Proceedings of 2019 IEEE
International Conference on Mechatronics and Automation
August 4 - 7, Tianjin, China

Face and Gender Recognition System Based on Convolutional Neural networks

Yuxiang Zhou and Hongjun Ni
School of Mechanical Engineering
Nantong University
9 Seyuan Rd, Jiangsu 226019, China
{1307060490@qq.com& ni.hj@ntu.edu.cn}

Fuji Ren and Xin Kang
Faculty of Engineering
University of Tokushima
2-1 Minamijyousanjima-cho, Tokushima 770-8506, Japan
ren@is.tokushima-u.ac.jp

***Abstract* - In the existing research, face features and gender attributes are separated, resulting in face recognition errors and gender recognition errors in complex backgrounds. In this work, we propose the Face and Gender Recognition System that uses convolutional neural networks (CNN). The system consists of two components: one is face recognition module and two is gender recognition module. Both face recognition module and gender recognition module use pre-trained CNN to extract face and gender features in the image. Specifically, in the face recognition module, we use the public datasets Labeled Faces in the Wild (LFW), YouTube Face (YTF) and VGGFace2 to train CNN, which improves the precision. In the gender recognition module, we use the public dataset Adience to train CNN and improve the best recognition accuracy from 91.80% to 93.22%.**

Index Terms - face recognition; gender recognition; CNN

I. Introduction

In past years, CNN have achieved great results in face recognition. With advanced networks architecture and keen learning methods, CNN has improved face recognition ability to unprecedented levels [1]. Generally, face recognition includes face recognition and face validation [2-4]. Recently, efforts have been made to explore the impact of face alignment methods based on larger datasets or new loss functions on face recognition performance.

Gender plays an important role in social interaction. In society, languages retain different names and grammatical rules for men or women [5], and at the same time retain different words when dealing with different kinds of people. This is the result of social interaction, and it depends on our ability to recognize individual characteristics at a glance, namely gender, face appearance. The ability to automatically and accurately estimate their accuracy from face images still do not meet the commercial applications. Although there is a clear relationship between good face recognition and gender recognition, there are far fewer automatic gender recognition systems developed from facial images.

Currently, the accuracy of face and gender recognition at this stage needs to be improved. We think that face recognition is often tied to other tasks. For example, when presenting a partial image of a face in a complex background, we will recognize identity and gender. This prompted us to explore the learning of related tasks for face recognition.

In this study, we combine face recognition with gender recognition to allow both functions to be implemented simultaneously. In terms of neural networks, we use improved CNN to improve recognition accuracy. By continuing to use Batch Normalization to stabilize the stable propagation of the training time, the model size is reduced by introducing a Global Average Pool (GAP) instead of the fully connected layer before the final image output, thereby improving recognition efficiency and accuracy.

There are two main contributions to this research: First, we propose the Face and Gender Recognition System that uses CNN to achieve face recognition in complex contexts while achieving gender recognition. Secondly, in Face and Gender Recognition System, we have improved the accuracy on all four datasets.

The rest of this paper is organized as follows: In Section II, we briefly review recent related work. Section III details the methods of Face and Gender Recognition System. In Section IV, we prove the effectiveness of the method from experiments. Finally, in Section V, we make conclusions and prospects.

II. Related Work

A. Face Recognition

There are so many works that has been done on face recognition. Conventional methods use the way of manually extracting features from images to represent faces, such as SIFT, LBP, and HOG [6-7], and then use these features in the classification step. Recently, the deep learning method, artificial neural networks, is used to realize the feature extraction of face recognition, and the appropriate features are automatically learned by training a large amount of image data. In neural networks, CNN was first adopted in the field of face recognition. As early as 1997, Lawrence et al. [8] first applied CNN to face recognition, but the number of image datasets used was limited. In past years, by building and using of datasets with a large amount of images, the research of face recognition has also made great progress.

Cao et al. [9] established the VGGFace2 dataset, which contains 3.31 million images from 9131 subjects. The LFW dataset [10] is consisted of 13,323 web images from 5,749 celebrities, and the YTF dataset collects 3,425 videos from 1,595 themes.

