As Time Goes by: Dynamic Face Aging Method Using CycleGAN

Wanzhou Liu
College of Computer Science & Technology
Qingdao University
Qingdao, Shandong Province, China
Corresponding author email：2019204400@qdu.edu.cn

Abstract

Achieving Image-to-image style transfer using various types of generative adversarial networks is a popular question in computer vision, with numerous works being proposed. However, the current typical research on the use of neural networks to achieve Image-to-image style transfer is to improve the output performance through the improvement of the model, and the intermediate process of style transfer has not been paid much attention. In this paper, based on the implementation of CycleGAN style translation, the gradients of non-leaf nodes are calculated and extracted by using the autograd function in PyTorch, and finally, the aging images of faces at different ages are dynamically generated by using the middle layer gradient. The experimental results based on the AGFW-v2 dataset tell us that this is capable for the initial research purpose, and the presentation of the process of style transfer intermediate enriches the stage changes of age. The processing of intermediate layer features in this study can also be used in areas such as Photo Enhancement, Image Dehazing and Face Swapping to achieve controllability of model processing results.

Keywords*—deep learning, style transfer, GAN, CycleGAN*

1. Introduction

The analysis and generation of images is a very important part of computer vision, with the help of computers, this project hopes to simulate the whole life of a person from adolescence to old age with just one picture, and to show the power of time with deep learning.

Making a reliable model is not such an easy job, since the aging trend I want to capture is easily affected by various conditions of the input image, such as facial expressions, the make-up of the character, or photographic settings. While it is indisputable that no one will not be interested in what they looked like in the past or the future, especially in a dynamic and gradual way.

Obtaining photos of one person that were taken at different ages is challenging and expensive. Therefore, unpaired image-to-image translation has gained a great deal of attention for applications in which paired data are unavailable or difficult to collect. A key problem of unpaired image-to-image translation is determining which properties of the source domain to preserve in the translated domain, and how to preserve them. Some many state-of-the-art methods and models have been used to implement image-to-image translations, including UNIT [1], DiscoGAN [2], DualGAN [3], and CycleGAN [4]. Also included is the recently updated ACL-GAN [5]. All those methods have unpaired training data using cycle-consistency loss as a constraint, while generally, the older CycleGAN still has better performance in processing more complex scenes.

There are quite a few models available in which to choose, but to accomplish this task, a single conversion from picture to picture alone will not work. Because a person's life cannot have only three stages: current picture, past, and future, but needs to be able to review or predict each of ages. Therefore, I decided to use CycleGAN and PyTorch's Autograd mechanics to automatically derive the gradient to extract the intermediate feature values from the training model and finally reach the target of face change in every decade.

2. Related Work

2.1 Generative Adversarial Network

GAN [6] includes two networks, a generative network and a discriminant network (Fig. 1). The generator will create some fake pictures and send them to the discriminator, then the discriminator will determine if these pictures are real pictures.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2021), edited by Jinping Liu, Yanping Chen, Feng Wu, Proc. of SPIE Vol. 12168, 121681C · © 2022 SPIE · 0277-786X · doi: 10.1117/12.2631420

And it will continue till the discriminator can't tell whether the picture is real. The core method of a GAN is based on the "indirect" training through the discriminator, which itself is also being updated dynamically. This enables the model to learn in an unsupervised manner.

The key to GAN's success is the idea of an adversarial loss that forces the generated images to be, in principle, indistinguishable from real photos. This loss is particularly powerful for image generation tasks, as this is exactly the objective that much of computer graphics aims to optimize.

2.2 Cycle Generative Adversarial Network

CycleGAN [4] comes from a paper in which is proposed a method that can capture the characteristics of one image domain and figure out how these characteristics could be translated into another image domain, all in the absence of any paired training examples. CycleGAN is essentially two symmetric GANS that form a circular network (Fig. 2). Two GANS share two generators with one discriminator each, meaning there are two discriminators and two generators. In particular, by forcing the translated images to fool the discriminator using a classical adversarial loss and translating those images back to the original images, cycle consistency ensures the translated images contain enough information from the original input images. This helps to build a reasonable mapping and generate high-quality results.

