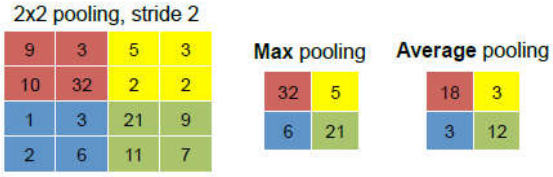

Fig. 2: Various forms of pooling (Figure adopted from Caffe Tutorial [23]).

C. Fully Connected Layers

The fully connection layers in these networks are exactly the same as the layers used in the multilayer perceptron .These layers are connected to the entire input space with weights and transform the feature maps values into a 1-dimensional feature vectors. In these layers, spatial information is lost, hence the use of convolutional layers after these layers is not meaningful. This layer is used for classification task (softmax) hence they are located at the end of network..

IV. Proposed Method

In this section, we explain the proposed method for optical character recognition.

A. Preprocessing

Since we want to use convolutional neural network architecture, hence the image size should match with the CNN input image size. The adopted CNN which is called VGGNet has the input image size of 224*224. So as a pre-processing, all images are resized to a constant size of 224*224 because our dataset images are in different sizes.

B. Feature Extraction

As mentioned before, CNNs are able to extract high-level features from a large amount of given data, much better than hand-crafted features. For this reason, we use one of the famous deep convolution neural networks which is called "VGGNet" to extract the features from the given images.

VGGNet [18] was released by VGG (Visual Geometry Group) from University of Oxford, its architecture was inspired by AlexNet [3] and ZFNet [19] but it has a significantly improvement over them only by changing the hyper-parameters. VGGNet is admired because of its simple architecture and improvement over ZFNet and AlexNet. Fig. 3 shows the architecture of the VGGNet. One of the main reasons for the improvements obtained by VGG net was that it was able to extract more discriminative features compared to its previous networks (i.e., AlexNet and ZFNet).

We feed the images into the network and extract the information in the last layer before the classification layer (i.e., FC2) and use it as the feature of the image for classification.

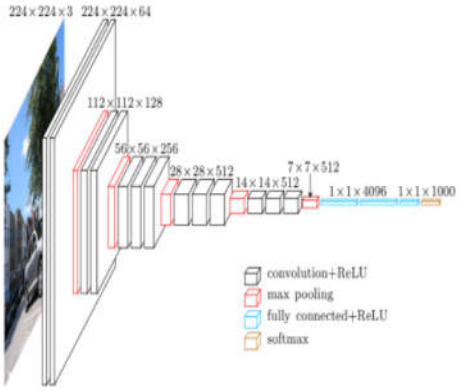

Fig. 3: VGGNet architecture

C. Classification

For the classification task we used a linear support vector machine (SVM). Our ideology was to extract high discriminative features and use a simple classifier for image classification. On the other hand, due to high number of extracted features SVM has been used to improves the performance of classification instead of a 10 classes softmax classifier in the VGG Net.

D. Dimensionality Reduction

Principal component analysis is an approach to factor analysis that extract important variables (in the form of components) from a large set of variables in a dataset. In fact, it extracts low-dimensional set of extracted features from a high-dimensional set to help record more important features with fewer variables. The principal component analysis considers the total variance in the data, and transforms the original variables into a smaller set of linear combinations and reduced the dimension .

Since the extracted feature from FC2 layer has 4096 dimensions, to prevent the curse of dimensionality, we apply PCA and select 100 and 150 components.

V. Experimental Results

In this section, we explain the experimental results obtained by the proposed method and some of the state of the art methods.

We fine-tuned the VGGNet on the HODA dataset by removing the last layer and replace it with a 10 class softmax. Then we feed all images into the network and extract the FC2 layer as the image descriptor.

After that, we applied PCA to reduce the 4096 dimension FCfeature. Finally, the result of PCA is fed tothe trained .SVM for classification.

A. Dataset

We use Hoda dataset as our experimental dataset which is gathered by E.kabir et al [7]. It is the first dataset of Persian handwritten digits that includes 10 classes of 0-9 Persian digits which is extracted from about 12000 registration forms of university entrance examination in Iran. It contains 60000 samples for train and 20000 samples for test. Fig. 4 shows some sample images of this database.