978-1-7281-1699-0/19/$31.00 ©2019 IEEE

Taigman et al. [11] used the Deep-Face neural networks, which consisted of 9 layers of CNN, and trained over 4.4 million face images, which includes more than 4,000 identities. It performed well in both the LFW and YTF datasets. What is more, Parkhi et al. [12] used VGG-Face networks including 16 layers networks, and trained over 2.6 million images to achieve higher accuracy in these benchmarks. He et al. [13] proposed the ResNet50 networks. He trained the model on the ImageNet 2012 classification dataset [14] , which was evaluated on the 50k validation image. Face recognition requires high quality and comprehensive benchmark sets such as LFW and VGGFace2, and such public datasets do not apply to gender assessment [15].

In this study, we use CNN to study face recognition and face orientation invariant face recognition in complex backgrounds. The datasets include LFW, VGGFace2 and YTF.

B. Gender Recognition

For many years, gender classification has received much attention because of its application in face recognition [16], human-computer interaction [17], and soft biometrics [18]. In the early stage of gender recognition research, Golomb et al. [19] first used neural networks to achieve gender classification, but the number of experimental image samples was only 90, and the accuracy of experimental results needs to be improved. Jia et al. [20] put forward a method to generate face image classifiers to train classifiers online on 4 million weakly marked face images. Recently, Antipov [21] proposed a CNN integration model and found that CNN requires much less training data to achieve more advanced performance. Their CNN collection trained 494,414 face images from the CASIA database [22]. Jia et al. [23] collected 5 million weakly labeled face images and studied the performance differences of different depth CNN on training data.

III. Methodology

As shown in Figure 1, our proposed Face and Gender Recognition System which is composed of two main modules: face recognition module and gender recognition module. When it is given a face image, Face and Gender Recognition System will try to identity information of the subject and its gender information separately, and finally output the image identity information and gender information simultaneously.

At present, face recognition is usually composed of four phases: detection, alignment, representation and classification. Given a face representation with a complex background, the system can still predict identity and gender by a classification algorithm.

A. Preprocessing

Firstly, the AdaBoost is used to detect the face region in complex background, and then the face expression image is preprocessed. The image size is first normalized to a size of 224 x 224 pixels and a 128-dimensional embedding is generated for each image. Face detection and alignment processing are used as pre-processing to prepare for subsequent image face recognition.

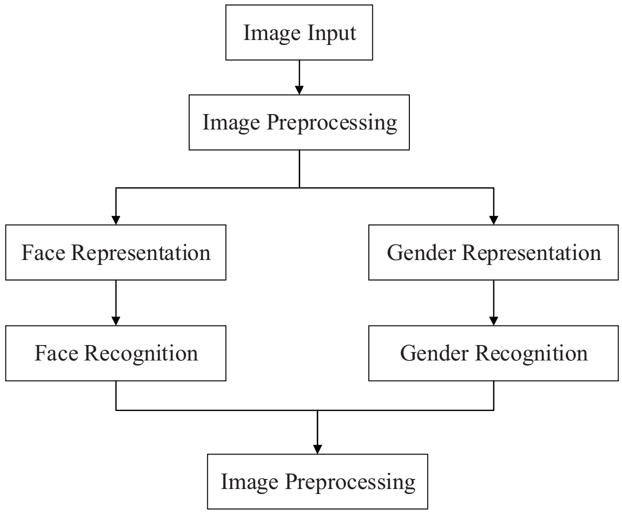

Fig. 1 Face and Gender Recognition System.

B. Representation

The methods of face feature extraction are summarized into two kinds: one is knowledge representation methods; the other is algebraic features or statistical learning.

Face and gender characterization greatly influences the ability of face and gender recognition, which is the important part of the research. In this study, we build the neural networks of our system based on the ResNet50 [14] networks to train face and gender datasets. The neural networks is shown in Figure 2. The neural networks we built is a 50-layer networks that trains over 3 million celebrity images. In the neural networks architecture, we use convolution to convolve features and use specific layer functions to get the characteristics of each layer. Since the networks structure of ResNet50 is different from other traditional convolutional networks structures, the forward propagation of signals and the back propagation of gradients become more complicated. In order to stabilize the forward propagation of the signal and the back propagation of the gradient during training, we continue to use Batch Normalization to solve this problem. Before the final image output, we use the Global Average Pool (GAP) instead of the fully connected layer to decrease the size of network model, thereby improving training speed and training accuracy of the neural networks.

