Face Recognition Based on Convolutional Neural Network

Musab Coşkun
Faculty of Electrical and Electronics Engineering
Fırat University
Elazığ, Turkey
musabcoskun@yandex.com

Ayşegül Uçar
Faculty of Mechatronics Engineering
Fırat University
Elazığ, Turkey
agulucar@firat.edu.tr

Özal Yıldırım
Faculty of Computer Engineering
Munzur University
Tunceli, Turkey
yildirimozal@hotmail.com

Yakup Demir
Faculty of Electrical and Electronics Engineering
Fırat University
Elazığ, Turkey
ydemir@firat.edu.tr

***Abstract*—Face recognition is of great importance to real world applications such as video surveillance, human machine interaction and security systems. As compared to traditional machine learning approaches, deep learning based methods have shown better performances in terms of accuracy and speed of processing in image recognition. This paper proposes a modified Convolutional Neural Network (CNN) architecture by adding two normalization operations to two of the layers. The normalization operation which is batch normalization provided acceleration of the network. CNN architecture was employed to extract distinctive face features and Softmax classifier was used to classify faces in the fully connected layer of CNN. In the experiment part, Georgia Tech Database showed that the proposed approach has improved the face recognition performance with better recognition results.**

Keywords—face recognition, convolutional neural network, softmax classifier, deep learning.

I. Introduction

Face recognition is the process of recognizing the face of a relevant person by a vision system. It has been a crucial human-computer interaction tool due to its usage in security systems, access-control, video surveillance, commercial areas and even it is used in social networks like Facebook as well. After rapid development of artificial intelligence, face recognition has once again attracted attention due to its non-intrusive nature and since it is main method of person identification for human when it is compared with other types of biometric techniques. Face recognition can be easily checked without the subject person's knowledge in an uncontrolled environment.

As the history of face recognition is surveyed, it is seen that it has been addressed in many research papers e.g. [1]–[6]. Traditional methods based on shallow learning have been facing challenges like pose variation, facial disguises, lighting of the scene, the complexity of the image background, and changes in facial expression as in references [7]–[17]. Shallow learning based methods only utilize from some basic features of images and depend on artificial experience to extract sample features. Deep learning based methods can extract more complicated face features [18]–[27]. Deep learning is making crucial advances in solving problems that have restricted the best attempts of the artificial intelligence community for many years. It has proven to be excellent at revealing complex structures in high-dimensional data and is therefore applicable to lots of domains of science, business and government. It addresses the problem of learning hierarchical representations with a single algorithm or a few algorithms and has mainly beaten records in image recognition, natural language processing, semantic segmentation and many other real world scenarios [28]–[35]. There are different deep learning approaches like Convolutional Neural Network (CNN), Stacked Autoencoder [36], and Deep Belief Network (DBN) [37], [38]. CNN is mostly used algorithm in image and face recognition. CNN is a kind of artificial neural networks that employs convolution methodology to extract the features from the input data to increase the number of features. CNN was firstly proposed by LeCun and was firstly applied in handwriting recognition [39]. His network was the origin of much of the recent architectures, and a true inspiration for many scientists in the field. Krizhevsky, Sutskever and Hinton achieved best results when they published their work in ImageNet Competition [40]. It is widely regarded as one of the most influential publications in computer vision and showed that CNNs outperform recognition performances compared to hand-crafted based methods. With computational power of Graphical Processing Units (GPUs), CNN has achieved remarkable cutting edge results over a number of areas, including image recognition, scene recognition, semantic segmentation, and edge detection.

The main contribution of this paper is to obtain a powerful recognition algorithm with high accuracy. In this paper, we developed a new CNN architecture by adding Batch Normalization process after two different layers.

The general structure of face recognition process in this paper is made up of three stages. It starts with pre-processing stage: colour space conversion and resize of images, continues with extraction of facial features, and afterwards extracted feature set is classified. In our system, Softmax Classifier is to realize final stage that is classification on the basis of the facial features extracted from CNN.

The rest of this paper is organized as follows. In section 2, CNN architecture is introduced. In section 3, the proposed algorithm is discussed. In section 4, the face database used in this paper is presented. The experimental results are given in section 5. Finally, we discuss conclusions in section 6.