Fig. 4: Sample images of HODA database.

B. Results

We compare the accuracy of our method with that of [20]. We took exactly the same number of training and test samples as [20]. The experiment was repeated 10 times for each train-test set and the average of the accuracy is reported. Table 1 shows the results of the proposed method for different training and testing rates. As we can see the proposed method is able gain higher accuracy compared to the method proposed by [20]. Furthermore, the obtained accuracy is higher than that of VGGNet with 10 class softmax fine tuned and without PCA dimensionality reduction. It shows that all building blocks of the proposed method are vital to gain this high accuracy.

Table 2 shows the confusion matrix of the proposed method. As we observe, the proposed algorithm has almost equal accuracy in all classes and does not have any bias toward a specific class.

Also in Fig. 5 we can see some of the misclassified images. As we can see, the correct labels of the images are hard to be determined even for a human.

Furthermore, Table 3 compares the accuracy of the proposed method with other five state of the art algorithms in this area. Again, we can see that the proposed algorithm outperforms other competing method and gain higher accuracy.

Table 1: The accuracy of the proposed method using different training size images

# of samples	Training samples	Test samples	Average accuracy [20]	VGG Net (10-classes SOFTMAX)	Proposed method	
					Without PCA	With PCA
60000	0.7	0.3	0.9922	0.9974	0.9982	0.9983
60000	0.5	0.5	0.9918	0.9965	0.9980	0.9981
60000	0.3	0.7	0.9907	0.9959	0.9978	0.9978
80000	0.7	0.3	0.9916	0.9964	0.9974	0.9976
80000	0.5	0.5	0.9915	0.9960	0.9975	0.9975
80000	0.3	0.7	0.9899	0.9957	0.9966	0.9965

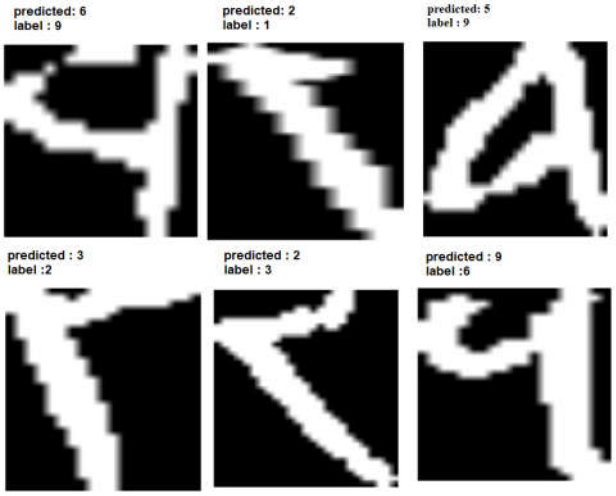

Fig. 5:Some wrong predicted samples

Table 2:Confusion matrix of 80000 samples (0.7 for training and 0.3 for test set) after applying PCA and SVM

	0	1	2	3	4	5	6	7	8	9
0	1688	1	0	0	0	0	0	2	0	0
1	0	1712	0	0	0	0	1	0	0	2
2	0	0	1761	2	0	0	0	0	0	0
3	0	0	6	1705	2	0	0	0	0	0
4	0	0	0	1	1629	0	1	0	0	0
5	1	0	0	0	0	1662	0	0	0	0
6	0	1	0	0	0	1	1715	0	0	1
7	0	0	0	0	0	0	2	1701	0	0
8	0	1	0	0	0	0	0	0	1745	0
9	0	0	0	0	0	0	3	0	0	1681

Table 3: Comparison of the proposed method with some state of the art OCR methods applied on HODA database.

[20]	[11]	[15]	[21]	[14]	Our method
98.75	97.52	99.02	99.30	99.31	99.69

VI. Conclusion

In this paper, a fine tuned convolutional neural network has been used for feature extraction on Hoda dataset and along with a PCA for dimensionality reduction and a SVM as classifier. As it is demonstrated by the results all of the building blocks of the proposed methods are essential for the improved obtained compared to state of the art algorithms performed on the same database. For the future work, we will focus on new architectures with higher accuracy and use non-linear classifiers for Persian handwritten digit recognition.

