ERGAN: High Perform GAN for Eyeglasses Removal

1st Mengyuan Cheng
Xi'an University of Posts and Telecommunications
Xi'an University of Posts and Telecommunications
Xi'an China
17868812211@163.com

2nd Xiaopeng Cao
Xi'an University of Posts and Telecommunications
Xi'an University of Posts and Telecommunications
Xi'an China
cxp2760@163.com

2021 16th International Conference on Intelligent Systems and Knowledge Engineering (ISKE) | 978-1-6654-0553-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISKE54062.2021.9755402

***Abstract*—To solve the occlusion of eyeglasses on the eye area, which brings great challenges to face recognition and gaze tracking. Based on the successful application of face attribute editing task in Generative Adversarial Network, this paper proposes a new eyeglasses removal method ERGAN. The latent space of GANs has rich semantic information. By manipulating latent codes, different attributes of images can be changed. The key is to find the appropriate semantic direction. First of all, in order to enable GAN to process real images, we introduce the more advanced GAN inversion model IDInvert, which can convert the given real image back to the latent space of the pre-trained GAN model. Then use our model ERGAN to remove eyeglasses attributes. Subsequently, to find the appropriate semantic direction of eyeglasses editing, we add the semantic direction of the eyeglasses instance to the semantic direction calculated by InterfaceGAN. Recombine the two semantic directions into the semantic direction of eyeglasses editing. Finally, experiments are performed on GANs generated images and real images. The experimental results show that our method is helpful to improve the accuracy of eyeglasses removal.**

Keywords—Generative Adversarial Networks, facial attribute editing, GAN inversion

I. Introduction

Line of sight tracking is one of the main research directions in the field of computer vision. Although much important progress has been made. But sight tracking still faces many challenges. These challenges come mainly from the occlusion of eyeglasses in face images. As the most important auxiliary tool in most people's study and life, eyeglasses have brought many conveniences to people. However, the occlusion of the eye area by eyeglasses poses great challenges to face recognition and line of sight tracking. Therefore, this paper takes the task of face image eyeglasses extraction [1] as the main research object.

Early eyeglasses removal methods for face images mainly use PCA for reconstruction, but the traditional eyeglasses removal methods are easy to cause obvious processing traces or image distortion. In 2014, Goodfellow et al. proposed Generative Adversarial Nets (GAN) [2]. Generating confrontation network (GAN)is a deep learning architecture for estimating how data points are generated in a probabilistic framework, which mainly includes two interacting neural networks: generator G and discriminant D. The goal of G is to generate false data that can be faked. The goal of D is to identify real data and false data, and the two are jointly trained in a confrontational process.

In recent years, with the development of generative adversarial networks, more and more people try to use GAN to process face images. GAN can generate high resolution and realistic 'fake' images and has rich semantic information in the latent space of intermediate features, which can change different attributes of images by manipulating latent codes. However, due to the lack of inference ability or encoder of GANs, manipulating the latent code is only applicable to the images generated by GANs. To apply GAN to real image editing, GAN inversion is proposed. GAN inverse mapping aims to convert a given image back to the hidden space of the pre-trained GAN model, and then the image can be faithfully reconstructed by the reverse coding of the generator. Many researchers have also proposed many new methods for face attribute editing. The purpose of the face attribute editing task is to change a given face image to its certain attributes, such as eyeglasses, age, gender, and expression. A successful face attribute editing not only requires accurate, high-quality output for the target attribute but also ensures that other attributes are not changed.

Inspired by the face attribute editing task, this paper will propose a framework ERGAN (High Perform GAN for Eyeglasses Removal) to remove the eyeglasses attribute based on the face attribute editing algorithm, aiming at the eyeglasses attribute in the face image, and ensuring the integrity of other face attribute features.