CycleGAN uses this cycle consistency loss to enable training without the need for paired data. In other words, it can translate from one domain to another without a one-to-one mapping between the source and target domain. This opens up the possibility to do a lot of interesting tasks like photo enhancement [7], image colorization [8], style transfer [9], etc. I hope to use CycleGAN to capture the changes of facial features during human aging and use it to achieve the effect of predicting people's facial change, to predict what they look like when they are young and decades after they grow old.

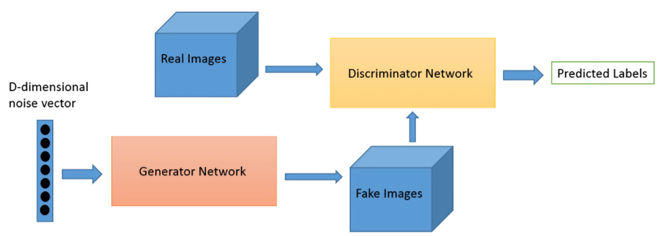

Figure 1. The Framework of GAN

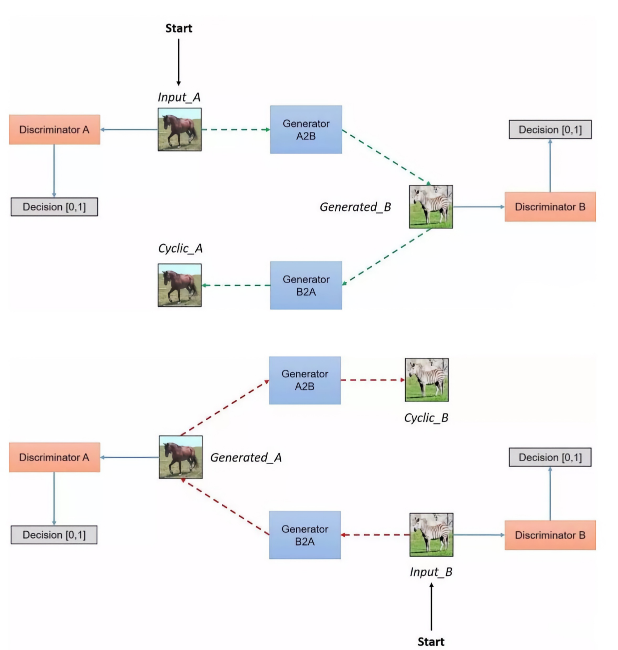

Figure 2. The Framework of CycleGAN

2.3 Autograd Mechanics in PyTorch

In the PyTorch documentation, autograd is a reverse automatic differentiation system. Conceptually, autograd records a graph recording all of the operations that created the data as user execute operations, giving a directed acyclic graph whose leaves are the input tensors and roots are the output tensors. By tracing this graph from roots to leaves, it can automatically compute the gradients using the chain rule.

Internally, autograd represents this graph as a graph of Function objects, which can be applied to compute the result of evaluating the graph. When computing the forwards pass, autograd simultaneously performs the requested computations and builds up a graph representing the function that computes the gradient. When the forwards pass is completed, it can evaluate this graph in the backwards pass to compute the gradients.

This research provides a method to implement intermediate layer feature extraction. With the help of automatic computation of gradients in PyTorch, the features of non-leaf nodes were extracted using the hook function, thus outputting the intermediate layer of image feature processing and achieving the transformation of images of different ages. It should be additionally noted that although care has been taken in finding the dataset to collect images of faces in their most natural state, women inevitably wear more or less make-up and this human factor can have an impact on the results, so the study ultimately decided to use only men without make-up for training and testing for the time being, and to use this as a basis for evaluating the method.