IV. Experiments

A. Datasets

As shown in Table I, the LFW, YTF and VGGFace2 datasets are used in the face recognition module. The Adience dataset is used in the gender recognition module. The datasets are a collection of real-world imaging conditions that varies in appearance, noise, posture, light, and so on.

When we do networks training in the YTF dataset, we do not use video, but the picture frame that the video is clipped.

Each person has an average of 2.15 videos.

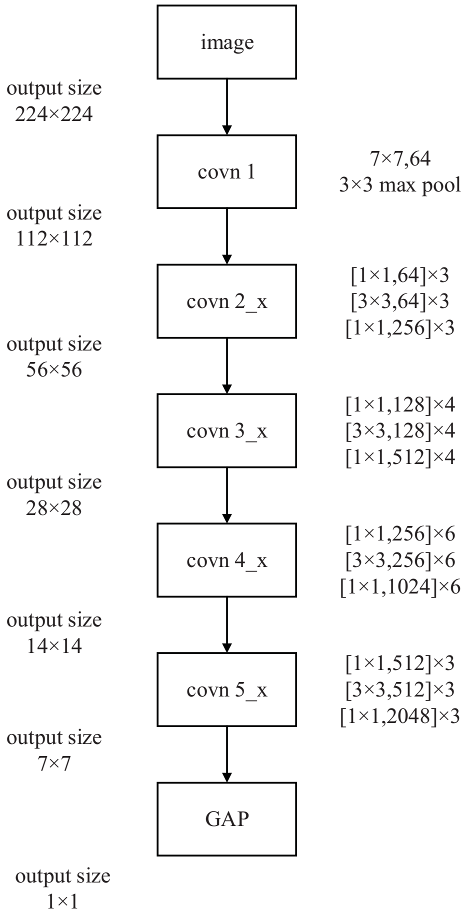

Fig. 2 Neural networks structure diagram used in this system.

TABLE I
FACE DATASETS FOR TRAINING AND TESTING IN FACE RECOGNITION MODULE AND GENDER RECOGNITION MODULE

	Datasets	Identity
Face Recognition	LFW	5,749
	YTF	1,595
	VGGFace2	9,131
Gender Recognition	Adience[15]	2,284

B. *Performance of Face Recognition Module*

We trained the face recognition module in LFW dataset, YTF dataset and VGGFace2 dataset. In LFW dataset, the experimental results are shown in Table II. From the table we find that after 200epoch training, the correct validation rate of the networks tends to be saturated, and the validation accuracy rate will not increase with the increase of epoch, but it will decrease slightly. When the networks was trained for 200epoch, the average validation accuracy was 0.9933, and the average validation loss was 0.0260.

In the YTF dataset, we can see experimental results in Table II. When the networks was trained for 200epoch, the average validation accuracy was 0.9983, and the average validation loss was 0.0089.

In the VGGFace2 dataset, we did 50epoch and 100epoch experiments. From Table II we can see experimental results. At 100epoch, the average validation accuracy was 0.9866 and the average verified loss was 0.0593.

We compare the results of previous reports with those calculated by our gender recognition system. As shown in Table III, the comparison of final results is performed in the same dataset. It can be seen that the correct rate in our work has been improved.

TABLE II
TRAINING AND VALIDATION ACCURACY, LOSS OF THE FACE RECOGNITION MODULE ON THE LFW, YTF VGGFACE2 DATASETS

Epoch	*Datasets*	*Acc*	*Loss*	*Val_acc*	*Val_loss*
50	LFW	98.35%	0.0522	91.12%	0.3226
100	LFW	99.57%	0.0176	96.43%	0.1266
200	LFW	99.76%	0.0116	99.33%	0.0260
100	YTF	99.60%	0.0164	99.51%	0.0127
200	YTF	99.84%	0.0071	99.83%	0.0089
50	VGGFace2	97.87%	0.0771	96.99%	0.0838
100	VGGFace2	98.68%	0.0582	98.66%	0.0593

TABLE III
FACE CLASSIFICATION ACCURACY RESULTS ON DIFFERENT DATASETS

Method	*Datasets*	*Accuracy*
Schroff et al.[24]	LFW	98.87%
Parkhi et al.[25]	LFW	98.95%
Ours	LFW	99.33%
Schroff et al.[24]	YTF	95.10%
Parkhi et al.[25]	YTF	97.30%
Ours	YTF	99.83%
Cao et al.	VGGFace2	98.00%
Ours	VGGFace2	98.66%

