Multi-pose Facial Expression Recognition Based on Unpaired Images

Bairu Chen, Yibo Gan, and Bing-Kun Bao(✉)

College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China
bingkunbao@njupt.edu.cn

Abstract. Giving machines the ability to perceive human emotions and enable them to recognize our emotional states is one of the important goals to realize human-computer interaction. In the past decades, facial expression recognition (FER) has always been a research hotspot in the field of computer vision. However, the existing facial expression datasets generally have the problems of insufficient data and unbalanced categories, leading to the phenomenon of over-fitting. To solve this problem, most methods employ the generative adversarial network (GAN) for data augmentation, and achieve good results in facial image generation. But these works focus only on facial identity or head poses, which are not robust for the transformation of facial expression recognition from the laboratory environment to unconstrained scenes. Therefore, we employ the disentangled representation learning to obtain facial feature representation, so as to reduce the impact of pose changes and identity biases on FER. Specifically, the generator uses the encoder-decoder structure to map each face image to two latent spaces: the pose space and the identity space. In each latent space, we disentangle the target attribute from other attributes, and then concatenate corresponding feature vectors to generate a new image with one person's identity and another person's pose. Experimental results on Multi-PIE and RAFD datasets show that the proposed method can obtain high quality generated images and effectively improve the recognition rate of facial expressions.

Keywords: Facial expression recognition · Generative adversarial network · Disentangled representation learning

1 Introduction

Facial expression is used to convey emotional states and intentions for human beings, universally, naturally and powerfully [5]. Due to the practical importance of FER in pain assessment, driver fatigue monitoring, lie detection and many other human-computer interaction systems [15], a great deal of research has been done on facial expression recognition, aiming to classify facial emotions into seven basic expressions, such as happy, angry, sad, disgust, fear, surprise and neutral.

Y. Peng et al. (Eds.): ICIG 2021, LNCS 12889, pp. 374–385, 2021.
https://doi.org/10.1007/978-3-030-87358-5_30

Although deep learning methods have made great progress in facial expression recognition [19,20,22], there are still huge challenges in FER application. On one hand, deep neural network needs sufficient training data to avoid overfitting. However, most of the existing face datasets are taken in the laboratory environment and annotated by hand, which is time-consuming and laborious. Besides, the category and number of facial expressions are limited, leading to the modest recognition rate of facial expressions. On the other hand, facial expression features vary with individual identity, lighting, head pose and other factors, which is also quite difficult for computer recognition.

At present, most of the methods are based on generative adversarial network for data augmentation, and have achieved good results in facial image generation. However, they usually focus only on facial identity or head poses, which are not robust. For example, in Fig. 1, (a) uses a face image as input to synthesize the subject's facial images with different expressions in any pose [27], and the identity of the output image is single. For (b), although using two face images with different identities as input to realize the exchange of facial expressions among different subjects, the influence of poses on facial expressions is ignored. In view of this, we need to take both pose changes and identity biases into account to further enrich the generated face images, so as to expand the face datasets and improve the expression recognition rate. Therefore, this paper focuses on the multi-pose facial expression recognition based on unpaired images.

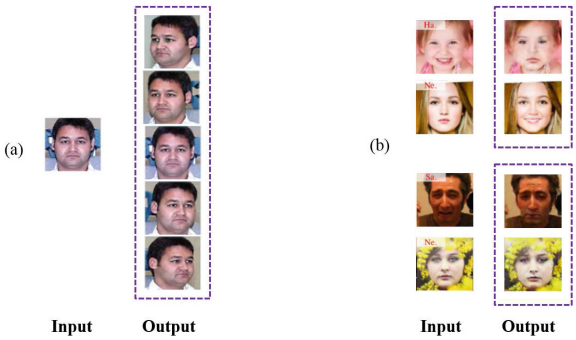

Fig. 1. Examples of face synthesis via CG-FER and GA-FER model respectively.

For multi-pose FER, traditional methods generally fall into three paradigms [27]: (1) Extract features unrelated to pose as facial expression representation. (2) Perform pose normalization for facial images. (3) Establish distinct classifiers for each specific pose. The success of these methods is largely attributed to the high quality of feature extraction. However, existing facial expression recognition methods are mostly based on manual visual features, which are susceptible to the influence of illumination and individual differences, and cannot well cope with the nonlinear facial changes caused by poses [1,6].