II. Related Works

In 2004, Wu et al. [3] removed glasses by statistical mapping with or without glasses, but difficult to remove frameless glasses. In 2005, Du et al. [4] proposed an spectacle removal algorithm for artificially synthesized face images with glasses. The adaptive binarization was used to detect the spectacle part, and then the recursive error compensation of PCA reconstruction was used to remove the spectacle. However, this method is sensitive to the change of image illumination. In 2006, Chen Xuerong et al. [5] extracted the feature vector in the face image to compensate, and then removed the glasses by classification and recognition. However, this method is only for infrared face samples and has no effect on ordinary color face images. In 2014, Li Gen et al. [6] proposed a method based on the combination of machine learning and local features of mind evolution to remove face occlusion and realize face recognition under the condition of occluded face, but its verification data set is small and does not have fine-grained features. In 2014, Mirza et al. [8] proposed a conditional generative adversarial network (CGAN), adding class information as a hidden vector, and considering the class c information when the generator and discriminator input. In 2015, Wu et al. [9] proposed a de-occlusion method based on sparse representation classification, and reconstructed the occluded

This work was supported by the Key Research and Development Program of Shaanxi Province(2021GY-181).

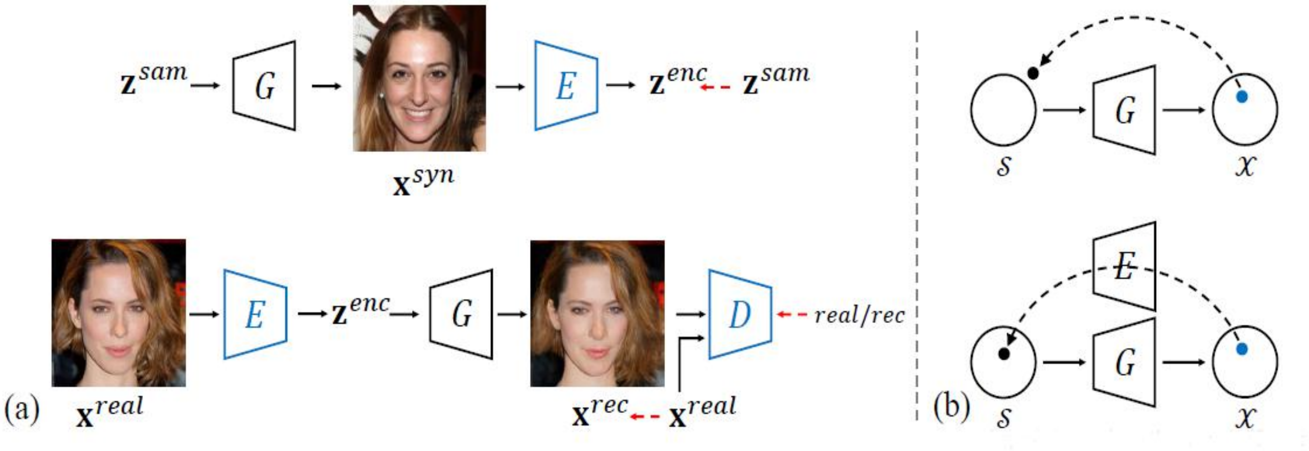

Fig. 1. Encoder Training in IDInvert

face image through the obtained sparse coefficient, but this method relied too much on the priori knowledge of glasses. In 2017, Zhang Xin and others [10] proposed a fan-shaped linear interpolation compensation method based on minimum distance search to remove glasses. The threshold segmentation based on semi-threshold and minimum variance filtering was used to determine whether there are glasses, but the processing of dark box glasses will have obvious processing traces. In 2017, Odena et al. [11] proposed AC-GAN, which uses auxiliary classifiers to maximize the logarithmic likelihood of the generated categories, so that the discriminant of the network can not only identify the true and false of the generated samples but also identify the categories of the generated samples. In 2017, Shen and Liu [12] proposed a generative adversarial network consisting of two image conversion networks and a discriminant network to modify a given face attribute by learning the corresponding residual images. In 2017, Zhu et al. [13] proposed cycleGAN. The network consists of two generators and two discriminators, forming a loop. In the case of inconsistent data training, it realizes style transfer, object deformation, seasonal change, etc. In 2018, Cao et al. [14] proposed an improved face restoration algorithm based on a generative adversarial network, adding a restoration layer to the network, so that the test image can generate the corresponding high-confidence image, which greatly reduces the network convergence time. In 2018, Choi et al. [15] proposed a unified generation confrontation network StarGAN for the multi-domain image to image conversion, which can perform image to image conversion for multiple domains only using a single model. In 2018, Zhang et al. [16] introduced the attention mechanism into the GAN framework and proposed SaGAN to ensure that only specific attribute regions are changed and other regions remain unchanged. In 2019, Karras et al. [17] proposed Style-GAN, which is based on style transfer. By modifying the input of each level, the visual characteristics expressed by the level are controlled without affecting other levels. In 2019, Lin et al. [18] proposed