3. Method

3.1 Baseline

The generative network for CycleGAN uses the feed-forward transformation networks with perceptual loss functions of Johnson et al. [10] who have shown impressive results for neural style transfer and super-resolution. This network contains three convolutions, several residual blocks [11], two fractionally-strided convolutions with stride $\frac{1}{2}$, and a convolution that maps the image to RGB. They used six blocks for 128×128 images and nine blocks for 256×256 and higher resolution training images. Similar to Johnson et al. [10], CycleGAN uses instance normalization [12]. For the discriminator network, CycleGAN uses 70×70 PatchGANs [13], which aim to classify 70×70 overlapping image patches to determine whether they are real or fake. Such a patch-level discriminator architecture has fewer parameters than a full-image discriminator and can work on images of arbitrary size in a fully convolutional manner [13]. The adversarial losses' function of CycleGAN is:

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y\sim p_{data}(y)}\left[log\, D_y(y)\right] + \mathbb{E}_{x\sim p_{data}(x)}\left[log\left(1 - D_Y\big(G(X)\big)\right)\right] \qquad (1)$$

CycleGAN uses two techniques to stabilize the training procedure of the model. First, for $\mathcal{L}_{GAN}$ (Equation 1), a least-squares loss [14] is used instead of the negative log likelihood objective, which is more stable during training and generates higher quality results. In particular, for a GAN loss $\mathcal{L}_{GAN}$, they train the G to minimize $\mathbb{E}_{x\sim p_{data}(x)}\left[\big(D(G(x) - 1)\big)^2\right]$ and train the D to minimize $\mathbb{E}_{y\sim p_{data}(y)}[(D(y) - 1)^2] + \mathbb{E}_{x\sim p_{data}(x)}\left[D\big(G(x)\big)^2\right]$. Secondly, to reduce the oscillation [15] of the model, it uses the generated image history to update the discriminator instead of using the image generated by the latest generator. There is an image buffer that stores 50 previously generated images.

Table 1: The properties of AGFW-v2 in comparison with other aging databases.

Database	# Images	# Subjects	Label type	Image Type	Subject type
MORPH - Album 1	1,690	628	Years old	Mugshot	Non-famous
MORPH - Album 2	55,134	13,000	Years old	Mugshot	Non-famous
FG-NET	1,002	82	Years old	In-the-wild	Non-famous
AdienceFaces	26,580	2,984	Age groups	In-the-wild	Non-famous
CACD	163,446	2,000	Years old	In-the-wild	Celebrities
IMDB-WIKI	523,051	20,284	Years old	In-the-wild	Celebrities
AgeDB	16,488	568	Years old	In-the-wild	Celebrities
AGFW	18,685	14,185	Age groups	In-the-wild/Mugshot	Non-famous
AGFW-v2	**36,299**	**27,688**	**Age groups**	**In-the-wild/Mugshot**	**Non-famous**

3.2 Implementation

The quality of the age representation in the images in the many existing face datasets can significantly affect the learning process of model aging. While selecting a dataset, the number of longitudinal face image samples per subject, the number of subjects, the range of total age samples and the distribution of the age should be all taken into account for better results. Many public databases for age estimation or age growth systems are very limited in terms of the total number of images, the number of images per target, and the longitudinal separation of target samples in the dataset, such as FG-NET, MORPH, and AgeDB. There are also datasets that are relatively larger in size, but with noisy age labels, i.e. CACD [16], IMDB-WIKI [17]. For this work, I chose to use an extension of Aging Faces in the Wild (AGFW-v2) [18]. Table 1 shows where this dataset differs from the others.

AGFW [19] first introduced 18,685 images, including an age range from 10 to 64 years. The second database of two times the size of the AGFW was generated based on the AGFW collection criteria. In contrast to other age-related databases, the majority of those collected in AGFW-v2 were not public figures and did not have significant make-up or facial grooming, which contributed to a more accurate aging effect during the learning process.

