Face Manipulation with Generative Adversarial Network

Longpeng Xu

School of Information Engineering, Guangdong University of Technology, Guangzhou, 510000, China
DanielXu@mail2.gdut.edu.cn

Abstract. Generative adversarial network has appeared as an effective image manipulation tool in recent years and has been widely used. The GAN-based manipulation of face images is also possible and tools including DeepFake are already misused. In this paper, we discuss the pros and cons of face manipulation with generative adversarial network. We find that this technique can be very useful for recovering masked face and further improving face recognition accuracy.

1. Introduction

Face image processing has come into People's Daily life. In terms of entertainment, creating a fake face for a family photo will add to the atmosphere of the party, covering the faces of relatives with the fake faces, most of the time will be funny. But if this "interesting" technology is used in criminal cases, the consequences could be disastrous. Some fans will exchange the faces of their favorite celebrities with the faces of some of the characters in the videos, in order to entertain themselves. If what these fans are doing is just a personal act, not uploading fake products to social media, then it is only their own entertainment products, which cannot be regulated. But often the manipulated videos are posted on social media, where the creators of the videos maliciously replace the faces of public figures in order to smear them. What's more, replacing one's face on some ugly videos is a great insult to one's character and causes irreparable damage to the image of a public figure. We need to know that the current DeepFake technology has been very mature, ordinary people can easily create fake faces through some software, and the victims may be ordinary people like us rather than just public figures.

In addition to videos and images, the creation of fake portraits can fool identification, circumvent government identification checks, and steal information and money from others, and even pretend to be others for fraud. Biometric information is difficult to be changed by ordinary people, so it is widely used in the financial field. In fact, the real risk to users is the security of "biometric information" that can spin out of control. Just as the static and dynamic face images collected by some companies through face-changing apps some time ago are typical biometric information, which are sensitive personal physiological characteristics information like fingerprints and iris. Because it is difficult for ordinary people to change, so it is widely used in identity authentication, transaction and payment links, such as the information once leaked, is bound to bring lasting and difficult to eliminate the impact on users. The public's concern that "face-swapping" software threatens payment security is based on such considerations.

In this paper, we discuss both the benefits and harms of face manipulation with generative adversarial network. We first give a short review of related work in Section 2. Then we present a case study of masked face manipulation based on GAN in Section 3. We further discuss the damage and detection of DeepFake in Section 4. We conclude this paper in Section 5.
2. Related Work

2.1. Face Transfer
Face transfer is the transfer of one person's facial expressions and head poses to another's face. A variety of methods have been developed; traditional methods are based on face modeling. But recently, Generative Adversarial Network (GAN) are used in face transfer.

Face transfer has many applications. A cartoon face of a person can be generated in [1]. The author proposes a landmark auxiliary CycleGAN to solve the problem of generating high-quality cartoon faces that capture the basic facial features of people. CycleGAN is used to generate the facial image of the target person with the corresponding head pose and source facial expression, and PatchGAN is used to explore the influence of different receiving field sizes on the generated image [2]. InterFaceGAN is proposed to explain the unwrapped face representation obtained by the latest GAN model, and a comprehensive analysis of the attributes of face semantics in the latent space [3].

2.2. Face Attribute Editing
In face attribute editing, given a face image, most of the facial features, face shape, hair, and skin color can be changed, leaving only a few features. The editing is conducted on the same face picture, instead of transferring the features to another face picture.

The author uses the age difference between two age groups to capture facial aging areas with different attention factors, and proposes a new type of conditional attention normalized GAN (CAN-GAN) for age synthesis. It aims to freely convert the input surface to the aging surface [4]. In order to use another facial image to control the pose, expression and facial features of a given facial image, CtrlFaceNet is provided to control the source facial image while preserving the identity and skin color. It also introduces a method to train the framework in a fully self-supervised mode using a large-scale dataset of unconstrained face images [5]. A fairness-aware facial Image-to-Image translation model called FairFaceGAN is introduced to separate the information about protected attributes and that of target attributes [6].

2.3. Face Completion
Face completion is to restore the incomplete or occluded face image into a clear and complete face image without occlusion. In general, to restore a complete face, pairs of occluded and unoccluded faces are used to train the generative network. Previous face completion methods include an encoder-decoder architecture, which uses naive synthetic occlusion for training and uses global and local adversarial losses. The author provides a face completion encoder/decoder based on a convolution operator with a gating mechanism, and has been trained through a large number of face occlusions [7]. In order to realize that the system can remove one object at a time to perform face de-occlusion in facial images, a GAN-based network is used, and partial convolution and vanilla convolution operations in the generator part of the GAN network are integrated [8]. An algorithm is based on a neural network that directly trains the content of the missing region, and combines the reconstruction loss, two confrontation loss and semantic analysis loss for training, so as to ensure the authenticity of the pixel and the consistency of the local and global content [9].