a new method RelGAN for the multi-domain image to image conversion, which mainly uses relative attributes and has good performance in facial attribute transfer and interpolation tasks. In 2019, He et al. [19] proposed a new face editing method DeepFaceEditing based on geometric and appearance decoupling, which can freely edit faces through sketches. In 2020, Bahng et al. [20] proposed an unpaired and unmarked multi-domain image-to-image conversion method for original unmarked raw data in order to reduce the necessity of labeled data in the field of face image translation. In 2020, Shen et al. [21] proposed InterfanceGAN to train the SVM classifier in order to edit the face attributes of the pre-trained GAN model. They separated the potential space into opposite semantic labels and output the normal vector of the support vector machine boundary as the semantic direction. In 2020, Voynov et al. [22] first proposed the use of unsupervised methods to effectively discover the effective interpretable moving direction of potential coding without the expensive cost. In 2020, Härkönen et al. [23] extended the current encoder-based inversion method by introducing an iterative refinement mechanism and proposed a residual-based encoder ReStyle. In 2021, Shen and Zhou Ti [24] developed a new algorithm to discover potential semantic directions, called SeFa, independent of any new type of training and sampling, and applicable to the most popular GAN models (e.g. PGGAN, StyleGAN, BigGAN, StyleGAN2).In 2021, Mao et al. [25] proposed an ERCNN network model for the removal of eyeglasses for fine-grained face recognition and achieved end-to-end eyeglasses removal by learning the mapping relationship between face images with and without eyeglasses. In 2021, Han et al. [26] proposed a new framework IALS to solve the problem of poor attribute decoupling, which dynamically searches the semantic direction of instance perception in GAN hidden space to promote the decoupling of attribute changes.

III. Method

This chapter introduces the new eyeglasses removal framework ERGAN proposed in this paper. Firstly, to make the GAN model be applied to real image editing, we introduce the GAN inversion mechanism and then edit the eyeglasses attributes of face images for the GAN-inverted graphics.

A. *GAN Inversion*

Due to the lack of inference ability or encoder of GANs, manipulating the latent code is only applicable to the images generated by GANs. So how to apply GAN to real image processing? One solution is to use GAN inversion. GAN inverse mapping aims to convert a given image back to the hidden space of the pre-trained GAN model, and then the image can be faithfully reconstructed by the reverse coding of the generator. The GAN inverse mapping enables the

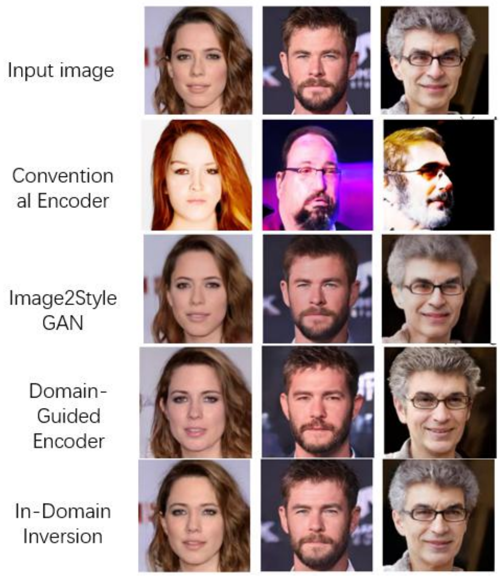

Fig. 2. IDInvert Experimental Results

controllable direction found in the existing hidden space to apply to real image editing without point-to-point supervision or high-cost optimization. After a real image is introduced into the hidden space, we can change its encoding in a certain direction to modify the image attribute of the response. With the rapid development of GAN and interpretable machine learning technology, GAN inverse mapping not only provides an alternative flexible image editing framework but also helps us understand the internal mechanism of the deep generation model.

The basic idea of GAN Inversion is divided into two steps. First, an encoder is trained with a dataset for rough calculation. Then use an optimization process to optimize. The existing method is not only the encoder is only the optimization process, but also the combination of the two. This paper cites the new GAN inversion framework IDInvert proposed by Zhu et al. [27].