For identity bias, existing methods are divided into two categories: one is to learn identity invariant features [17,25,28]; the other is to minimize identity bias by learning distinct model for each specific person [4,26]. The former relys on identity-related image pairs which are difficult to obtain in the real world; the latter is also infeasible due to the lack of annotated facial images.

To address the above issues, we take both pose changes and identity biases into account, proposing the multi-pose FER based on unpaired images. In this paper, the disentangled representation learning is adopted to map each face image to two latent spaces: pose space and identity space. When inputting two images with different identities, a new facial image is generated by concatenating one person's identity vector and another person's pose vector into the decoder. In addition, a classifier is embedded in our model to facilitate image synthesis and facial expression recognition.

To sum up, this paper makes the following contributions:

(1) An end-to-end generative adversarial network is proposed to realize the pose exchange between different identity face images and facial expression recognition.
(2) The disentangled representation learning is applied to obtain facial feature representation, so as to reduce the impact of pose changes and identity biases on FER.

2 Related Work

2.1 Facial Expression Recognition

Facial expression is an important way to convey emotions in nonverbal communication. The process of FER is divided into three steps: image preprocessing, feature extraction and facial expression classification. How to extract the features of facial images effectively is the key step of facial expression recognition. The early FER is to manually extract facial expression features by designing feature extraction algorithm. It includes appearance model algorithm based on landmarks location for face modeling, as well as extraction algorithm based on local features, such as Local Binary Pattern (LBP) [31], Weber Local Descriptor (WLD) [29], multi-feature fusion, Garbor wavelet transform [23] and so on. These artificially designed extraction methods may lose part of the information of images, and are not robust enough for illumination and image scale.

With the success of deep neural network in the field of image classification and recognition, the research on FER based on deep learning has been carried out one after another. However, the training of deep convolutional neural network requires abundant data, and the lack of data will make the model unable to obtain sufficient information, resulting in the phenomenon of over-fitting. Therefore, it is extremely urgent to study the expansion of facial expression datasets. Yan et al. [24] proposed to use GAN for virtual expression images synthesis. Nirkin et al. [18] proposed the FSGAN based on RNN [2] to exchange faces. Zhang et al. [27] proposed a joint pose and expression model to generate

images with different expressions in any pose. Nowadays, improving the expression recognition rate via generating a large number of facial images has become one of the research hotspots, but these jobs either focus on pose or identity, which are not suitable for unconstrained scenes. Therefore, this paper overcomes both pose changes and identity biases to generate facial expression images with one person's identity and another person's pose.

2.2 Generative Adversarial Network (GAN)

In 2014, Ian Goodfellow [8] applied the idea of generative adversarial learning to unsupervised learning and proposed a new generative model, GAN. The network uses an unsupervised method to learn the distribution of samples, and generates highly realistic composite data, which is widely used in the field of image. GAN is a kind of neural network which is trained by game theory. In other words, GAN is trained by adversarial learning of generator and discriminator. Using this characteristic of GAN to train the existing datasets, it can generate the artificial samples which are very similar to training samples, just meeting the requirement of expanding the facial expression datasets. Therefore, it is worth studying how to use adversarial network to effectively expand facial expression datasets to alleviate the impact caused by insufficient data and unbalanced categories.

Kaneko et al. [12] added an additional filter structure to CGAN to control the generation of different facial attributes. Similarly, Choi et al. [3] proposed that StarGAN could also be used to generate different facial attributes, such as different hair colors, different ages, different expressions, etc. Although there are many work related to face synthesis, most of them are unable to decouple multiple facial attributes at the feature level, and are confined to paired images of the same person for training. Therefore, this paper proposes a multi-pose facial expression recognition based on unpaired images to decouple facial pose and identity information at the feature level.