C. *Performance of Gender Recognition Module*

We put the gender recognition module into the Adience dataset for training, and we did 50epoch and 100epoch experiments. The experimental results are shown in Table IV. The experiment found that after 200 epoch training, the correct recognition rate of the networks tends to be saturated, and the recognition accuracy rate will not increase with the increase of epoch, and the correct rate is almost stable. At 100epoch, the average validation accuracy was 0.9322 and the average validation loss was 0.2781.

We compare the results of previous reports with those calculated by our gender recognition system. The results are shown in Table V. The previous reports and the experiments in our work were all conducted in the Adience dataset. It is obvious that the correct rate in our work has been greatly improved.

TABLE IV
TRAINING AND VALIDATION ACCURACY, LOSS OF THE FACE RECOGNITION MODULE ON THE ADIENCE DATASET

Epoch	*Acc*	*Loss*	*Val_acc*	*Val_loss*
50	0.9699	0.0950	0.9159	0.2945
100	0.9805	0.0672	0.9322	0.2781

TABLE V
GENDER CLASSIFICATION ACCURACY RESULTS ON THE ANDIENCE BENCKMARK

Method	*Accuracy*
Eidinger et al.	77.8%
Liao et al.[26]	78.63%
Hassner et al.[27]	79.3%
Levi et al.[28]	86.8%
Dehghan et al.[29]	91%
Gurnani et al.[30]	91.8%
Ours	93.22%

We did a test of the Face and Gender Recognition System in the VGGFace2 dataset and found some images that were misclassified by the system, as shown in Figure 3. Faces in the image are misidentified, and most of them have interference such as angle, illumination, and occlusion. The first three faces are associated with the face angle problems that cause the face recognition errors. The fourth face is with an illumination problem that causes the face and gender recognition errors. The last two faces are with the occlusion problems that cause the face recognition errors.

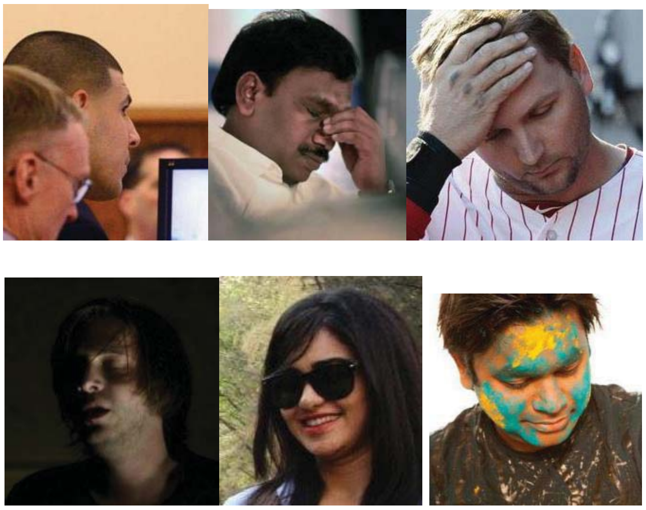

Fig. 3 Common false positives in Face and Gender Recognition System.

V. CONCLUSION AND FUTURE WORK

Our proposed Face and Gender Recognition System realizes the combination of image face recognition and gender recognition module, which enables not only face recognition but also gender recognition in complex background. Based on the ResNet50 neural networks, we use the global average pool (GAP) instead of the fully connected layer before final output, followed by the softmax layer, which reduced the size of the networks. By constructing such a simple structure, the accuracy of the system recognition has been improved.

In this study, we conduct experiments in different datasets to determine the applicability of the same neural networks to both face recognition and gender recognition modules. In the face recognition module, we achieved an increase in accuracy in all three face recognition datasets. In the gender recognition module, we compare our results with previous studies in same datasets. We achieve the best results that are ever achieved on the Adience dataset, by increasing the best achievable accuracy from 91.80% to 93.22%. Although our work has improved the accuracy of face and gender recognition, there is still much room for improvement in the recognition accuracy of gender recognition.