The data set used in traditional encoder training is ' fake ' images generated by GANs model, such as image2stylegan, but IDInvert uses real face images to train encoder. This framework uses Flicker-Face-HQ (FFHQ) high-quality face image dataset [2] to train the encoder. Cut face images of 1024 × 1024 pixels in FFHQ dataset into 256 × 256 pixels. The encoder training model for IDInvert is shown in Figure 1.

IDInvert uses styleGAN, which is the same as the addition of latent code in image2stylegan. The encoder generates W+ of 18 * 512, which is added to the synthesis network of styleGAN layer by layer through AdaIN. MLP was not used in the experiment, only a synthesis network was used and it needed to be fixed.

The model compared with the advanced models such as image2stylegan, the effect of the model was the best. The experimental results are shown in Fig. 2. It can be seen from the results that the model can well retain some details of the real image, such as hair and wrinkles on the face, which can provide guarantee for the subsequent generation of high-quality face images in this paper.

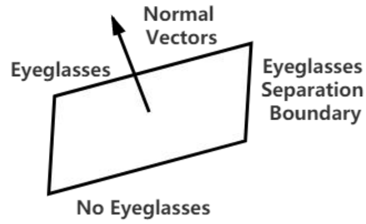

Fig. 3. Semantic Direction of Eyeglasses Attribute in InterfaceGAN

B. *InterfaceGAN Semantic Direction*

To achieve this goal, current methods required carefully designed loss functions [9,29,30] introduction of additional attribute labels or features [31,32] or special architectures [33] to train new models. However, the synthesis resolution and quality of these models are far behind those of native GANs, like PGGAN [34] and StyleGAN [17]. Different from previous learning-based methods, the InterfaceGAN explores the interpretable semantics inside the latent space of fixed GAN models, and turns unconstrained GANs to controllable GANs by varying the latent code.

Given a well-trained most advanced GAN model, the generator of the model can be expressed as a certain function :

$$g : Z \to X \tag{1}$$

Here, $Z \subseteq R^d$ denotes the d-dimensional latent space, for which Gaussian distribution $N(0, I_d)$ commonly used. X represents the face image space, each sample x has a piece of certain semantic information, such as the eyeglasses attribute g processed in this paper. The generator maps the potential vector Z to the face image and then realizes the face attribute editing by moving the potential vector Z along a certain direction in the potential space. The key to this step is to find the right semantic direction for attribute editing d_x.Correspondingly, the key of this study is to find out the appropriate semantic direction of eyeglasses.

The InterfaceGAN [21] method is used to find the semantic direction of eyeglasses attribute editing. The model is based on the assumption that for any binary attribute, there is a hyperplane in the potential space so that all samples from the same side have the same attribute. The InterfaceGAN method is mainly divided into the following three steps to find the semantic direction. Firstly, random sampling from the generated huge image corpus. Secondly, it labels the attributes of these images using a set of CNN binary classifiers $H(\cdot)$ [28]. Finally, the generated samples are used to train an SVM to separate each attribute label in the potential space of GAN, and the normal vector output of the SVM boundary is taken as the semantic direction of each attribute. As shown in Figure 3, the plane is a boundary with or without eyeglasses, and its normal vector is the semantic direction of the eyeglasses attribute.

C. *Eyeglasses-Instance Semantic Direction*

When editing the eyeglasses' attributes, it is found that when the eyeglasses ' attributes change, other people's face attributes, such as age, will also change. This is because there is a deviation in the data distribution in the training data set (most of the elderly in the FFHQ data set

TABLE I. AUC CHANGING WITH PARAMETERS

α	0	0.2	0.4	0.6	0.8	1.0
AUC	0.7956	0.8298	0.8765	0.9033	0.8931	0.8631

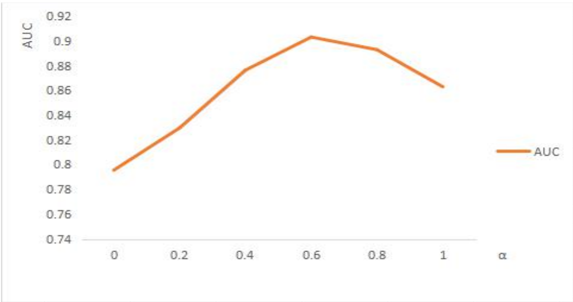

Fig. 4. AUC chart

are equipped with eyeglasses). Therefore, in order to ensure that AUC changing with parameters other attributes of the face image remain unchanged after removing the eyeglasses, it is necessary to decouple the attribute editing. Therefore, direction normalization is introduced to ensure that one or more conditional attributes b are retained when editing the original attribute g :