978-1-5386-1750-2/17/$31.00 ©2017 IEEE

II. METHODOLOGY

CNNs are a category of Neural Networks that have proven very effective in areas such as image recognition and classification. CNNs are a type of feed-forward neural networks made up of many layers. CNNs consist of filters or kernels or neurons that have learnable weights or parameters and biases. Each filter takes some inputs, performs convolution and optionally follows it with a non-linearity[41]. A typical CNN architecture can be seen as shown in Fig.1. The structure of CNN contains Convolutional, pooling, Rectified Linear Unit (ReLU), and Fully Connected layers.

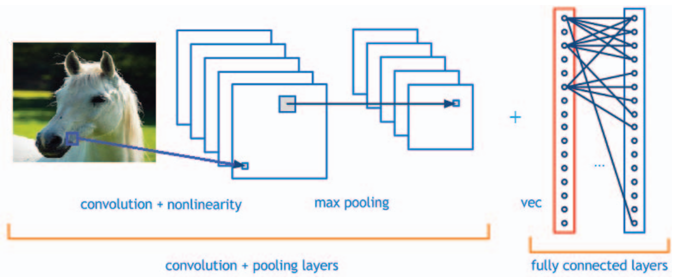

Fig. 1. A traditional Convolutional Neural Networks design.

A. Convolutional Layer

Convolutional layer performs the core building block of a Convolutional Network that does most of the computational heavy lifting. The primary purpose of Convolution layer is to extract features from the input data which is an image. Convolution preserves the spatial relationship between pixels by learning image features using small squares of input image. The input image is convoluted by employing a set of learnable neurons. This produces a feature map or activation map in the output image and after that the feature maps are fed as input data to the next convolutional layer.

B. Pooling Layer

Pooling layer reduces the dimensionality of each activation map but continues to have the most important information. The input images are divided into a set of non-overlapping rectangles. Each region is down-sampled by a non-linear operation such as average or maximum. This layer achieves better generalization, faster convergence, robust to translation and distortion and is usually placed between convolutional layers.

C. ReLU Layer

ReLU is a non-linear operation and includes units employing the rectifier. It is an element wise operation that means it is applied per pixel and reconstitutes all negative values in the feature map by zero. In order to understand how the ReLU operates, we assume that there is a neuron input given as x and from that the rectifier is defined as $f(x) = max(0,x)$ in the literature for neural networks.

D. Fully Connected Layer

Fully Connected Layer (FCL) term refers to that every filter in the previous layer is connected to every filter in the next layer. The output from the convolutional, pooling, and ReLU layers are embodiments of high-level features of the input image. The goal of employing the FCL is to employ these features for classifying the input image into various classes based on the training dataset. FCL is regarded as final pooling layer feeding the features to a classifier that uses Softmax activation function. The sum of output probabilities from the Fully Connected Layer is 1. This is ensured by using the Softmax as the activation function. The Softmax function takes a vector of arbitrary real-valued scores and squashes it to a vector of values between zero and one that sum to one.

III. THE PROPOSED ALGORITHM

The block schema of the proposed CNN recognition algorithm is given in Fig. 2. The algorithm is mainly carried out in three steps as below:

1) Resize the input images as 16x16x1, 16x16x3, 32x32x1, 32x32x3, 64x64x1, and 64x64x1.

2) Build a CNN structure with eight layers made up of convolutional, max pooling, convolutional, max pooling, convolutional, max pooling, convolutional, and convolutional layers respectively.

3) After extracting all features, use Softmax classifier for classification.

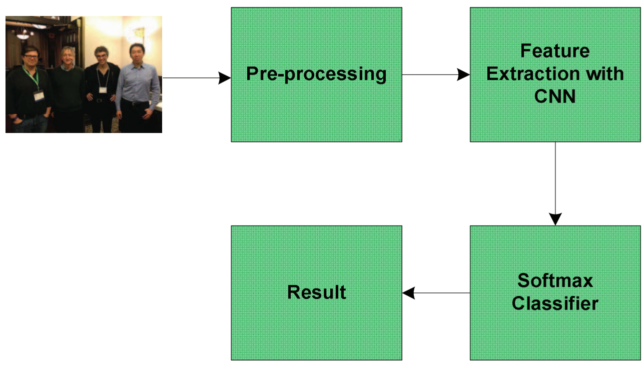

Fig. 2. The block schema of the proposed algorithm.

In Fig. 3, the structure of feature extraction block of the proposed CNN is illustrated.