3 Proposed Method

3.1 Facial Expression Synthesis and Recognition

We propose an end-to-end facial expression recognition model to realize pose interchange among different facial images with different identities based on generative adversarial network (GAN). The framework of the model is shown in Fig. 2. Before passing a facial image into the model, we first use the "lib face detection" algorithm with 68 points to detect the face. After preprocessing, we input the face image into the encoder to learn the identity and pose representation, and then the corresponding feature vectors are concatenated together and input into the decoder. By adversarial learning between generator and discriminator, a new image with one person's identity and another person's pose can be generated when two face images with different identities are input to the model. Then the original images and the generated images are fed into the classifier for multi-pose facial expression recognition.

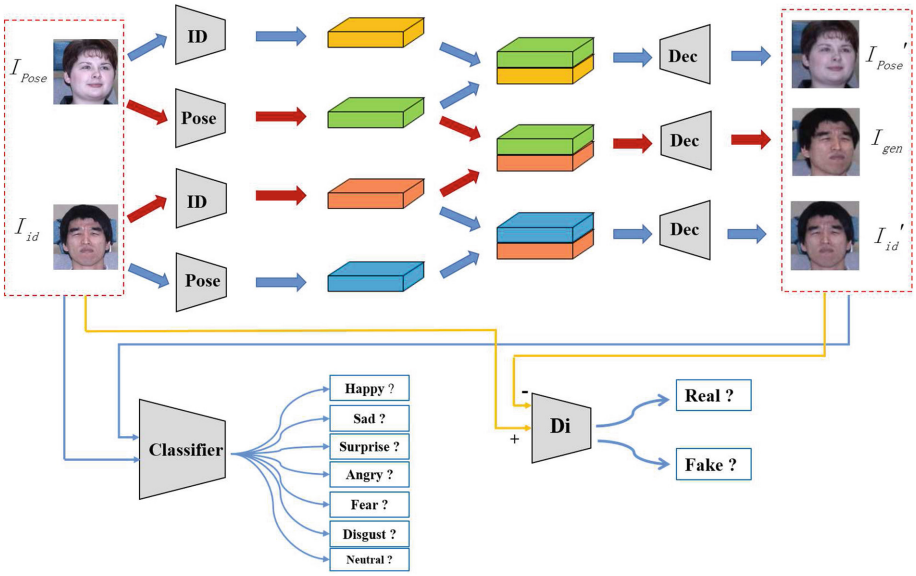

Fig. 2. The overall architecture of our facial expression synthesis and recognition model.

3.2 Network Architecture

Generator. To decouple pose and identity, we employ a two-branch generator network, processing two input streams separately. As shown in Fig. 2, the identity encoder specifically captures identity information, while the pose encoder exclusively captures pose information. Then concatenate the identity and pose features together and feed them into the decoder. In summary, the generator can be expressed as:

$$I_{gen} = G\left(E_i\left(I_{id}\right), E_p\left(I_{pose}\right)\right) \tag{1}$$

Where I_{pose} and I_{id} represent the pose reference image and identity reference image respectively. E_p is the pose encoder, which is composed of continuous Conv - Norm - ReLU blocks, while E_i is the identity encoder, consisting of consecutive Conv - ReLU blocks.

Discriminator. In order to make the model work, we also need a discriminator for adversarial training. With the size of input images larger, the receptive field of a single discriminator is limited. To solve this problem, we use multi-scale discriminators: D_1 denotes the discriminator working at a larger scale, guiding the generator to synthesize facial tiny details, while D_2 is responsible for processing the overall image content to avoid distortion of the generated images.

Classifier. For the classifier, we adopt a deep model, which ensures that the features are not affected by interference factors at each layer, and always maintains the discriminative information related to the recognition task. In this paper, VGGNet-19 network [21] is adopted as the classifier.