$$d_{g|b} = d_g - < d_g, d_b > d_b \quad (2)$$

In this paper, the semantic direction for eyeglasses is added to the InterfaceGAN method. That is, the semantic direction of this paper is redefined as the combination of InterfaceGAN semantic direction and eyeglasses classifier semantic direction :

$$d'_{X_g}(z) = \alpha d_{X_g} + (1-\alpha) d_{X_g}(z) \quad (3)$$

Here,α ∈ [0,1]Change (2) to:

$$d'_{g|b}(z) = d'_g(z) - < d'_g(z), d'_b(z) > d'_b(z) \quad (4)$$

To make the generator G map the hidden vector z to the image, the attribute classifier H maps the x ∈ X in the image to an attribute label. To connect the hidden space and attribute space of GAN, we combine the generator G and classifier H, H(G(·)).For attribute X_g, we can search for the instance-specific semantic direction for instance z, denoted by $d_{X_g}(Z)$, via minimizing the following loss:

$$\begin{aligned}&\arg\min_{d_{X_g}(Z)} D(H(G(z + d_{X_g}(z))), y) \\ &= \arg\min_{d_{X_g}(Z)} (-y \log(H(G(z + d_{X_g}(z)))) \\ &- (1-y)\log(1-(H(G(z + d_{X_g}(z)))))\end{aligned} \quad (5)$$

where y is the target attribute label and L(·,·) is the classification loss with binary cross-entropy function.

We simply use gradient descent to search for $d_{X_g}(Z)$ in:

$$\begin{aligned}d_{X_g}(z) &= \frac{-\nabla_z D(H(G(z)), y)}{\|\nabla_z D(H(G(z)), y\|_2} \\ &= (2y-1)\frac{\nabla_z H(G(z))}{\|\nabla_z L(H(G(z))\|_2}\end{aligned} \quad (6)$$

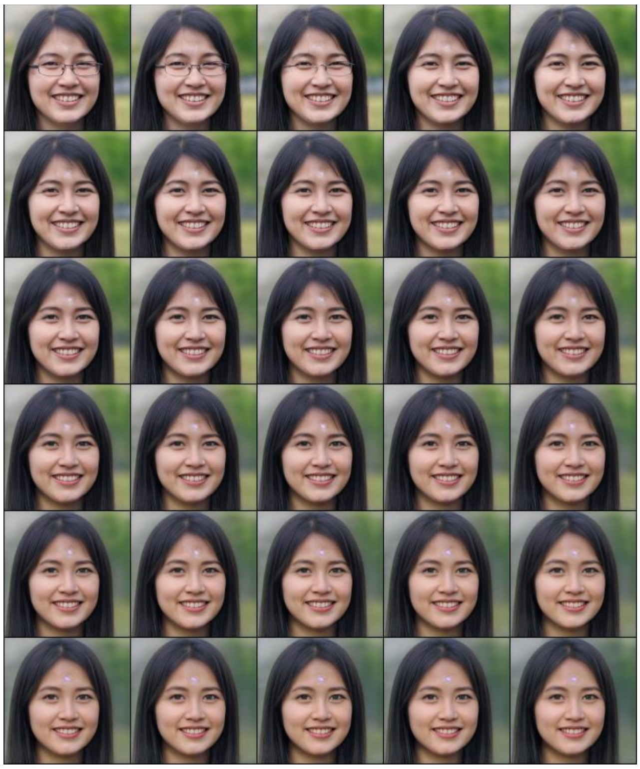

Fig. 5. 30 iterations of face image

The determination of coefficients α will be outlined in the experimental environment.

IV. Experiments

This chapter introduces the experimental environment of the ERGAN model proposed in this paper, the solution of coefficients in the semantic direction formula of eyeglasses attributes, and the experimental results on GAN-generated images and real images.

A. Experimental Environment

- Python 3.7 and the basic Anaconda3 environment.
- PyTorch 1.x with GPU support (a single NVIDIA GTX 1060 is enough).
- The tqdm library to visualize the progress bar.

