2017 2nd International Conference on Manufacturing Science and Information Engineering (ICMSIE 2017)
ISBN: 978-1-60595-516-2

Robust Face Recognition Based on Convolutional Neural Network

Ying Xu, Hui Ma, Lu Cao*, He Cao, Yikui Zhai,
Vincenzo Piuri and Fabio Scotti

ABSTRACT

Face recognition via deep learning has achieved a series of breakthrough in recent years. The deeply learned features are required to be discriminative and generalized enough for identifying new unseen classes without label prediction. Therefore, this paper improved the performance of deep learning by aligning the training data and enhancing the preprocessing. Moreover, the designed model, named Lightened Convolutional Neural Network model, and the center loss layer jointly to enhance the discriminative of the designed network features. The network is trained on the self-expanding CASIA-WebFace database and tested on the Labeled Faces in the Wild (LFW) database. Experimental results show that the proposed network model brings significant improvement in the accuracy of face recognition, compared with the original CNN model.[1]

INTRODUCTION

During the last decades, deep learning, as an emerging multi-layer neural network algorithm of machine learning field and artificial intelligence (AI), has obtained great achievements. However, there are still many issues to be solved urgently, such as the construction and selection of the depth structure, the training method of the deep network, the parallel calculation under the large-scale data, the multi-information fusion and so forth.

[1]Ying Xu, Hui Ma, Lu Cao, He Cao, Yikui Zhai, School of Electronic and Information on Engineer, Wuyi University, Jiangmen, China, 529020.
Vicenzo Piuri, Fabio Scotti, Department of Computer Science, University' degli Studi di Milano, Crema, Italy, 26013.
*Lu Cao (caolu20001742@163.com) is the corresponding author.

Face recognition based on deep learning [1, 2] has achieved a series of breakthroughs in recent years. DeepFace [3] was used to obtain a feature extractor by training a 4.4M face image and firstly used it for face verification tasks. As an extension of DeepFace, Web-Scale [4] selects an effective training set from a large data set by semantic bootstrapping. The FaceNet [1] network, designed with the introduction of triplet loss in the CNN model, was trained about 8M different identities of 100-200 person's face images.

APPROACHES

The network model architecture includes 4 convolution layers, 4 maximum feature map activation functions, 4 maximum pooling layers and 2 fully connected layers. The input of the proposed network model is a 144 × 144 grayscale image. To increase the number of samples, we randomly crop the image into 128 × 128 and set the image as the first convolution layer's input during the training phase. As the input of the center loss function, the first fully connection layer fc1 represents a 512-dimensional face image. The second fully convolution layer fc2 is used as the input of the softmax loss function, and the output size of softmax loss function is set to 15511 (the size of CASIA-WebFace face database).

Network Architecture

The improved lightened convolution neural network architecture is shown in Table.1. we train the proposed model under the jointly supervision of softmax loss and center loss, with maximum feature map activation function.

Maximum Feature Map Activation Function (MFM)

Maxout network that the maximum feature map activation function (MFM) is derived in order to replace the close representation of the ReLU sparse representation. Maxout is a non-linear activation function, which has the characteristics of fitting arbitrary convex function and converging the network to better global solution. Given an input convolution layer as shown in Fig.1, the maximum feature map activation function can be expressed as:

$$f_{ij}^{k} = \max_{1 \le k \le n}(C_{ij}^{k}, C_{ij}^{k+n}) \tag{1}$$

Where the input convolution layer of the channel is $2n$, $1 \le i \le h$, $1 \le j \le w$, through MFM, the equation (1) outputs $f \in R^{h \times w \times n}$. The gradient of the activation function by equation (1) can be expressed as:

$$\frac{\partial f}{\partial C^{k}} = \begin{cases} 1, & \text{if } C_{ij}^{k} \ge C_{ij}^{k+n} \\ 0, & other \end{cases} \tag{2}$$

where $1 \le k \le 2n$ and

$$k = \begin{cases} k' & 1 \le k' \le n \\ k' - n & n+1 \le k' \le 2n \end{cases} \tag{3}$$