ACKNOWLEDGMENT

This research has been partially supported by JSPS KAKENHI Grant Number 15H01712 and A Priority Academic Program Development of Jiangsu Higher Education Institution (PAPD) and Key Research and Development Program of Jiangsu (Industry Prospects and Common Key Technologies) (BE2018093).

REFERENCE

[1] Y. Sun, D. Liang, X. Wang, and X. Tang. Deepid3: Face recognition with very deep neural networks. *arXiv preprint arXiv*:1502.00873, 2015.

[2] F. Ren, Y Q. Li, M. Hu, et al. Face recognition method based on multi features description and local fusion classification decision. *Opto-Electronic Engineering*, vol. 43, no. 9, pp. 1–8, 2016.

[3] Z. Huang, F. Ren. Face expression recognition based on multi-regional D-S evidences theory fusion. *IEEJ Trans Electr Electron Eng*, vol. 12, no. 2, pp. 251–261, 2017.

[4] F. Ren, Z. Huang. Automatic face expression learning method based on humanoid robot XIN-REN. *IEEE Trans Hum Mach Syst*, vol. 46, no. 6, pp. 810–821, 2016.

[5] G. Levi and T. Hassncer. Age and gender classification using convolutional neural networks. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 34–42, 2015.

[6] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 529–534, 2011.

[7] C. Lu and X. Tang. Surpassing human-level face verification performance on LFW with GaussianFace. *arXiv Preprint arXiv*:1404.3840, 2014.

[8] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back. Face recognition: A convolutional neural-networks approach. *IEEE Transactions on Neural networks*, vol. 8, no. 1, pp. 98–113, 1997.

[9] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. In FG, 2018.

[10] C. Lu and X. Tang. Surpassing human-level face verification performance on LFW with GaussianFace. *arXiv Preprint arXiv*:1404.3840, 2014.

[11] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1701– 1708, 2014.

[12] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. *British Machine Vision Conference (BMVC)*, pp. 41.1–41.12, September 2015.

[13] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *arXiv preprint arXiv*:1512.03385, 2015.

[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *arXiv*:1409.0575, 2014.

[15] E. Eidinger, R. Enbar, and T. Hassner. Age and gender estimation of unfiltered faces. Trans. on Inform. Forensics and Security, 9(12), 2014.

[16] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar. Attribute and simile classifiers for face verification. *ICCV*, 2009.

[17] K. Ricanek and T. Tesafaye. MORPH: A longitudinal image database of normal adult age-progression. *Conference Automatic Face and Gesture Recognition. IEEE*, pp. 341–345, 2006.

[18] M. Demirkus, D. Precup, J. J. Clark, and T. Arbel. Soft biometric trait classification from real-world face videos conditioned on head pose estimation. *Conference Comput. Vision Pattern Recognition workshops. IEEE*, pp. 130–137, 2012.

[19] S. Jia and N. Cristianini. Learning to classify gender from four million images. Pattern Recogn. Lett. vol. 58, no. C, pp. 35–41, 2015.

[20] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. Sexnet: A neural networks identifies sex from human faces. *Neural Information Processing System*, pp. 572–579, 1990.

[21] G. Antipov, S. A. Berrani, and J. L. Dugelay. Minimalistic cnnbased ensemble model for gender prediction from face images. *Pattern Recognition Letters*, 2015.

[22] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *Eprint Arxiv*, 2014.

[23] S. Jia, T. Lansdall-Welfare, and N. Cristianini. Gender classification by deep learning on millions of weakly labelled images. *IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 462–467, 2016.

[24] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. *CVPR*, 2015.

[25] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. *British Machine Vision Conference*, 2015.

[26] Z. Liao, S. Petridis, and M. Pantic. Local Deep Neural networks for Age and Gender Classification. *arXiv preprint arXiv*:1703.08497, 2017.

[27] T. Hassner, et al. Effective face frontalization in unconstrained images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[28] G. Levi, and H. Tal. Age and gender classification using convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015.

[29] A. Dehghan, et al. Dager: Deep age, gender andemotion recognition using convolutional neural networks. *arXiv preprint arXiv*:1702.04280, 2017.

[30] A. Gurnani, K. Shah, V. Gajjar, V. Mavani, and Y. Khandhediya. SAF-BAGE: Salient approach for face soft-biometric classification–age, gender, and face expression. 2018.