Face Translation based on Semantic Style Transfer and Rendering from One Single Image

Peizhen Lin
Shenzhen Institute of Advanced Technology (SIAT) of the Chinese Academy of Sciences (CAS)
pz.lin@siat.ac.cn

Baoyu Liu
College of Computer Science and Software Engineering, Shenzhen University
1910272003@email.szu.edu.cn

Lei Wang*
Shenzhen Institute of Advanced Technology (SIAT) of the Chinese Academy of Sciences (CAS)
lei.wang1@siat.ac.cn

Zetong Lei
Nanjing University
lzt@smail.nju.edu.cn

Jun Cheng
Shenzhen Institute of Advanced Technology (SIAT) of the Chinese Academy of Sciences (CAS),
jun.cheng@siat.ac.cnCorresponding author: Lei Wang.

ABSTRACT

Many avatar characters have been animated in films or games, which always need a lot of time for post-processing with the computer graphics technologies. In recent years, lots of deep learning based methods have been proposed for face translation and image generation, which always require a large amount of data for training. However, there are few samples for special characters' prototype. In this paper, we present one face translation framework for translating human faces to that with visual effects from one single prototype image. The proposed framework consists of three modules. We first design one module to generate semantic face mask–the semantic mask generating (SMG) module. According to the semantic mask, the face color tone can be changed to that of the prototype. So we design the semantic color transfer (SCT) module. For the local textures, we design the deformation and rendering (DR) module. Experiments show that the proposed framework can generate images with prototype's visual effects while preserving the original person's identification and expression information.

CCS CONCEPTS

• **Applied computing**; • **Computer graphics**; • **Machine learning**;

KEYWORDS

Deep learning, Face translation, Mask Generative Adversarial Networks,

ACM Reference Format:
Peizhen Lin, Baoyu Liu, Lei Wang*, Zetong Lei, and Jun Cheng. 2021. Face Translation based on Semantic Style Transfer and Rendering from One Single Image. In *2021 10th International Conference on Software and Computer Applications (ICSCA 2021), February 23–26, 2021, Kuala Lumpur, Malaysia.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3457784.3457811

ICSCA 2021, February 23–26, 2021, Kuala Lumpur, Malaysia

ACM ISBN 978-1-4503-8882-5/21/02…$15.00
https://doi.org/10.1145/3457784.3457811

1 INTRODUCTION

In the production of modern movies, especially the kind of science fiction (sci-fi), special visual effect is a very important component, but expensive and ambitious. Actors have to wear motion and expression capture devices, then visual effect artists use a lot of computer graphics (CG) technologies to make the vivid characters appear in the film. The filmmakers usually need a lot of manpower and material resources, spend large costs to achieve good visual effects, and they always need a lot of time for makeup or coping with the CG technologies.

The Generative Adversarial Network (GAN) [1] is proposed in recent years, and it is particularly good at handling the transformation of images due to its power of adversarial training of generation and discrimination. Various kinds of GAN have been proposed to improve the network's stability and realness of the generated images. For example, pix2pix [2] implements transformation between paired pictures, while CycleGAN [3] makes changing between unpaired pictures. In addition, MaskGAN [4] can use the mask generated by semantic segmentation of the face area and combine it with the face image itself to synthesize the face of any specified shape. However, these methods either need paired datasets for training or limited to human face translation.

In this paper, we propose a new framework for face translation with semantic style transfer and rendering, which consists of three modules. The first is the Semantic Mask Generating (SMG) module, designed to generate a face mask for the target person as a semantic segmentation map, so that controllable operations on the face can be realized in the following. The second is the Semantic Color Transfer (SCT) module, designed to transfer the face color tone to that of the prototype source image, and the transferring is limited to specific areas according to the face semantic mask. The third is the Deformation and Rendering (DR) module, aiming to achieve deformation and render the target human face image with the special features of the source image. Only in the first module, a face

dataset is needed for mask generation. Only one prototype image is used as supervision, and satisfactory results have been obtained by the proposed framework.

2 RELATED WORKS

In this section, we will review about related works, including image synthesis and style transfer, facial conversion and translation, face modeling and rendering.