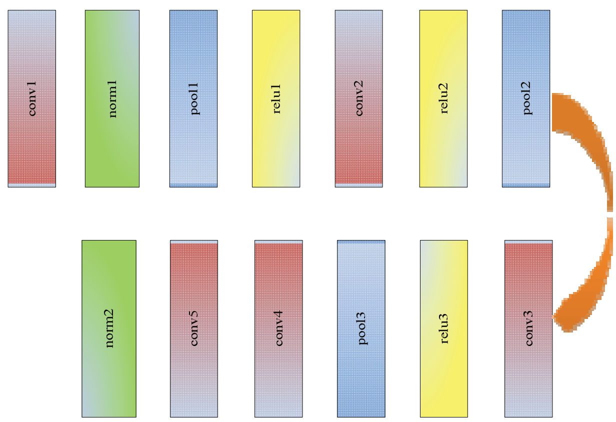

Fig. 3. The structure of feature extraction block of the proposed CNN.

IV. DATABASE

Georgia Tech face database contains images of 50 individuals taken in two or three sessions between 06/01/99 and 11/15/99 at different times at the Centre for Signal and Image Processing at Georgia Institute of Technology. Each individual in the database is represented by 15 coloured JPEG images with cluttered background taken at resolution 640×480 pixels. The average size of the faces in these images is 150×150 pixels. The images show frontal and/or tilted face with different facial expressions, lighting conditions and scale. All of the face regions in the images in the database were resized as 82×94. Fig. 4 presents some face images of different subjects from the GT database [42].

Fig. 4. Some face images of different subjects of the Georgia Tech database.

V. EXPERIMENTAL RESULTS

We designed our CNN with Beta23 version of MatConvNet software tool. After pre-processing stage, size of each image was changed as 16x16x1, 16x16x3, 32x32x1, 32x32x3, 64x64x1, and 64x64x3. 66% of images were assigned as training set, 34% as test set. We implemented different tests by making changes in image size, learning rate, batch size, and etc. CNN was trained for 35 epochs. Performance of the proposed CNN was evaluated according to top-1 and top-5 errors. Top-1 error rate checks if the top class is the same as the target label and top-5 error rate checks if the target label is one of your top five predictions. A brief structure of the proposed algorithm is depicted in Table 1. The results are better than those in the literature that use shallow learning techniques like in references [43-45].

TABLE I. THE PARAMETERS OF THE PROPOSED ALGORITHM

Input Image size	Number of Epoch reached at the highest rate for Top-1 Error	Number of Epoch reached at the highest rate for Top-5 Error	Batch Size	Learning Rate	Top-1 Error (Accuracy Rate)	Top-5 Error (Accuracy Rate)
16x16x1	20	24	10	0.001	88.8	98.4
16x16x3	34	21	20	0.001	93.2	98.8
32x32x1	21	17	30	0.001	90	98
32x32x3	17	31	20	0.001	92.8	98.8
64x64x1	19	18	30	0.001	92.4	98.4
64x64x3	21	28	10	0.001	94.8	98.8

The performance of the proposed CNN architecture in terms of top-1 error rate is shown in Figure 5. As seen from the Figure 5, the lowest top-1 error rate was obtained from 64x64x3 sized image. This result matters when it is aimed to find target label of any subject in the database. Top-5 error rate is presented in Figure 6 and the lowest rate was achieved from all of the images with three channels.

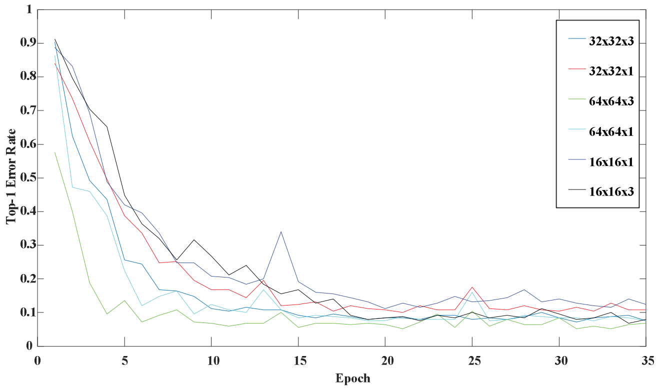

Fig. 5. Top-1 error rate of the proposed CNN.