3.3 Loss Functions

Decoupling Loss. We use decoupled learning mechanism to control image generation based on identity and pose latent spaces. As shown in Fig. 2, given two images I_{id} and I_{pose} with different identities, the identity information of I_{id} and the pose information of I_{pose} need to be retained respectively to generate image $I_{gen} = G\left(E_i\left(I_{id}\right), E_p\left(I_{pose}\right)\right)$. In order to supervise that the image generated by the model is consistent with the input image in terms of basic facial features, we minimize the L1 distance between I_{gen} and I_{id} :

$$L_{dis} = \left\| I_{id} - G\left(E_i\left(I_{id}\right), E_p\left(I_{pose}\right)\right)\right\|_1 \tag{2}$$

Reconstruction Loss. Because there are many potential solutions to minimize decoupling loss, it alone may not achieve the desired goal. Therefore, in order to guarantee that the pose/identity encoder only encodes the pose/identity information, we require model reconstruction I_{pose} and I_{id} :

$$L_{recon} = \left\| I_{id} - G\left(E_i\left(I_{id}\right), E_p\left(I_{id}\right)\right)\right\|_1 + \left\| I_{pose} - G\left(E_i\left(I_{pose}\right), E_p\left(I_{pose}\right)\right)\right\|_1 \tag{3}$$

Classifier Loss. After synthesizing the facial images, original images and generated images are fed into the classifier for facial expression recognition. Here, we adopt a softmax cross-entropy loss for constraint:

$$L_c = -E\left[-y^e \log C\left(G\left(I\right), y^e\right) - y^e \log C\left(I, y^e\right)\right] \tag{4}$$

Adversarial Loss. In generative adversarial network, the generator is responsible for making the generated sample distribution fit the real sample as much as possible, while the discriminator is to judge whether the input sample comes from the real or the generated. In order to promote the antagonistic game between generator and discriminator, we use multi-scale discriminators. D_1 and D_2 work for different image scales, so as to improve the quality of the generated image and make them more realistic visually. Specifically, we train the generator and discriminator to optimize the following objective function:

$$L_{adv} = \sum_{i=1}^{2} \min_{G} \max_{D_i} E\left[\log D_i\left(I, y\right) + \log\left(1 - D_i\left(G\left(I, y\right), y\right)\right)\right] \tag{5}$$

Total Loss. Combined with all the above loss functions, our model forms the following min-max optimization problem:

$$\min_{G,C} \max_{D_i} \alpha L_{dis} + \beta L_{recon} + L_c + L_{adv} \tag{6}$$

Where, α and β represent the weight of each loss. The whole learning and training process is an iterative optimization of generator and discriminator.

4 Experimental Results

4.1 Datasets

The proposed multi-pose facial expression recognition model based on unpaired images is evaluated on the following two public facial expression datasets:

Fig. 3. An example of Multi-PIE images

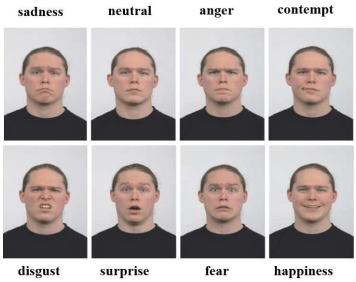

Fig. 4. An example of RAFD images

Multi-PIE. [9] The Multi-PIE is used to evaluate face recognition in a controlled environment with changes in pose, illumination, and expression. It contains images of 337 subjects from 15 different perspectives, with six different emotions: disgust, neutral, scream, smile, squint or surprise. In our experiment, we use 270 subjects in five different poses (−30, −15, 0, +15, +30). Similar to [7], we selected 6,124 face images as training data and 1,531 images as test data. Figure 3 is an example of facial expression images in Multi-PIE.

RAFD. [14] The dataset includes images of 67 subjects of different ages, genders and skin colors in eight different expressions (anger, disgust, fear, happiness, sadness, surprise, contempt and neutral). Each facial expression image corresponds to three different eye directions, and a total of 8,040 facial images are used. Figure 4 shows an example of eight basic expression images. Following the setting in [16], we selected 4,824 facial images of 8 kinds of expressions under three poses (−45, 0, +45), including 3,888 training images and 936 testing images.

4.2 Implementation Details

We construct the network according to the flowchart shown in Fig. 2. First, we employ the "lib face detection" algorithm to cut out the faces, and then resize the images as 224 × 224. The face pixel value is normalized to [−1,1]. To make the model training stable, we design the structure of generator and discriminator with reference to [30]. Specifically, the pose encoder is composed of continuous Conv - Norm - ReLU blocks, while the identity encoder consists of consecutive Conv - ReLU blocks, which are respectively responsible for decoupling the identity and pose characteristics. The decoder, on the other hand, consists of seven deconvolution layers that convert the concatenated vectors into the generated image $I_{gen} = G\left(E_i\left(I_{id}\right), E_p\left(I_{pose}\right)\right)$. For the discriminator D_i, we apply batch normalization following each convolution layer. We use Adam optimizer [13] to train our model, and the learning rate is 0.0002.