2.1 Image Synthesis and Style Transfer

Image synthesis and style transfer have been researched and applied in many areas of computer vision and graphics [5], [6], [7]. In recent years, this task is still a topic of general interest [8, 9]. By learning a mapping from observed image and random noise vector to output image, as well as loss function to train this mapping, conditional GAN [10] provides a general-purpose solution to image-to-image translation problem. It can synthesize photos from label or edge maps, colorize images. However, the results are limited to low-resolution and still far from realistic. A high resolution image synthesis solution is proposed in [2]. In [11], a method is introduced for game character auto-creation. The character customization is formulated under a facial similarity measurement and parameter searching paradigm, while an imitator is also used to generate face images.

2.2 Facial Conversion and Translation

Facial feature conversion includes the conversion of facial details, such as changing a person's mouth from closed to open or turning a non-smile face into a smiling face. These transformations usually use a single vector to mark the attributes of a specific image, and then make the neural network learn the corresponding content. For example, StarGAN [12] uses a single generator to process the mutual conversion of multiple image domains. Such kind of method is used to change certain facial details of the faces. Although sometimes there are some changes towards the sci-fi movie level, using these methods to achieve large-scale changes in face shape is still very difficult, since we need to mark them in the vector, and there are few images for training.

Some methods are suitable for automatic make-up. For example, BeautyGAN [13] uses histogram matching to achieve the cosmetic effect of a specific position on the target face. PSGAN [14] uses the makeup distillation network to realize the migration of makeup effect even if the pose changes greatly. However, when style transfer other than makeup transfer is needed, these methods are ineffective.

In recent years, face swapping technology has made great progress. The Deepfakes method [15] uses one realistic face exchange technology to accomplish face replacement frame-by-frame to form a video. FaceSwap [16] uses 3D Morphable Model (3DMM) [17] to swap one texture face into different people's faces. And changing the parameters of the 3DMM can drive the expression and posture of the driver's face.

The mask-based face transformation method aims to change the shape or features of the face as a whole while maintaining the details of the face features. MaskGAN [4] is the first method to use facial mask for this target. It can generate new faces that match the geometric features of the face mask and have the texture features of the original face. SEAN [18] ameliorates MaskGAN by using only selected parts of the mask to process these positions on the original face. SEAN uses its Semantic Region-Adaptive Normalization to control the style of each semantic region of the face with the extracted style vectors. It can combine several images' style on the desired output image.

2.3 Face Modeling and Rendering

Face modeling has been researched for some decades, and high-fidelity results have been achieved by sophisticated 3D facial capture system with animation algorithms [19], [20]. However, these systems are too complex for mainstream application. In recent years, some methods have been proposed to extract facial shapes and appearances from a single image [21], [22]. Various deep neural networks have also been applied to improve the face modeling quality [23], [24], [25]. In [26], a convolution neural network (CNN)-based encoder learns to extract semantically parameters, while the decoder is an expert-designed generative model. A code vector of face pose, shape, expression, skin reflectance or scene illumination is input to the decoder for reconstruction. However, facial hair and occlusions are challenging to handle by these methods.

A dynamic 3D avatar creating algorithm is developed in [27] to handle hairstyles and headwear, while generating expressive facial animations. However this method requires 32 input images. By using polygonal strips for hair rendering, a framework is introduced in [28] to generate a complete 3D avatar from a single image.

3 OUR APPROACH

The proposed framework is designed to translate human faces to that with visual effects from one single prototype image. It consists of three modules, i.e. Semantic Mask Generating (SMG), Semantic Color Transfer (SCT), Deformation and Rendering (DR), as shown in Figure 1. The first module is designed to generate face semantic mask so that controllable operations can be facilitated on the face's different parts. The second is aiming to transfer the source image's color tone according to the semantic mask. The third is to transfer the special shape and texture features to the face. Taking one person's photo as input, the proposed framework will generate the translated image with the prototype source image's style, while preserving the target's identity information and expression. We will introduce the three modules' details in this section as follows.

3.1 Semantic Mask Generating

The first module is designed for automatic semantic face mask generation to segment different parts of the face, e.g. hairs, eyebrow, nose, lips, mouth, eyes, etc. Due to the feature representation ability of deep neural networks (DNNs), we use a semantic segmentation network for the face mask generation.