2.4. Other Applications
Nowadays, People can easily edit an image by some applications. We can change its brightness, contrast, gray value or modify other details, even turn a real scene into a cartoon style. The author introduces a method called ChipGAN, which is an architecture based on an end-to-end generative confrontation network to convert photos into Chinese ink painting style [10]. A machine-assisted system called DeepMoney has been developed to distinguish between fake and real banknotes. In order to achieve this goal, the author adopted the latest machine learning model called Generative Adversarial Network (GAN) [11]. Fashion synthesis inspired a new type of conditional GAN architecture called Poly-GAN. It is an application that automatically places clothing on an image of a human body model in any posture.
Poly-GAN allows adjustments to multiple inputs and is suitable for many tasks, including image alignment, image stitching and inpainting. This technology makes it very convenient for people to choose clothes without trying them on [12].

GAN is not only applied in image processing, but also in other fields. GAN can be applied in Taxi-Based mobility demand. The author in [13] proposes a deep learning framework that aims to predict the needs of taxi passengers while considering space, time and external dependencies and uses a conditional generative confrontation network (CGAN) model.

3. Case Study of Face Manipulation based on GAN
In 2020, COVID-19 virus spread rapidly around the world. Mask has become a necessity in people's daily life. At present, facial recognition is widely applied in lots of areas. But now people wear masks hiding their faces, the conventional facial recognition technology become ineffective. Such as facial security checks at train stations and facial access control. Therefore, improving existing facial recognition technology to recognize faces even when people are wearing masks is becoming important. Recently, most facial recognition technologies are designed based on deep learning. Image editing methods base on deep learning have achieved effective results for removing object in an image. But it doesn’t work well with large, complex objects, especially in facial images [14]. According to previous studies, applying GAN to restore the face will be beneficial to improve the recognition accuracy. In this part, we use a public face dataset and a GAN-based face recovery scheme to further validate this opinion.

3.1. Dataset Description
In this part, we use the famous LFW data set. Labeled Faces in the Wild (LFW) [15] is a database of face photographs designed for studying the problem of unconstrained face recognition. This database was created and maintained by researchers at the University of Massachusetts, Amherst. 13,233 images of 5,749 people were detected and centered by the Viola Jones face detector and collected from the Internet. The dataset we used specifically is the deep-funneled version, which produced superior results for most face verification algorithms.

Because many people in the LFW dataset have only a few photos, and it is not conducive to the subsequent face recovery and recognition, we only use the seven people, each of which has at least 70 images. As for the situation that there may be a small sample in practical application, that is, a person has only a small number of photos, we leave it to the future research to explore. We use a total of 1,288 images, and the original image resolution is 125 x 94. In the following experiments, in order to reduce the overhead of computing resources, we unified the image size to 64×64. We divided the data into training set and test set with a ratio of 80%:20%.

In Figure 1, we show the original face image.

![Figure 1. The original face image.](image1)
![Figure 2. The occluded face image.](image2)
![Figure 3. The recovered face image.](image3)
3.2. Methods

3.2.1. Simulation of Masked Face. Since the input image we applied already corresponded to the result after face alignment, in order to simulate the situation in which the face was occluded by a mask, we simply cropped the lower half of the image, which roughly corresponded to the effect of information loss in the lower half of the face.

We show the occluded face image in Figure 2.

3.2.2. Restoration of Masked Face. In order to recover the complete face from the masked face, we need to take the complete face as input and the masked face as output, and train the generator part of a GAN model. Here we apply the existing work: https://github.com/ahmetmeleq/Face-Completion---Occlusion-Restoration-GAN

The details of the GAN model can be referred to the code in this link. We apply the original face corresponding to the training set and the cropped face as the input and output of the GAN model to carry out model training.

We show the recovered face image in Figure 3.

3.2.3. Face Recognition Comparison. In this part, we try to use a convolutional neural network model for face recognition. This is a classification problem, with seven people corresponding to the seven classes. The convolutional neural network model we applied uses 4 convolutional layers and 2 max pooling layers. Every two convolutional layers is followed by a max pooling layer. After the two max pooling layers, we apply the Dropout operation, and the probability parameter is set as 0.25. 32 5×5 convolution kernels are applied in the first two convolutional layers, while 64 5×5 convolution kernels are applied in the last two convolutional layers. After the convolution and pooling layer, we apply a fully connected network for classification, which contains two fully connected layers. The first fully connected layer contains 256 neurons, applying ReLU as the activation function, and the second fully connected layer contains 7 neurons, corresponding to 7 categories, applying Softmax as the activation function. The Dropout operation with a probability of 0.25 is also applied between the two full connection layers.