AGFW-v2 [18] was collected from three sources. The first is a search engine, and most of the images come from the everyday lives of non-celebrities. It also contains publicly available metadata related to the age of the person being collected. The second part comes from publicly accessible mugshot images, which are passport-style photographs with ages provided by services. Finally, data from the Productive Aging Laboratory is also included. In total, AGFW-v2 consists of 36,299 images divided into 11 age groups spanning 5 years. All images will be re-sized to 128 x 128 to train the CycleGAN model.

In order to generate images of more age groups of a person, I first differentiated the dataset after removing the makeup optimization in more detail, classifying the dataset from 10 to 60 years old, and transforming the age into labels for training. A vector of 1*6 was set up and corresponded to ages 10 to 60 in order. Then the Label and picture features are executed concat operation and sent to the network generator without changing the decoder. with the help of the function register_forward_pre_hook(vis), to print the feature map. The experiment aims to generate the face images with elder ages using the 10s pictures. By the controlling of the age vector, we can dynamically generate the faces with different ages.

This allows the extraction of non-leaf node gradients by the hook function through the automatic computation of gradients in PyTorch, thus outputting the intermediate layer for the processing of picture features and finally outputting pictures of different ages. First, with the age vector, I send the images into the generator to generate the aging face with predefined ages, then, the discriminator determines the confidence of the fake image. This step is the same as the test step without the discriminator. This method can generate 20s' - 60s' aging images by the given vector. I train 50 epochs with the learning rate set to 0.002. The SGD is selected as the optimizer for this research and finally, it turns out that the network was optimized with convergence.

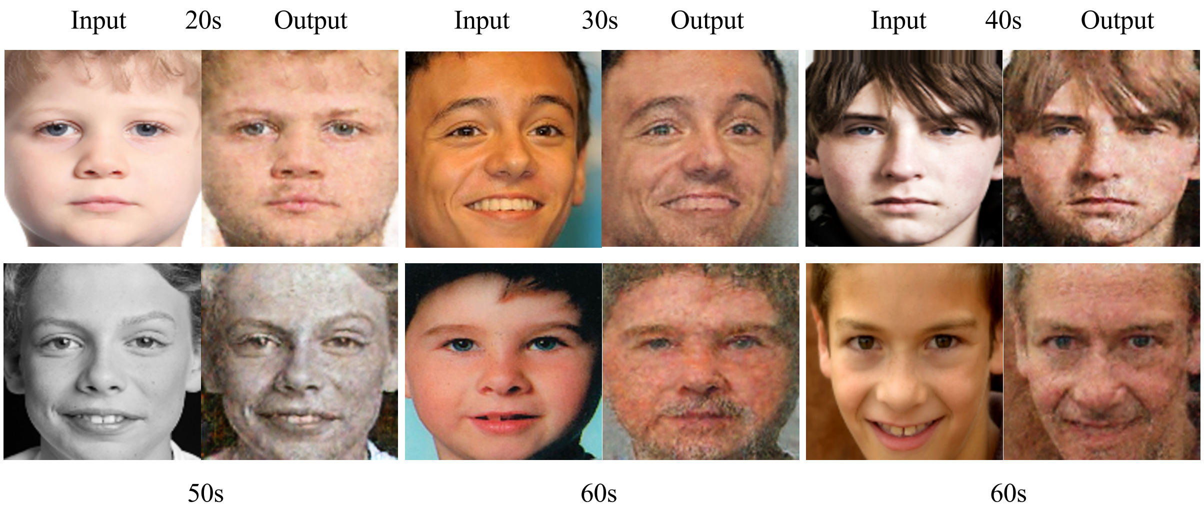

Figure 3. Examples of changes for different age groups.