B. Model Training

For generator G, we select the advanced GAN model StyleGAN, then train StyleGAN on FFHQ face dataset and test our model on its w-space. FFHQ dataset is a high-quality face dataset with 70000 high-definition face images in PNG format with 1024 × 1024 pixels. It has different race, age, expression, personality, hairstyle, skin color, face type, face posture, etc. 70000 high-definition face images in PNG format are aligned and cut into 256 × 256 pixels. For attribute classifier H, it is the ResNet-18 [28] network trained on the CelebA-HQ [17] dataset.

C. Parameter Identification

To determine the value of α ∈ [0,1] in (3),we use the DT proposed by et al. [26] to test different values. The x axis in the DT curve represents the conversion accuracy, and the y axis represents the accuracy of attribute decoupling. Therefore, the larger the area under the DT curve, the higher

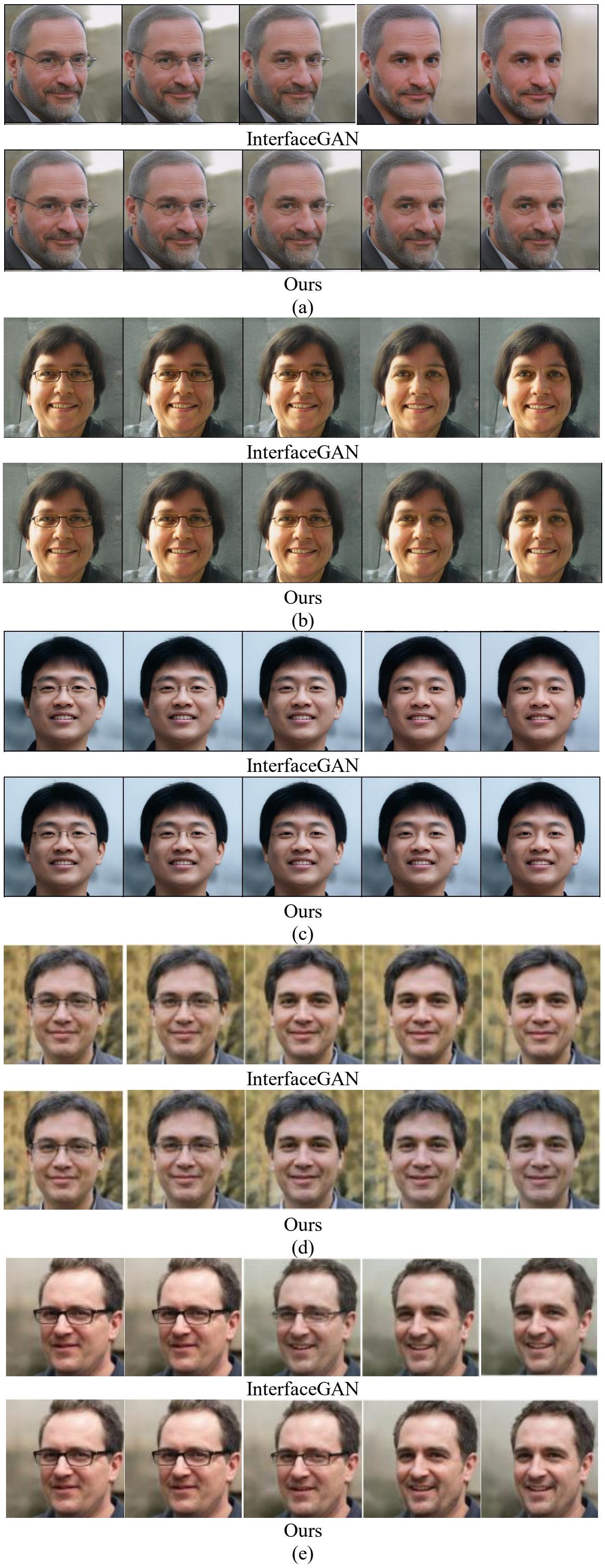

Fig. 6. Experimental results of comparison with InterfaceGAN on GAN-generated images