We compared the effectiveness of face recognition under five settings as shown in Table 1. We use Adam as the optimizer, and apply EarlyStopping mechanism to ease the overfitting problem. We further divide 10% of the data from the data set as the validation set. After 10 rounds of training, if the results on the validation set did not improve, we will stop the training.

3.3. Results

We give a summary of experiment results in Table 1.

| Experiment Number | Training Image  | Test Image | Test Accuracy |
|-------------------|----------------|------------|---------------|
| #1                | Original Faces | Original Faces | 0.8566         |
| #2                | Original Faces | Occluded Faces | 0.7132         |
| #3                | Original Faces | Recovered Faces | 0.1047        |
| #4                | Occluded Faces | Occluded Faces | 0.8411         |
| #5                | Recovered Faces | Recovered Faces | 0.8760        |

By comparing experiments 1-3, we find that the accuracy of the model trained directly on the original image dataset was still good (about 85%), but the performance of the masked face or the recovered face was greatly reduced (about 70% and 10%). By comparing experiments 4 and 5, we find that the retraining model is beneficial to the recognition of the obscured face or the restored face (up to 84% and 87% respectively). The face recovery method is better than the model trained directly with the rest of
the face (87%>84%), but this is only the result we gain from a small dataset and further validation is needed in the following studies.

The detailed confusion matrices for each experiment in the test set are shown in Figure 4-8, respectively.

Figure 4. The confusion matrix in the test set for the Experiment #1.

Figure 5. The confusion matrix in the test set for the Experiment #2.

Figure 6. The confusion matrix in the test set for the Experiment #3.

Figure 7. The confusion matrix in the test set for the Experiment #4.
4. Detection of DeepFake Images

While artificial intelligence techniques have been successfully used in many applications and brought a huge progress [16-19], the misuse of these techniques also raised a lot of concerns in recent years [20]. DeepFake information would have spread all over the world. In 2018, Deepfake videos have proliferated in the United States, many celebrities' faces being transferred to porn stars to be made into pornographic videos and circulated on social media [21]. In 2018, it is reported that Ali Bongo, the Gabon President in central Africa, has not been seen for a long time due to a stroke, but he suddenly appeared on TV on December 31, delivering a new speech to the people in Morocco. Bongo's political opponents immediately accused the video of Bongo's New Year's message as a deeply fabricated one, raising questions about the President's ability to continue to carry out his duties and triggering a coup by Gabon's military a week later [22].

From the perspective of the international security, DeepFake can be applied to plan false propaganda by criminals, and they will be able to control the direction of public opinion. DeepFake can also increases the chance of miscalculation and the risk of war, for example, forging the inflammatory remarks of a country's political leaders and celebrities can intensify the contradictions between countries. In addition, DeepFake can also erode the international trust system, using it to manufacture disseminates fabricated videos for the purpose of defamatory propaganda and subverting the government's credibility [23]. From the perspective of an individual, DeepFake digital images can be applied to steal other people's accounts, infringe the reputation of the people and embezzle the identity of others to make loans [24].

The detriment and influence of DeepFake has spread all over the world. The detection and defense of DeepFake has become one of the hot issues concerned by governments, enterprises and even individuals all over the world [25].

NIZAM et al. [26] proposed a new method aimed at extracting DeepFake fingerprints from images. This method is based on an expectation maximization algorithm that has been trained to detect and extract fingerprints representing the convolutional traces (CT) left by GAN in the image generation process. Compared with the latest technology in the DeepFake detection task, CT has a higher discriminative ability, can achieve better results, and also proves robustness to different attacks.

Wang et al. [27] proposed FakeSpotter, which is based on monitoring the behavior of neurons to discover artificial faces synthesized by AI. Research on neuron coverage and interaction has successfully
shown that they can be used as test standards for deep learning systems, especially in environments that are subject to adversarial attacks.

5. Conclusion

In this paper, we present both a quantitative analysis and the conceptual discussion of face manipulation based on generative adversarial network. We find that the GAN-based recovered faces are easier to identify, compared with occluded faces, e.g., masked faces, which shows the benefit of face manipulation. But we also point out the potential damages caused by DeepFake, which should be carefully detected and handled.