We design a network based on U-net [29], and take the paired face photo-mask data of the CelebA-HQ [4] as training set, then we get a network model for the face parsing network. The face semantic mask will be generated by this model-based network. Using this network will avoid the expense of manually segmenting and labeling faces, since it learns the feature representation of different face parts. It guarantees that the following face color transfer can be implemented on faces that are not in the training

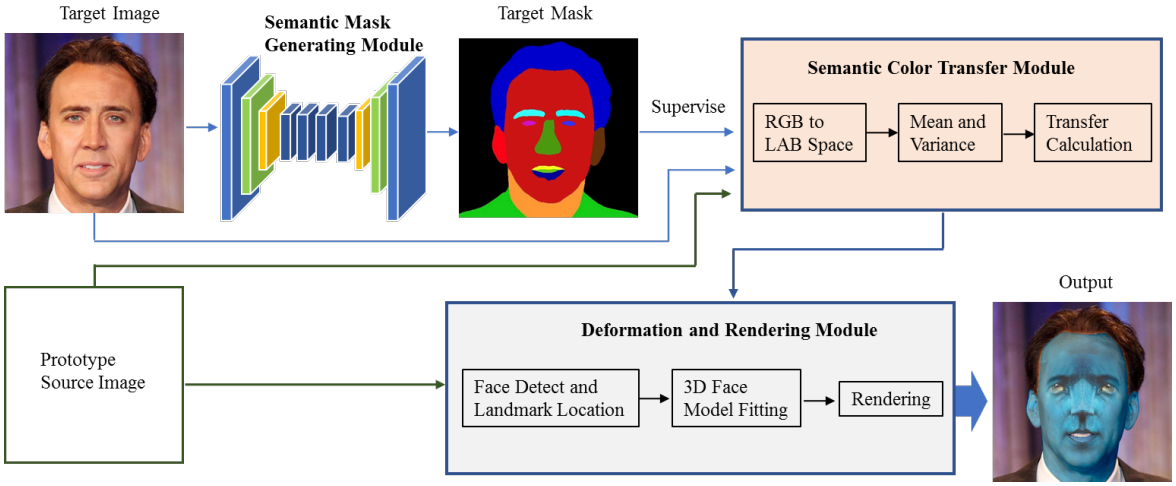

Figure 1: The framework of our method, which consists of three modules, i.e. semantic mask generation module, semantic color transfer module, deformation and rendering module. One single prototype source image is needed as supervision, and the target person's face image will be translated to that with visual effects.

dataset, making our method more applicable for practical use. The semantic face mask will be used for the next module.

3.2 Semantic Color Transfer

The second module is used to transfer the face skin color to that of the source image, since the color is one important feature for face. Instead of using deep learning-based methods which perform well but need a large amount of training data, we use a statistical computation method.

The input for this module includes two images and their masks. First, a photo of a normal person's face is needed, referred as the *Target Image*. And the corresponding mask of this face photo, is named as the *Target Mask*. This is obtained from the Semantic Mask Generating module. The second is a *Source* Image that provides the color style to transfer, as well as its mask.

The *Target Mask* will teach the face Color Transfer module which parts of the original face need color transferring. In our experiments, the module only performs color transfer on the parts such as the person's facial skin, ears, nose, eyebrows, mouth, and neck. As for other parts suggesting by the facial mask instructions, such as hair, background, clothes, eyes, teeth, etc., we choose to retain its original RGB values.

Taking the mask into consideration, we modified the method of [30] for our semantic face color transfer. In general, we realize color transfer by the following four steps:

1) Convert the input image's RGB value into the LAB space.

2) Calculate the mean and variance of each channel in the LAB space of the *Target Image* according to its mask, represented as $\mu^L_{T_{mask}}, \mu^A_{T_{mask}}, \mu^B_{T_{mask}}, \sigma^L_{T_{mask}}, \sigma^A_{T_{mask}}, \sigma^B_{T_{mask}}$, respectively, and *Source Image* according to corresponding semantic mask, represented as $\mu^L_{S_{mask}}, \mu^A_{S_{mask}}, \mu^B_{S_{mask}}, \sigma^L_{S_{mask}}, \sigma^A_{S_{mask}}, \sigma^B_{S_{mask}}$, respectively.