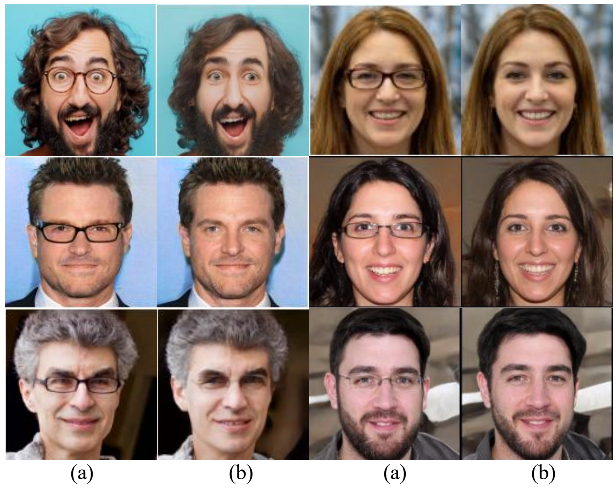

Fig. 7. Experimental results of real images

the conversion accuracy and decoupling accuracy of the converted image. The corresponding step size is set to 0.2 to test the effect to select the value. The average area under curve (AUC) under different DT curves is shown in Table 1 and Figure 4.

The image shows that when $\alpha = 0.6$, the image editing accuracy is the best, so we determine it.

Determine the step of iterations. Eyeglasses attribute editing in face images includes two tasks. One is the eyeglasses attribute is completely removed, and the other is other attributes are retained intact. In general, the higher the step of iterations, the better removal of spectacle attributes in face images. But other attributes will have some changes. The complete processing results of 30 iterations are shown in the Figure 5. We can find that when the iteration step is 4, the eyeglasses attribute is completely removed and other attributes remain intact. When the step of iterations is 6, we find that other attributes (such as eye size) have changed with the increase of iteration step. Consequently, it is very important to select the appropriate step of iterations. After many experiments, we found that when the step of iterations is 10, the eyeglasses attributes in the face image could be completely removed, and other attributes remained intact.

D. *Experimental Result*

Test our model using images generated by the GAN model. The images are generated by StyleGAN [17]. The experimental comparison results with InterfaceGAN are shown in Figure 6 and the first image is the original image.

The method is also tested on real images. Firstly, the IDInvert model introduced is used to map the real image to the potential space, and then the eye attribute is removed by inputting the ERGAN model proposed in this paper. The experimental results are shown in the Figure 7. Here, (a) are the original images, and (b) are the processed images.

It can be seen from the graph that our method is superior to the InterfaceGAN method for the image generated by GAN. It can preserve other properties in the image except for the eyeglasses. The InterfaceGAN method makes the person in the image look younger and loses some original attributes. The results of real image processing are not good for StyleGAN, but the retention of image details is still improved.

V. CONCLUSION

Aiming at the eyeglasses removal task of the face image, this paper proposes a new framework ERGAN. When looking for the semantic direction of eyeglasses attributes, the semantic direction of the eyeglasses instance is added to the semantic direction of the InterfaceGAN method to form a new semantic direction. In order to edit the real image, the IDInvert model is also introduced. It can be seen from the experimental results that this method is not complete enough for some details, and the processing effect of real images needs to be improved.

REFERENCES

[1] Y. Saito, Y. Kenmochi, K. Kotani. Estimation of Eyeglasses Facial Images Using Principal Component Analysis[C]. International Conference on Image Processing. Piscataway, NJ: IEEE 1999.

[2] Ian Goodfellow, Jean PougetAbadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In NeurIPS, 2014. 1, 2.

[3] Chenyu Wu, Ce Liu, Heungyueng Shum, Yingqing Xu, Zhengyou Zhang. Automatic eyeglasses removal from face images.[J]. IEEE transactions on pattern analysis and machine intelligence,2004,26(3):

[4] Cheng Du and Guangda Su. Eyeglasses removal from facialimages[J]. Pattern Recognition Letters, 2005, 26(14):2215–2220.doi: 10.1016/j.patrec.2005.04.002.

[5] Xuerong Chen, Zhongliang Jing, Shaoyuan Sun. A method to deal with eyeglasses interference in infrared face recognition[J]. Journal of Shanghai Jiaotong University, 2006, 40(3):113-116.

[6] Gen Li and Wenhui Li. Face occlusion recognition based onMEBML[J]. Journal of Jilin University: Engineering andTechnology Edition, 2014, 44(5): 1410–1416.

[7] Xiaodong Jia and Jiangling Guo. Eyeglasses removal fromfacial image based on phase congruency[C]. Proceedings ofthe 3rd IEEE International Congress on Image and SignalProcessing, Yantai, China, 2010: 1859–1862, doi:10.1109/CISP.2010.5647366.