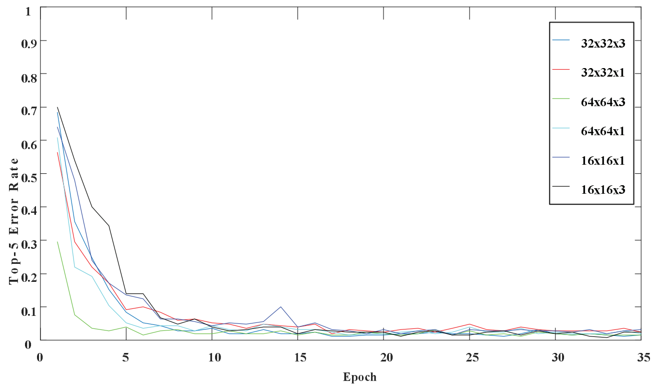

Fig. 6. Top-5 error rate of the proposed CNN.

IV. CONCLUSION

This paper presents an empirical evaluation of face recognition system based on CNN architecture. The prominent features of the proposed algorithm is that it employs the batch normalization for the outputs of the first and final convolutional layers and that makes the network reach higher accuracy rates. In fully connected layer step, Softmax Classifier was used to classify the faces. The performance of the proposed algorithm was tested on Georgia Tech Face Database. The results showed satisfying recognition rates according to studies in the literature.

REFERENCES

[1] S. G. Bhele and V. H. Mankar, "A Review Paper on Face Recognition Techniques," Int. J. Adv. Res. Comput. Eng. Technol., vol. 1, no. 8, pp. 2278–1323, 2012.

[2] V. Bruce and A. Young, "Understanding face recognition," Br. J. Psychol., vol. 77, no. 3, pp. 305–327, 1986.

[3] D. N. Parmar and B. B. Mehta, "Face Recognition Methods & Applications," Int. J. Comput. Technol. Appl., vol. 4, no. 1, pp. 84–86, 2013.

[4] W. Zhao et al., "Face Recognition: A Literature Survey," ACM Comput. Surv., vol. 35, no. 4, pp. 399–458, 2003.

[5] K. Delac, Recent Advances in Face Recognition. 2008.

[6] A. S. Tolba, A. H. El-baz, and A. A. El-Harby, "Face Recognition : A Literature Review," Int. J. Signal Process., vol. 2, no. 2, pp. 88–103, 2006.

[7] C. Geng and X. Jiang, "Face recognition using sift features," in Proceedings - International Conference on Image Processing, ICIP, pp. 3313–3316, 2009.

[8] S. J. Wang, J. Yang, N. Zhang, and C. G. Zhou, "Tensor Discriminant Color Space for Face Recognition," IEEE Trans. Image Process., vol. 20, no. 9, pp. 2490–501, 2011.

[9] S. N. Borade, R. R. Deshmukh, and S. Ramu, "Face recognition using fusion of PCA and LDA: Borda count approach," in 24th Mediterranean Conference on Control and Automation, MED 2016, pp. 1164–1167, 2016.

[10] M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces," Journal of Cognitive Neuroscience, vol. 3, no. 1. pp. 72–86, 1991.

[11] M. O. Simón, "Improved RGB-D-T based face recognition," IET Biometrics, vol. 5, no. 4, pp. 297–303, Dec. 2016.

[12] O. Dniz, G. Bueno, J. Salido, and F. De La Torre, "Face recognition using Histograms of Oriented Gradients," Pattern Recognit. Lett., vol. 32, no. 12, pp. 1598–1603, 2011.

[13] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 2, pp. 210–227, 2009.

[14] C. Zhou, L. Wang, Q. Zhang, and X. Wei, "Face recognition based on PCA image reconstruction and LDA," Opt. - Int. J. Light Electron Opt., vol. 124, no. 22, pp. 5599–5603, 2013.

[15] Z. Lei, D. Yi and S. Z. Li, "Learning Stacked Image Descriptor for Face Recognition," IEEE Trans. Circuits Syst. Video Technol., vol. 26, no. 9, pp. 1685–1696, Sep. 2016.

[16] P. Sukhija, S. Behal, and P. Singh, "Face Recognition System Using Genetic Algorithm," in Procedia Computer Science, vol. 85, 2016.

[17] S. Liao, A. K. Jain, and S. Z. Li, "Partial face recognition: Alignment-free approach," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 5, pp. 1193–1205, 2013.

[18] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Learning Deep Representation for Face Alignment with Auxiliary Attributes," IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 5, pp. 918–930, 2016.

[19] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2518–2525, 2012.