3) Calculate the new LAB values for the *Target Image* according to its mask as follows,

$$l'_{T_{mask}} = \lambda_L \frac{\sigma^L_{T_{mask}}}{\sigma^L_{S_{mask}}} \left(l_{T_{mask}} - \mu^L_{T_{mask}} \right) + \mu^L_{S_{mask}}$$

$$\alpha'_{T_{mask}} = \lambda_A \frac{\sigma^A_{T_{mask}}}{\sigma^A_{S_{mask}}} \left(\alpha_{T_{mask}} - \mu^A_{T_{mask}} \right) + \mu^A_{S_{mask}}$$

$$\beta'_{T_{mask}} = \lambda_B \frac{\sigma^B_{T_{mask}}}{\sigma^A_{S_{mask}}} \left(\beta_{T_{mask}} - \mu^B_{T_{mask}} \right) + \mu^B_{S_{mask}}$$

where $l'_{T_{mask}}$, $\alpha'_{T_{mask}}$, $\beta'_{T_{mask}}$ represent the new LAB values, and λ_L, λ_A, λ_B represent control parameters for LAB channels.

4) Convert the LAB values to RGB space.

The new image obtained in this way is color-transformed in the specified parts of the *Target Image* so that these parts have the source image's color style. In the meantime, those parts that do not need to be changed suggested by the facial mask remain the same as before.

3.3 Deformation and Rendering

After the above two modules, we have got an image with the prototype character's color tone in the designated face area. But another important factor has not been taken into account, i.e. the special shape and texture features of the prototype characters.

The Deformation and Rendering (DR) module is shown in Figure 2, we further process the results produced by the above module to make it have similar shape and texture features. Both the prototype *Source* image and the color transfered *Target* image will be first detected whether faces exist and the face landmarks are located. Then a 3D face model is fitted to the detected face, and rendered by the textures of the *Source* image at the corresponding face coordinates of the *Target* image.

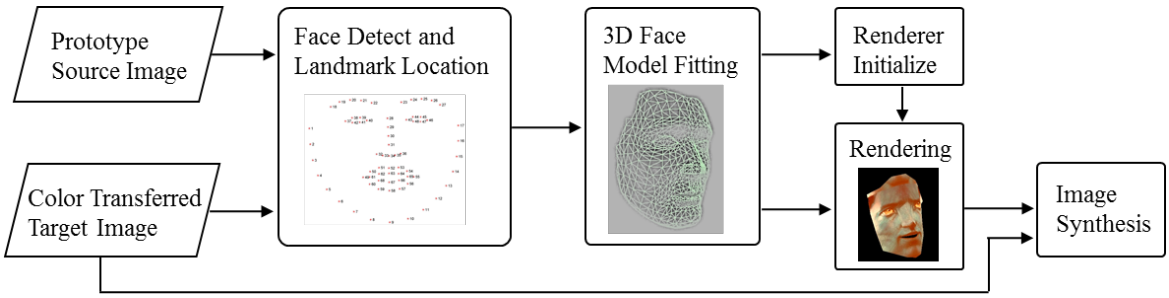

Figure 2: The Deformation and Rendering (DR) module

Specifically, our DR module uses the 3D morphable model [17] together with the landmark detection method to perform translation of special facial features. First we detect the facial area in the source image, and locate 68 face landmarks. Then we build a 3D model that fits the landmarks. And we initialize one renderer based on the 3D model coordinates as well as the source and target's textures.

The same procedures of face detection and landmark location will be necessary for the target image, i.e. the photo of normal person, so a set of landmarks of this human will also be detected and saved. Through non-linear fitting of the landmarks to the 3D face model, the person's 3D face shape will be obtained, and rendered with the above initialized renderer. So that the source information is rendered to the target 3D face. Then the rendered image is blended with the target image to obtain a complete face image to preserve the target's identity information. To be specific, the 3D face model is projected onto the image space using the following equation.

$$\mathrm{s} = \mathrm{aP}\left(S_0 + \sum_{i=1}^{n} w_i \times S_i\right) + t$$

where s is the projected shape of the image's 3D face model, a is a scaling parameter, P represent the first two rows of a rotation matrix that rotates the 3D face shape, S_0 is the neutral face shape, $w_i (1 \leq i \leq n)$ represents the blendshape weights, $S_i (1 \leq i \leq n)$ are the blendshapes, t is a 2D translation vector, and n is the number of blendshapes. Our module uses this equation to project the 3D face model onto image space, and harnesses Gauss Newton method to minimize the difference between the landmarks and the projected shape. As a result, the facial features on the *Source Image* will be transfered to the target image.