4.3 Quantitative Results

4.3.1 Experiments on the Multi-PIE Dataset

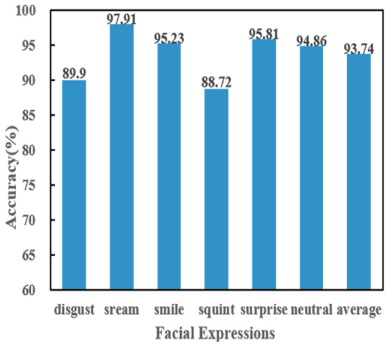

Fig. 5. Facial expression recognition results of Multi-PIE dataset

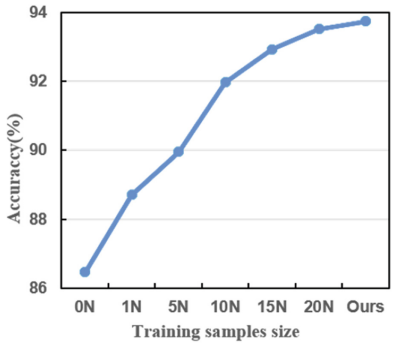

Fig. 6. Influence of the number of training samples on recognition rate

The facial expression recognition results of our proposed model on Multi-PIE is shown in Fig. 5, with an average FER accuracy of 93.74% in the last column. As can be seen from the graph, compared to other emotions, scream, smile, surprise and neutral are easier to be recognized with accuracy over 93%, which texture changes are more obvious. Among all expressions, the most difficult to be recognized is squint, with the recognition rate of only 88.72%.

Table 1 shows the comparison between our method and the current state-of-the-art methods on the Multi-PIE dataset. The average FER accuracy of each method under all poses are listed in the last column. These methods can be

divided into two categories: (1) face expression recognition based on manual features (KNN, LDA, LPP, D-GPLVM, GPLRF, GMLDA, GMLPP, MVDA, DS-GPLVM); (2) face expression recognition based on deep learning (ResNet50 [10], DesNet121 [11], CG-FER, GA-FER).

Table 1. Comparison with existing methods on Multi-PIE dataset.

Methods	Poses					Average
	−30	−15	0	+15	+30	
kNN	80.88	81.74	68.36	75.03	74.78	76.15
LDA	92.52	94.37	77.21	87.07	87.47	87.72
LPP	92.42	94.56	77.33	87.06	87.68	87.81
D-GPLVM	91.65	93.51	78.70	85.96	86.04	87.17
GPLRF	91.65	93.77	77.59	85.66	86.01	86.93
GMLDA	90.47	94.18	76.60	86.64	85.72	86.72
GMLPP	91.86	94.13	78.16	87.22	87.36	87.74
MvDA	92.49	94.22	77.51	87.10	87.89	87.84
DS-GPLVM	**93.55**	**96.96**	82.42	89.97	90.11	90.60
ResNet50	87.54	87.71	84.21	85.90	87.54	86.58
DesNet121	87.71	87.88	84.54	86.23	86.89	86.65
CG-FER	90.76	94.72	89.11	93.09	91.30	91.80
GA-FER	93.07	93.77	**92.83**	92.21	**95.13**	93.40
Ours	93.42	94.78	92.81	**93.15**	94.53	**93.74**

For the first category, the FER results of KNN, LDA, LPP, D-GPLVM, GPLRF, GMLDA, GMLPP, MVDA and DS-GPLVM are provided by [10]. The FER accuracy of our method on the Multi-PIE dataset is 93.74%, which is 17.59% $\sim$ 3.14% higher than that of the methods in [10]. For the second category, compared with the four deep learning-based methods, our model still has an improvement of 7.16% $\sim$ 0.34%. Here, GA-FER is trained for increasing the FER rate, through the synthesis of a large number of facial images under different expressions. Different from this method, ours mainly focus on the variable head poses, generating an image with one person's identity and another person's pose. Moreover, except adversarial loss and classifier loss, the proposed model also adopts L1 loss to constrain the discrepancy between generated images and original images. In general, these factors lead to an improvement in expression recognition rate with our method.