[20] S. Lawrence, C. L. Giles, Ah Chung Tsoi, and A. D. Back, "Face recognition: a convolutional neural-network approach," IEEE Trans. Neural Networks, vol. 8, no. 1, pp. 98–113, 1997.

[21] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in Procedings of the British Machine Vision Conference 2015, pp. 41.1-41.12, 2015.

[22] Z. P. Fu, Y. N. ZhANG, and H. Y. Hou, "Survey of deep learning in face recognition," in IEEE International Conference on Orange Technologies, ICOT 2014, pp. 5–8, 2014.

[23] X. Chen, B. Xiao, C. Wang, X. Cai, Z. Lv, and Y. Shi, "Modular hierarchical feature learning with deep neural networks for face verification," Image Processing (ICIP), 2013 20th IEEE International Conference on. pp. 3690–3694, 2013.

[24] Y. Sun, D. Liang, X. Wang, and X. Tang, "DeepID3: Face Recognition with Very Deep Neural Networks," Cvpr, pp. 2–6, 2015.

[25] G. Hu, "When Face Recognition Meets with Deep Learning: An Evaluation of Convolutional Neural Networks for Face Recognition," 2015 IEEE Int. Conf. Comput. Vis. Work., pp. 384–392, 2015.

[26] C. Ding and D. Tao, "Robust Face Recognition via Multimodal Deep Face Representation," IEEE Trans. Multimed., vol. 17, no. 11, pp. 2049–2058, 2015.

[27] A. Bharati, R. Singh, M. Vatsa, and K. W. Bowyer, "Detecting Facial Retouching Using Supervised Deep Learning," IEEE Trans. Inf. Forensics Secur., vol. 11, no. 9, pp. 1903–1913, 2016.

[28] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3367–3375, 2015.

[29] P. Pinheiro and R. Collobert, "Recurrent convolutional neural networks for scene labeling," Proc. 31st Int. Conf., vol. 32, no. June, pp. 82–90, 2014.

[30] W. Shen, X. Wang, Y. Wang, X. Bai, and Z. Zhang, "DeepContour: A deep convolutional feature learned by positive-sharing loss for contour detection," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 07–12, pp. 3982–3991, June 2015.

[31] M. A. K. Mohamed, A. El-Sayed Yarub, and A. Estaitia, "Automated Edge Detection Using Convolutional Neural Network," Int. J. Adv. Comput. Sci. Appl., vol. 4, no. 10, pp. 11–17, 2013.

[32] Dan Cireşan, "Deep Neural Networks for Pattern Recognition."

[33] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural Language Processing (Almost) from Scratch," J. Mach. Learn. Res., vol. 12, pp. 2493–2537, 2011.

[34] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," Proc. 25th Int. Conf. Mach. Learn., pp. 160–167, 2008.

[35] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 640–651, 2017.

[36] R. Xia, J. Deng, B. Schuller, and Y. Liu, "Modeling gender information for emotion recognition using Denoising autoencoder," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings, pp. 990–994, 2014.

[37] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," Neural Comput., vol. 18, no. 7, pp. 1527–1554, 2006.

[38] Y. Bengio, "Learning Deep Architectures for AI," vol. 2, no. 1, 2009.

[39] Y. LeCun, "Backpropagation Applied to Handwritten Zip Code Recognition," Neural Comput., vol. 1, no. 4, pp. 541–551, Dec. 1989

[40] A. Krizhevsky, I. Sutskever, and H. E. Geoffrey, "ImageNet Classification with Deep Convolutional Neural Networks," Adv. Neural Inf. Process. Syst. 25, pp. 1–9, 2012.

[41] A. Uçar, Y. Demir, and C. Guzelis, "Object Recognition and Detection with Deep Learning for Autonomous Driving Applications," Simulation, pp. 1-11, 2017.

[42] "Georgia Tech face database," 10-Jan-2017 [Online]. Available: http://www.anefian.com/research/face_reco.htm.

[43] Nischal K N, Praveen Nayak M, K Manikantan, and S Ramachandran, "Face Recognition using Entropy-augmented face isolation and Image folding as pre-processing techniques," 2013 Annual IEEE India Conference (INDICON), 2013.

[44] Katia Estabridis, "Face Recognition and Learning via Adaptive Dictionaries," IEEE Conference on Technologies for Homeland Security (HST), 2012.

[45] Qiong Kang and Lingling Peng, "An Extended PCA and LDA for color face recognition," International Conference on Information Security and Intelligence Control (ISIC), 2012.