4 EXPERIMENT

In this section, we will show experimental results to illustrate the efficiency of the proposed framework. Also we will compare with state-of-the-art methods, including deep learning and GAN-based approaches.

4.1 Experiment Setting

Based on the pre-trained face parsing network, we can generate new masks for unseen person's photo. So our framework doesn't need more learning parameters.

We collected photos of some actors as the target images. In the CelebA-HQ [4] dataset, there are many movie stars' photos and corresponding face masks that can be used directly. However, we are not using them, and the photos in our experiment were not deliberately collected from the dataset. Instead, we collected from the Internet to evaluate the proposed method's generality. For the *Source Image*, we synthesize one image of *Na'vi* based on the character of the movie *Avatar* as supervision.

4.2 Experimental Results

We have tested a lot of images, and some results have been given in Figure 3, from which we can see that our method has realized face style transformation. In general, the images transferred by our method have special facial features, such as their eyeballs and noses that are different from human. Remarkably, some dot patterns typically seen in *Na'vis* face can be found in the generated face, indicating some fine details of the source image are also kept using our pipeline. At the same time, our method has fairly preserved the original person's features and expressions. Different persons' individual characteristics have been preserved.

We have also taken one prototype image of the '*Talos*' as supervision for experiments, and the results have been given in Figure 4. From Figure 4, we can see that photo-realistic images have been generated. The distinct features of *Talos* have been translated to the human face while the identity and expression information of the person has been preserved.

4.3 Comparison with State-of-the-art Methods

We also compared with existing state-of-the-art methods that can also be used for face translation. The experimental results have been shown in Figure 5, from which we can see that the Histogram Matching [31] method only matches the original RGB histogram to that of the prototype image, without transforming the facial features, and the face color is not similar with that of *Na'vis*.

MaskGAN [4] can successfully transform the human face to have similar shape of *Na'vis*, but it transfers neither the color tone nor the special face features. And unnatural deformation has happened on the eyes and nose. The FaceSwap [18] method brings a lot of *Na'vis* face features to the human face, but it has not changed the whole style of the human face. The Ebsynth [32] method performs well in color transformation and identity preservation, but there are artifacts and the new face looks unreal. The same problem also arises in CycleGAN [3] and pix2pixHD [2] methods. As can be seen, they both make the human face acquire *Na'vis* color tone, but the results are more or less blurry with artifacts. This is a typical problem encountered when using GAN for face transformation, as