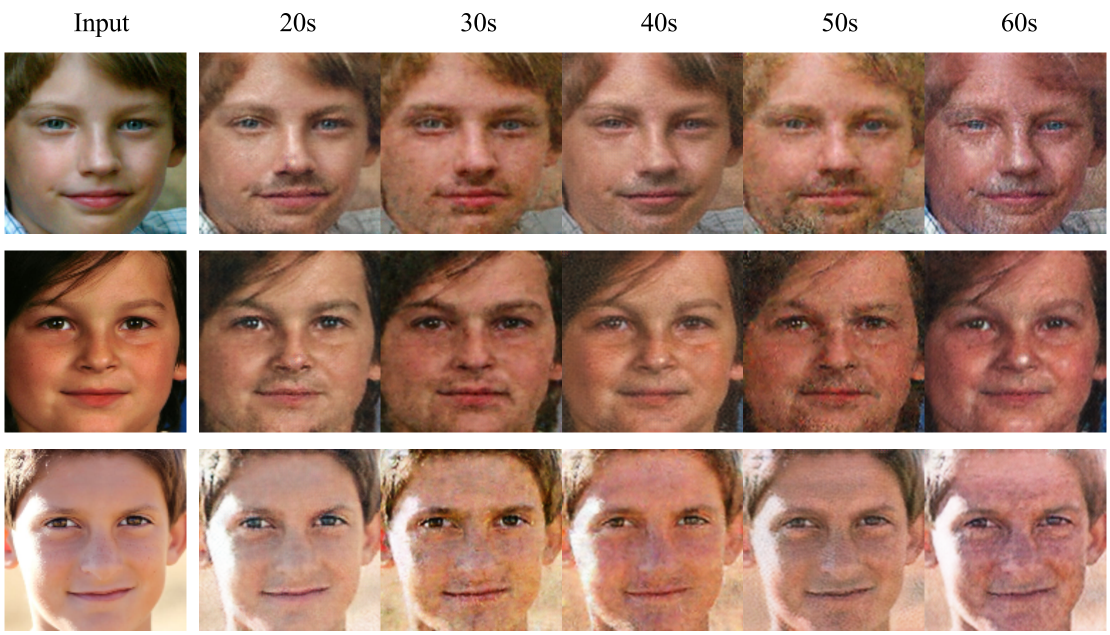

Figure 4. Outputs of a person's complete process of change from childhood to old.

4. Results

From Fig.3, we can clearly see the difference in how much the degree of change is at different ages, where the greater the difference between the target age and the current age, the more pronounced the change will be. The trained model will significantly change the color of the hair, also for facial hair, the model will treat it differently, it will choose to reduce the hair of the older person and increase its beard and eyebrows, and change the color of them. At the same time, the skin tone of people's faces will change, existing wrinkles will deepen and new ones may appear.

In the case of children moving into their twenties, the change is seen to deepen the color of the hair, add less beard and darken the eyebrows, in the case of middle age it deepens the wrinkles in the picture and fades the color of the eyebrows, and in the case of older age the hair becomes much lighter than before in color and the wrinkles in the face more pronounced.

This trending change is more visually apparent in the process of dynamically generating as complete an aging image of a person as possible. For the first set of images the hair color begins to lighten and thin as he reaches middle age, the beard becomes thicker, and again turns from black to white after age 50, while the skin tone also changes somewhat. The person in the second set of images has the most pronounced beard at age 50, but it becomes light again in the simulation at age 60, and the wrinkles on the face maintain a continuous change from inconspicuous to pronounced. The third group of images also produced significant changes in appearance with respect to age, but the beard was not a feature of change in this group, but rather the hairline and the degree of facial luster, with a significantly receding hairline and some dulling of facial luster seen in the later period.

Unfortunately, the arrangement of people's facial features did not change more significantly, which is not quite consistent with the fact that the ability to handle geometric transformations is something that needs to be improved, including CycleGAN and all generative adversarial networks. Also, the processing of the image affects its clarity, and it is more evident during the image processing that the subject of the photo changes while additionally changing the unnecessary background.

5. Conclusion

This work has presented a method, which is to extract the training features of the middle layer of the neural network by recording the gradient changes in training, and finally to achieve the style migration results of dynamically generated images.