TABLE I. THE PROPOSED CONVOLUTION NEURAL NETWORK ARCHITECTURE.

proposed model		
Layers	Size	Parameters
conv1	$(9,48)_{1,0}\times 2$	7.6K
mfm1	$(9,48)_{1,0}$	
pool1	$2/_{2,0}$	
conv2	$(5,96)_{1,0}\times 2$	4.8K
mfm2	$(5,96)_{1,0}$	
pool2	$2/_{2,}0$	
conv3	$(5,128)_{1,0}\times 2$	6.4K
mfm3	$(5,128)_{1,0}$	
pool3	$2/_{2,0}$	
conv4	$(4,192)_{1,0}\times 2$	6K
mfm4	$(4,192)_{1,0}$	
pool4	$2/_{2,0}$	
fc1	512	2,457K
fc2	15511	2,707K
softmax	15511	

Equation (3) indicates that the gradient of the 50% active layer is zero. Therefore, MFM can get a sparse gradient.

MFM activation function is the output of the two-convolution features map of the maximum node. By the using of the total statistical method, MFM activation function can obtain a compact representation and a sparse gradient at the same time, and in this way, we can achieve the choice of variables and realize the dimensionality reduction. In addition, the MFM activation function also can be considered as a sparse connection between two convolution layers.

EXPERIMENTS

Configuration:

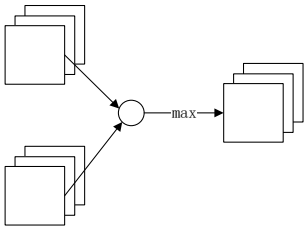

Figure 1. MFM function operations.

Using the open source deep learning Caffe [5] library to improve the CNN model and train, the training process hardware environment are Intel 80G RAM and GPU for Tesla K20c. Considering the current database with different face image size, angle changes, illumination and so forth, we introduce face image alignment and data enhancement preprocessing before training. All the faces images and their landmarks are detected by the recently proposed algorithms [6]. Due to the influence of illumination, some of the face image features may be highlighted or weakened, correspondingly, this cause may result to the feature is characterized by a higher or a lower gray values on the digital face images. One of the effectively approach is enhance the R, G, B three-channel, that means, adding a random number of [-50,50] to each channel in the original database.

Experiments on LFW

Improved lightened convolution neural network model is evaluated by (1: 1) [7] and the probe-gallery identification protocol (1: N) [8] on the unconstrained environment face recognition benchmark LFW.

For Face Verification Protocol (1: 1), 13,233 face images of LFW database can generate positive sample pairs and negative sample pairs. Selecting 3,000 pairs from these positive pairs and 3,000 pairs from negative samples pairs randomly, and comparing the 6,000 pairs of positive and negative samples on the two images of the similar distance. Through the similar scores of two images in positive and negative samples to calculate the accuracy, and using the correct reception rate as the model performance.

For the probe-gallery authentication protocol (1:N), we can divide it into two new protocols—closed task and open task:

For closed task: The gallery contains 4,249 images of 4,249 identities totally. The probe contains 3,033 images of 600 identities. Making the cosine similarity

measure between the probe and the image feature in the gallery and calculating Rank-1.

For open task: Setting the probe selected from the closed task as a gallery in the open task—3,033 images of 600 identities. Using the remaining images as probe, so probe contains 10,200 images from 5,749 identities totally, which includes 600 genuine probe and 9,600 impostors. Making each image in the probe match the similarity of all images in the gallery, and when the impostor's false alarm rate is 1%, calculating Rank-1.

The experimental results are shown in Table II. As can be seen from Table II, our model outperform DeepFace model with greater disparity, and improve the performance on LFW with 97.36% to 99.23%. The results of designed model in this paper on LFW verification protocol is competitive with those of DeepFace [1], DeepID-2+ [9], VGG [10], Model A [11] and Model B [11] for single net, by using a novel lighten design. For the probe-gallery identification protocol, our model's performance also outperforms over the results of VGG with 95.16% and 62.37%, respectively.