Besides, we train the classifier with different number of face images on the Multi-PIE dataset to evaluate the influence of data size on recognition rate. The overall performance of our model is shown in Fig. 6, where $\mathrm{m} \times \mathrm{N}$ (denoted as mN) means that m times generated images are selected, and then they train our model with the original images. $0 \times \mathrm{N}$ means that only the original images are

used to train the classifier. As can be seen from the graph, with the number of training samples increasing, the FER accuracy is improved, which further indicates the necessity of data augmentation, and also verifies the effectiveness of our facial expression synthesis model.

4.3.2 Experiments on the RAFD Dataset

The results of different poses and expressions on the RAFD dataset are shown in Table 2, which the last column and the bottom row respectively represent the average recognition accuracy of each pose and each expression. The bottom right 81.56% represents the average FER results of our model on RAFD. Of the eight emotions, happiness, anger, surprise and disgust are more likely to be recognized, while the most difficult to be recognized are fear and neutral. To explain this phenomenon, we look closely at face images and find that fear and neutral facial movements are relatively few compared to other expressions, making them difficult to be recognized.

Table 2. Facial expression recognition results on RAFD dataset.

	Happiness	Sadness	Anger	Surprise	Disgust	Fear	Contempt	Neutral	Average
−45	90.5	77	92	85	85	67	75.5	63	79.38
0	95	80	94.5	92	92	72.5	80.5	67	84.19
+45	92	74.5	90.5	87.5	92.5	70.5	77	64.5	81.13
Average	92.5	77.17	92.33	88.17	89.83	70	77.67	64.83	81.56

Table 3. Comparison with existing methods on RAFD dataset.

Methods	Poses		Average
	Angle	Number	
sLDA	(−45, 0, +45)	3	63.3
TDP	(−45, 0, +45)	3	63.7
Multi-SVM	(−45, 0, +45)	3	66.13
TDP-Zhang	(−45, 0, +45)	3	75
VGG16	(−45, 0, +45)	3	72.82
VGG19	(−45, 0, +45)	3	73.5
ResNet50	(−45, 0, +45)	3	74.75
DenseNet121	(−45, 0, +45)	3	74.2
Ours	(−45, 0, +45)	3	**81.56**

In the experiment, we compare our model with existing methods on RAFD dataset, including SLDA, TDP, Multi-SVM, TDP-Zhang (related results are provided in [16]) and VGG16 [21], VGG19 [21], ResNet50 [10] and DenseNet121

[11]. As can be seen from Table 3, our model has a great improvement compared with the methods based on manual features (such as sLDA and TDP), which proves that the features extracted by deep learning methods can achieve better results when dealing with nonlinear facial changes. At the same time, for classic deep learning models such as VGG16 and VGG19, our method still has an improvement of 8.74% $\sim$ 6.81%, fully indicating the synthesis of effective facial expression images via generative adversarial network plays an important role in promoting the FER task.

5 Conclusion

Based on generative adversarial network, we present an end-to-end model for face image synthesis and expression recognition simultaneously. Through the disentangled representation learning, we extract the facial identity information and pose, so as to overcome the challenges brought by pose changes and identity biases. Experiments on Multi-PIE and RAFD demonstrate the effectiveness of the proposed model. In the future, we will further consider the impact of illumination and occlusion on expression recognition, and consider how to expand the seven basic expressions to more complex and varied expressions, so that the study of expression recognition will be closer to real-world scenes.

Acknowledgment. This work was supported by the National Key Research & Development Plan of China 2020AAA0106200, the National Natural Science Foundation of China under Grant 61936005, 61872424, the Natural Science Foundation of Jiangsu Province (Grants No BK20200037).