In this study, the dataset was trained several times for different age groups, and with the help of automatic computation of gradients in PyTorch, the features of non-leaf nodes were extracted using the hook function, thus outputting the intermediate layer of image feature processing and achieving the transformation of images of different ages.

Experiments on AGFW-v2 dataset show our method is capable of handling dynamic and staged aging problems of the human face. This method can show more details of the model's processing and changes of the image, which is important in some changes that need to show more intermediate processes, such as how the landscape changes throughout different seasons of the year, choosing the level of image processing in Photo Enhancement [20] and Image Dehazing [21] autonomously, or even selection of the level of detail processing for face changes or expression changes [22].

REFERENCES

[1] Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. In *Advances in neural information processing systems* (pp. 700-708).

[2] Liu, C., Chang, X., & Shen, Y. D. (2020). Unity style transfer for person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6887-6896).

[3] Li, Y., Shi, T., Chen, W., Zhang, Y., Wang, Z., & Li, H. (2020). Unsupervised Style Transfer via Dualgan for Cross-Domain Aerial Image Classification. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1385-1388). IEEE.

[4] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232).

[5] Zhao, Y., Wu, R., & Dong, H. (2020, August). Unpaired image-to-image translation using adversarial consistency loss. In *European Conference on Computer Vision* (pp. 800-815). Springer, Cham.

[6] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

[7] You, Q., Wan, C., Sun, J., Shen, J., Ye, H., & Yu, Q. (2019, July). Fundus image enhancement method based on CycleGAN. In *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 4500-4503). IEEE.

[8] Anwar, S., Tahir, M., Li, C., Mian, A., Khan, F. S., & Muzaffar, A. W. (2020). Image colorization: A survey and dataset. *arXiv preprint arXiv:2008.10774*.

[9] Hu, C., Ding, Y., & Li, Y. (2020, June). Image Style Transfer based on Generative Adversarial Network. In *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)* (Vol. 1, pp. 2098-2102). IEEE.

[10] Johnson, J., Alahi, A., & Fei-Fei, L. (2016, October). Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision* (pp. 694-711). Springer, Cham.

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[12] Jing, Y., Liu, X., Ding, Y., Wang, X., Ding, E., Song, M., & Wen, S. (2020, April). Dynamic instance normalization for arbitrary style transfer. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 04, pp. 4369-4376).

[13] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

[14] Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., & Paul Smolley, S. (2017). Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2794-2802).

[15] Goodfellow, I. (2016). Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*.

[16] Li, P., Hu, Y., Li, Q., He, R., & Sun, Z. (2018, August). Global and local consistent age generative adversarial networks. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 1073-1078). IEEE.

[17] Rothe, R., Timofte, R., & Van Gool, L. (2018). Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, *126*(2), 144-157.

[18] Duong, C. N., Luu, K., Quach, K. G., Nguyen, N., Patterson, E., Bui, T. D., & Le, N. (2019). Automatic face aging in videos via deep reinforcement learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10013-10022).

[19] Nhan Duong, C., Luu, K., Gia Quach, K., & Bui, T. D. (2016). Longitudinal face modeling via temporal deep restricted boltzmann machines. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5772-5780).

[20] Omiya, M., Horiuchi, Y., Simo-Serra, E., Iizuka, S., & Ishikawa, H. (2019, June). Optimization-Based Data Generation for Photo Enhancement. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 1890-1898). IEEE.

[21] Ren, W., Pan, J., Zhang, H., Cao, X., & Yang, M. H. (2020). Single image dehazing via multi-scale convolutional neural networks with holistic edges. *International Journal of Computer Vision*, *128*(1), 240-259.

[22] Peng, B., Fan, H., Wang, W., Dong, J., & Lyu, S. (2021). A Unified Framework for High Fidelity Face Swap and Expression Reenactment. *IEEE Transactions on Circuits and Systems for Video Technology*.