TABLE II. MODEL VALIDATION ON THE LFW DATABASE.

Method	**Images**	**#Net**	**Accuracy**	**TPR@FAR =0.1%**	**Rank-1**	**TPR@FAR =0.1%**
DeepFace [3]	4M	3	97.35%	-	-	-
DeepID-2+[9]	-	1	98.70%	-	-	-
FaceNet[1]	200M	-	99.63%	-	-	-
VGG[10]	4M	1	97.27%	81.90%	74.10%	52.01%
Model A[11]	-	1	97.77%	84.37%	84.79%	63.09%
Model B[11]	-	1	98.13%	87.13%	89.21%	69.46%
Our model	1.5M	1	97.36%	83.5%	91.28%	68.62%

CONCLUSIONS

In this paper, we designed a lightened convolution neural network model which added a central loss layer. By aligning face image, the network training, enhancing the preprocessing, compare with DeepFace, DeepID-2+ and VGG, the proposed network model in this paper can obtain competitive recognition accuracy, which is a satisfying result on LFW database. Though FaceNet can achieved the best result of 99.63%, but the proposed network is designed as a more on lightened one which can be applied on mobile device in the future.

ACKNOWLEDGMENTS

This work is supported by National of Nature Science Foundation Grant (No.61372193, No.61771347), Guangdong Higher Education Outstanding Young Teachers Training Program Gran(No.SYQ2014001), Characteristic Innovation Project of Guangdong Province(No.2015KTSCX 143, 2015KTSCX145, 2015KT SCX148), Youth Innovation Talent Project of Guangdong Province(No.2015KQ NCX172, No.2016KQNCX171, 2015KQNCX165), Science and Technology Project of Jiangmen City (No.201501003001556,No.201601003002191), and China National Oversea Study Scholarship Fund.

REFERENCES

1. Schroff, F., Kalenichenko, D., Philbin. 2015. "Facenet: A unified embedding for face recognition and clustering, " J. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 815–823,
2. Cevikalp, Hakan, and Bill Triggs. 2010. "Face recognition based on image sets." Computer Vision and Pattern Recognition. pp: 2567-2573.
3. Taigman, Yaniv, et al. 2014. "Deepface: Closing the gap to human-level performance in face verification." Proceedings of the IEEE conference on computer vision and pattern recognition. pp: 1701-1708.
4. Taigman, Yaniv, et al. 2015. "Web-scale training for face identification." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp: 2746-2754.
5. Jia, Yangqing, et al. 2014. "Caffe: Convolutional architecture for fast feature embedding." Proceedings of the 22nd ACM international conference on Multimedia. ACM, pp: 675-678.
6. Zhang K., Zhang Z., Li Z., et al. 2016. "Joint face detection and alignment using multitask cascaded convolutional networks," J. IEEE Signal Processing Letters, 23(10): 1499-1503.
7. Huang G.B., Ramesh M., Berg T., et al. 2007. "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," R. Technical Report 07-49, University of Massachusetts, Amherst.
8. Best-Rowden L., Han H., Otto C., et al.2014. "Unconstrained face recognition: Identifying a person of interest from a media collection," J. IEEE Transactions on Information Forensics and Security, 9(12): 2144-2157.
9. Sun, Yi, Xiaogang Wang, and Xiaoou Tang. 2015. "Deeply learned face representations are sparse, selective, and robust." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp: 2892-2900.
10. Parkhi O.M., Vedaldi A., Zisserman A. 2015. "Deep Face Recognition." BMVC. Vol. 1, No. 3, p. 6.
11. Wu, Xiang, Ran He, and Zhenan Sun. 2015. "A lightened CNN for deep face representation. arXiv preprint." arXiv preprint arXiv: 1511.02683, 4.