References

1. Bengio, Y., Courville, A.C., Vincent, P.: Unsupervised feature learning and deep learning: a review and new perspectives. CoRR, p. 2012 (2012)
2. Berglund, M., Raiko, T., Honkala, M., Kärkkäinen, L., Vetek, A., Karhunen, J.T.: Bidirectional recurrent neural networks as generative models. In: NIPS, pp. 856–864 (2015)
3. Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: StarGAN: unified generative adversarial networks for multi-domain image-to-image translation. In: CVPR, pp. 8789–8797 (2018)
4. Chu, W.S., De la Torre, F., Cohn, J.F.: Selective transfer machine for personalized facial expression analysis. TPAMI **39**(3), 529–545 (2016)
5. Darwin, C., Prodger, P.: The Expression of the Emotions in Man and Animals. Oxford University Press, Oxford (1998)
6. Ding, C., Tao, D.: A comprehensive survey on pose-invariant face recognition. TIST **7**(3), 1–42 (2016)
7. Eleftheriadis, S., Rudovic, O., Pantic, M.: Discriminative shared Gaussian processes for multiview and view-invariant facial expression recognition. TIP **24**(1), 189–204 (2014)
8. Goodfellow, I.J., et al.: Generative adversarial networks. arXiv preprint: 1406.2661 (2014)

9. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. Image Vis. Comput. **28**(5), 807–813 (2010)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
11. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: CVPR, pp. 4700–4708 (2017)
12. Kaneko, T., Hiramatsu, K., Kashino, K.: Generative attribute controller with conditional filtered generative adversarial networks. In: CVPR, pp. 6089–6098 (2017)
13. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint: 1412.6980 (2014)
14. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H., Hawk, S.T., Van Knippenberg, A.: Presentation and validation of the Radboud faces database. Cogn. Emotion **24**(8), 1377–1388 (2010)
15. Li, S., Deng, W.: Deep facial expression recognition: a survey. In: IEEE Transactions on Affective Computing. https://doi.org/10.1109/TAFFC.2020.2981446
16. Mao, Q., Zhang, F., Wang, L., Luo, S., Dong, M.: Cascaded multi-level transformed Dirichlet process for multi-pose facial expression recognition. Comput. J. **61**(11), 1605–1619 (2018)
17. Meng, Z., Liu, P., Cai, J., Han, S., Tong, Y.: Identity-aware convolutional neural network for facial expression recognition. In: FG, pp. 558–565. IEEE (2017)
18. Nirkin, Y., Keller, Y., Hassner, T.: FSGAN: subject agnostic face swapping and reenactment. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7184–7193 (2019)
19. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition (2015)
20. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: CVPR, pp. 815–823 (2015)
21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint: 1409.1556 (2014)
22. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: CVPR, pp. 1701–1708 (2014)
23. Tian, Y.l., Kanade, T., Cohn, J.F.: Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. In: FG, pp. 229–234. IEEE (2002)
24. Yan, Y., Huang, Y., Chen, S., Shen, C., Wang, H.: Joint deep learning of facial expression synthesis and recognition. TMM **22**(11), 2792–2807 (2019)
25. Yang, H., Ciftci, U., Yin, L.: Facial expression recognition by de-expression residue learning. In: CVPR, pp. 2168–2177 (2018)
26. Yang, H., Zhang, Z., Yin, L.: Identity-adaptive facial expression recognition through expression regeneration using conditional generative adversarial networks. In: FG, pp. 294–301. IEEE (2018)
27. Zhang, F., Zhang, T., Mao, Q., Xu, C.: Joint pose and expression modeling for facial expression recognition. In: CVPR, pp. 3359–3368 (2018)
28. Zhang, K., Huang, Y., Du, Y., Wang, L.: Facial expression recognition based on deep evolutional spatial-temporal networks. TIP **26**(9), 4193–4203 (2017)
29. Zhang, Z., Wang, L., Zhu, Q., Chen, S.K., Chen, Y.: Pose-invariant face recognition using facial landmarks and weber local descriptor. Knowl. Based Syst. **84**, 78–88 (2015)
30. Zhang, Z., Song, Y., Qi, H.: Age progression/regression by conditional adversarial autoencoder. In: CVPR, pp. 5810–5818 (2017)
31. Zhong, L., Liu, Q., Yang, P., Liu, B., Huang, J., Metaxas, D.N.: Learning active facial patches for expression analysis. In: CVPR, pp. 2562–2569 (2012)