References

[1] LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. nature 521 (7553), 436-444. DOI: http://dx. doi. org/10.1038/nature14539, 2015.

[2] Deng, L., et al. Recent advances in deep learning for speech research at Microsoft. in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. 2013. IEEE.

[3] Krizhevsky, A., I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. in Advances in neural information processing systems. 2012.

[4] Chen, C., et al. Deepdriving: Learning affordance for direct perception in autonomous driving. in Proceedings of the IEEE International Conference on Computer Vision. 2015.

[5] Esteva, A., et al., Dermatologist-level classification of skin cancer with deep neural networks. Nature, 2017. 542(7639): p. 115..

[6] Ebrahimpour, R. and S. Hamedi, Hand written digit recognition by multiple classifier fusion based on decision templates approach. World Academy of Science, Engineering and Technology, 2009. 57: p. 560-565.

[7] Khosravi, H. and E. Kabir, Introducing a very large dataset of handwritten Farsi digits and a study on their varieties. Pattern recognition letters, 2007. 28(10): p. 1133-1141

[8] Kiani, K. and E.M. Korayem. Classification of Persian handwritten digits using spiking neural networks. in 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI). 2015. IEEE.

[9] Shayegan, M.A. and S. Aghabozorgi, A new dataset size reduction approach for PCA-based classification in OCR application. Mathematical Problems in Engineering, 2014. 2014.

[10] Salimi, H. and D. Giveki, Farsi/Arabic handwritten digit recognition based on ensemble of SVD classifiers and reliable multi-phase PSO combination rule. International Journal on Document Analysis and Recognition (IJDAR), 2013. 16(4): p. 371-386.

[11] Ebrahimpour, R., et al., Recognition of Persian handwritten digits using Characterization Loci and Mixture of Experts. International Journal of Digital Content Technology and its Applications, 2009. 3(3): p. 42.46.

[12] Zamani, Y., et al. Persian handwritten digit recognition by random forest and convolutional neural networks. in 2015 9th Iranian Conference on Machine Vision and Image Processing (MVIP). 2015. IEEE.

[13] Safdari, R. and M.-S. Moin. A hierarchical feature learning for isolated Farsi handwritten digit recognition using sparse autoencoder. in 2016 Artificial Intelligence and Robotics (IRANOPEN). 2016. IEEE.

[14] Khorashadizadeh, S. and A. Latif, Arabic/Farsi Handwritten Digit Recognition usin Histogra of Oriented Gradient and Chain Code Histogram. International Arab Journal of Information Technology (IAJIT), 2016. 13(4).

[15] Alaei, A., P. Nagabhushan, and U. Pal. Fine classification of unconstrained handwritten Persian/Arabic numerals by removing confusion amongst similar classes. in 2009 10th International Conference on Document Analysis and Recognition. 2009. IEEE.

[16] Sze, V., et al., Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE, 2017. 105(12): p. 2295-2329.

[17] Guo, Y., et al., Deep learning for visual understanding: A review. Neurocomputing, 2016. 187: p. 27-48.

[18] Simonyan, K. and A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[19] Zeiler, M.D. and R. Fergus. Visualizing and understanding convolutional networks. in European conference on computer vision. 2014. Springer.

[20] Farahbakhsh, E., E. Kozegar, and M. Soryani. Improving persian digit recognition by combining data augmentation and AlexNet. in 2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP). 2017. IEEE.

[21] Ghanbari, N., 2019. A Review of Research Studies on the Recognition of Farsi Alphabetic and Numeric Characters in the Last Decade. In Fundamental Research in Electrical Engineering (pp. 173-184). Springer, Singapore.

[22] Girshick, R., et al. Rich feature hierarchies for accurate object detection and semantic segmentation. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[23] Jia, Y., et al. Caffe: Convolutional architecture for fast feature embedding. in Proceedings of the 22nd ACM international conference on Multimedia. 2014. ACM.