References

[1] Wu R, Gu X, Tao X, et al. (2019) Landmark assisted CycleGAN for cartoon face generation. arXiv preprint arXiv:1907.01424.
[2] Xu R, Zhou Z, Zhang W, et al. (2017) Face transfer with generative adversarial network. arXiv preprint arXiv:1710.06090.
[3] Shen Y, Yang C, Tang X, et al. (2020) InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs. arXiv preprint arXiv:2005.09635.
[4] Shi C, Zhang J, Yao Y, et al. (2020) CAN-GAN: Conditioned-Attention Normalized GAN for Face age Synthesis. Pattern Recognition Letters.
[5] Zeno B, Kalinovskiy I, Matveev Y, et al. (2020) CtrlFaceNet: Framework for Geometric-driven Face Image Synthesis. Pattern Recognition Letters.
[6] Hwang S, Park S, Kim D, et al. (2020) FairFaceGAN: Fairness-aware Facial Image-to-Image Translation[C]. Proceedings of the The 31st British Machine Vision Conference.
[7] Mathai J, Masi I, AbdAlmageed W. (2019) Does generative face completion help face recognition? [C]/2019 International Conference on Biometrics (ICB). IEEE, 1-8.
[8] Din N U, Javed K, Bae S, et al. (2020) Effective Removal of User-Selected Foreground Object from Facial Images Using a Novel GAN-Based Network. IEEE Access, 8: 109648-109661.
[9] Li Y, Liu S, Yang J, et al. (2017) Generative face completion[C]/Proceedings of the IEEE conference on computer vision and pattern recognition, 3911-3919.
[10] He B, Gao F, Ma D, et al. (2018) ChipGAN: A generative adversarial network for Chinese ink wash painting style transfer[C]/Proceedings of the 26th ACM international conference on Multimedia, 1172-1180.
[11] Ali T, Jan S, Akhodre A, et al. (2019) DeepMoney: counterfeit money detection using generative adversarial networks. PeerJ Computer Science, 5: e216.
[12] Pandey N, Savakis A. (2019) Poly-GAN: Multi-Conditioned GAN for Fashion Synthesis. arXiv preprint arXiv:1909.02165.
[13] Yu H, Chen X, Li Z, et al. (2019) Taxi-based mobility demand formulation and prediction using conditional generative adversarial network-driven learning approaches. IEEE Transactions on Intelligent Transportation Systems, 20(10):3888-3899.
[14] Din N U, Javed K, Bae S, et al. (2020) A novel GAN-based network for unmasking of masked face. IEEE Access, 8: 44276-44287.
[15] Huang G B, Mattar M, Berg T, et al. (2008) Labeled faces in the wild: A database forstudying face recognition in unconstrained environments.
[16] Jiang W, Zhang L. (2018) Geospatial data to images: A deep-learning framework for traffic forecasting. Tsinghua Science and Technology, 24(1): 52-64.
[17] Jiang W. (2020) Applications of deep learning in stock market prediction: recent progress. arXiv preprint arXiv:2003.01859.
[18] Jiang W, Zhang L. (2020) Edge-siamnet and edge-triplenet: New deep learning models for handwritten numeral recognition. IEICE Transactions on Information and Systems, 103(3): 720-723.
[19] Jiang W. (2020) Time series classification: nearest neighbor versus deep learning models. SN Applied Sciences, 2(4): 1-17.

[20] Tolosana R, Vera-Rodriguez R, Fierrez J, et al. (2020) Deepfakes and beyond: A survey of face manipulation and fake detection. arXiv preprint arXiv:2001.00179.

[21] Terrifying high-tech porn: Creepy 'deepfake' videos are on the rise. By John Brandon, Fox News https://www.foxnews.com/tech/terrifying-high-tech-porn-creepy-deepfake-videos-are-on-the-rise

[22] Corin Faife, InAfrica, (2019) Fear of State Violence Informs Deepfake Threat. https://blog.witness.org/2019/12/africa-fear-state-violence-informs-deepfake-threat/.

[23] Miao Z. (2020) Horrible DeepFake Technology (in Chinese). http://www.globalview.cn/html/global/info_35303.html.

[24] Xinhua Net. (2020) The Results of Face Information Leak (in Chinese). http://www.xinhuanet.com/fortune/2019-09/02/c_1124951913.htm.

[25] Liang R, Li M, Chen K, et al. (2020) A Survey of Audiovisual Deepfake Detection Techniques. Journal of Cyber Security, 005(002):1-17.

[26] Din N U, Javed K, Bae S, et al. (2020) Effective Removal of User-Selected Foreground Object from Facial Images Using a Novel GAN-Based Network. IEEE Access, 8: 109648-109661.

[27] Wang R, Juefei-Xu F, Ma L, et al. (2020) FakeSpotter: A simple yet robust baseline for spotting AI-synthesized fake faces//International Joint Conference on Artificial Intelligence (IJCAI).