[8] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.

[9] Congzhong Wu, Qufen Liu, and ZHAN Shu. A faceocclusion removal method based on sparse representation[J]. Journal of Hefei University of Technology, 2015, 38(5):615–617, 680.

[10] X. Zhang, Y. U. Hong, M. A. Xin-Li, et al. Fan Iinear Interpolation Compensation Method Based on Minimun Distance Search to Remove Glasses[J]. Journal of Chinese Computer Systems, 2017.

[11] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In ICML, pages 2642–2651, 2017.

[12] Wei Shen and Rujie Liu. Learnin gresidual images for face attribute manipulation. In CVPR, pages 1225–1233, 2017.FERNÁNDEZ A, GARCÍA R, USAMENTIAGA R, et al. Glasses detection on real images based on robustalignment[J]. Machine Vision and Applications, 2015, 26(4):519–531. doi: 10.1007/s00138-015-0674-1.

[13] Junyan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In ICCV,pages 2242–2251, 2017.

[14] Zhiyi Cao, Shaozhang Niu, Jiwei Zhang. Research onface reduction algorithm based on generative adversarialnets with semi-supervised learning[J]. Journal of Electronics & Information Technology, 2018, 40(2): 323–330. doi:10.11999/JEIT170357.

[15] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multidomain image-to-image translation. In CVPR, pages 8789–8797, 2018.

[16] Gang Zhang, Meina Kan, Shiguang Shan, and Xilin Chen. Generative adversarial network with spatial attention for face attribute editing. In ECCV, pages 422–437, 2018.

[17] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. CVPR, pages 4401–4410, 2019. 2, 3, 4, 6, 7, 8.

[18] Yujing Lin, Powei Wu, Chehan Chang, Edward Chang, and Shih-Wei Liao. Relgan: Multidomain image-to-image translation via relative attributes.In ICCV, pages 5914–5922, 2019.

[19] Zhenliang He, Wangmeng Zuo, Meina Kan, Shiguang Shan, and Xilin Chen. Attgan: Facial attribute editing by only changing what you want. IEEE Transactions on Image Processing, 28(11):5464–5478, 2019.

[20] Hyojin Bahng, Sunghyo Chung, Seungjoo Yoo, and Jaegul Choo. Exploring unlabeled facesfor novel attribute discovery. In CVPR, pages 5821–5830, 2020.

[21] Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the latent space of gans for semantic face editing. In CVPR, pages 9243–9252, 2020.

[22] Andrey Voynov and Artem Babenko. Unsupervised discovery of interpretable directions in the gan latent space. In ICML, 2020.

[23] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. Ganspace: Discovering interpretable gan controls. In NIPS, 2020.

[24] Yujun Shen and Bolei Zhou. Closedform factorization of latent semantics in gans. arXiv preprint arXiv:2007.06600, 2020.

[25] Liang Mao, Yueju Xue, Yinhui Wei, and Tingting Zhu. An eyeglass removal method for fine-grained face recognition[J]. Journal of Electronics & Information Technology,2021,43(05):1448-1456.

[26] Y Han, Yang J, Y Fu. Disentangled Face Attribute Editing via Instance Aware Latent Space Search. 2021.

[27] J. Zhu, Y. Shen, D. Zhao, and B. Zhou, "In-domain gan inversion for real image editing," in ECCV, 2020.

[28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, pages 770–778, 2016.

[29] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In NeurIPS, 2016. 2.

[30] Luan Tran, Xi Yin, and Xiaoming Liu. Disentangled representation learning gan for pose-invariant face recognition. In CVPR, 2017. 2.

[31] Guillaume Lample, Neil Zeghidour, Nicolas Usunier, Antoine Bordes, Ludovic Denoyer, and Marc ′ Aurelio Ranzato. Fader networks: Manipulating images by sliding attributes. In NeurIPS, 2017. 2.

[32] Xi Yin, Xiang Yu, Kihyuk Sohn, Xiaoming Liu, and Manmohan Chandraker. Towards large-pose face frontalization in the wild. In ICCV, 2017. 2.

[33] Chris Donahue, Akshay Balsubramani, Julian McAuley, and Zachary C. Lipton. Semantically decomposing the latent spaces of generative adversarial networks. In ICLR, 2018. 2.

[34] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In ICLR, 2018. 1, 2, 3, 4, 6, 8.