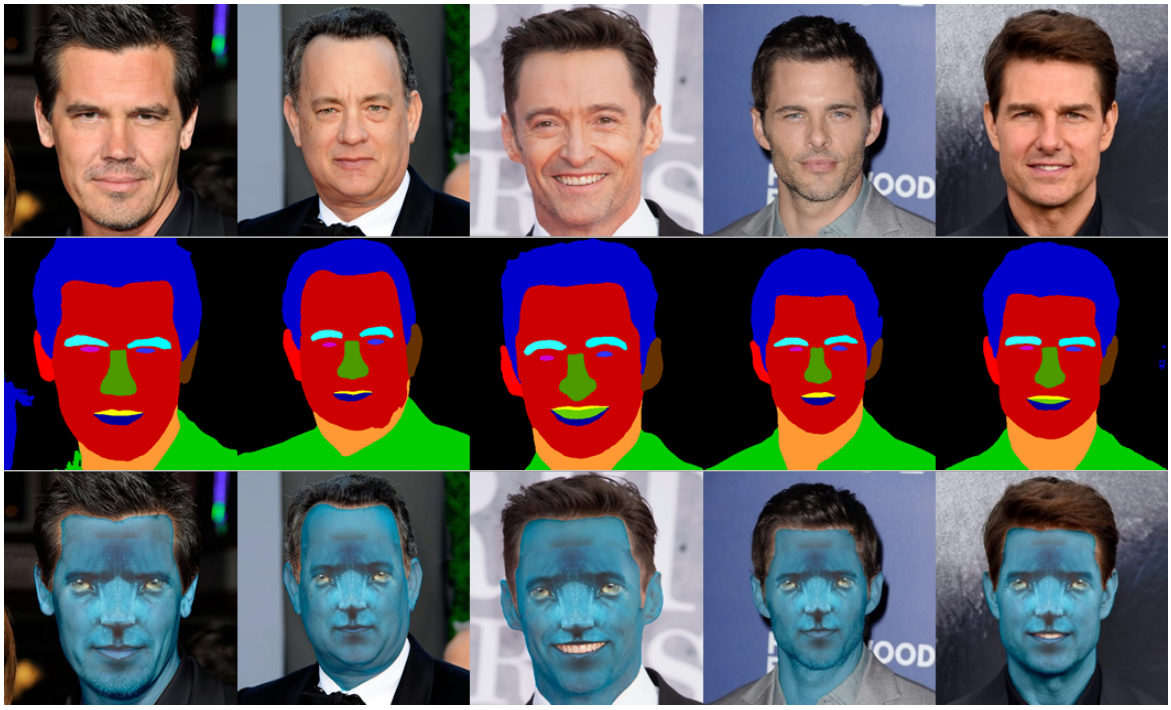

Figure 3: Experimental results on different persons. From top to down: original human face, generated face semantic mask, deformation and rendered results.

both of them are built on GAN. Especially when training data is very few for the network to learn, the generated images' quality degraded seriously. In contrast, our result has both *Na'vis* color and facial features, which are more vivid than other methods. At the same time, our method can retain better identity information and expression of the original person.

We also compare these methods quantitatively, by using the criteria of Frechet Inception Distance (FID) score [33] to evaluate the generated image's quality. FID score is a measurement of the distance between the feature vector of the real image and the generated image. Lower scores mean the images have higher similarity.

In our experiment, the FID score is calculated between the generated images and a set of genuine *Na'vis* photos. The results have been listed in Table I, from which it can be seen that the FID score of our method is lower than others except for CycleGAN and Ebsynth. The reason is that CycleGAN [3] takes advantage of the background alternation. It learns different styles between two set of images, so it can create some changes in the background. While our method does nothing to the background, thus having a higher score. Ebsynth [32] performs well in the overall color adaptation, but the quality and texture of its results are far from satisfactory.

Considering the influence of background on FID scores, it's unfair to fully assess the methods' ability for the face transformation. Using the masks generated in our SMG module, we calculate the FID scores without background, and the results have also been listed in Table 1. It can be seen that, our method obtains a better FID score. It is lower than most methods, only slightly higher than CycleGAN [3] which uses a large image-set for adversarial training. The large-scale training helps CycleGAN learn style level details, and this kind of learning method typically consumes a lot of sources and time. While our method can generate the result from a single image in seconds. Besides, although our method's FID score is slightly higher, but the subjective quality of our results is much better, which are more natural and have less artifacts.

5 CONCLUSION AND FUTURE WORKS

We propose a new framework for face translation with only one prototype image as supervision. The framework includes three modules–SMG, SCT and DR. Our method is still far from perfection. Although the output images have achieved the facial features of the prototype source image, we did not deal with the original person's hair, which part is still retained. Moreover, we did not accomplish geometric adjustments to the character's whole face. In the future, we will improve the method by generating corresponding hair and ear models, and more 3D results will be generated. Transfer learning will be further studied and used for the target.

ACKNOWLEDGMENTS

This work was supported in part by the National Key R&D Program of China (2018YFB1308000), the National Natural Science Foundation of China (U1713213, 61772508, 61976143), in part by Shenzhen Technology Project (JCYJ20170413152535587), CAS Key Technology Talent Program, Guangdong Technology Program (2016B010108010, 2017B010110007), Shenzhen Engineering Laboratory for 3D Content Generating Technologies (NO. [2017] 476),

Figure 4: Experimental results with one prototype image of '*Talos*' as supervision. Top: original person's face image. Mid: generated face with visual effects. Bottom: Zoomed-in parts of the generated face.

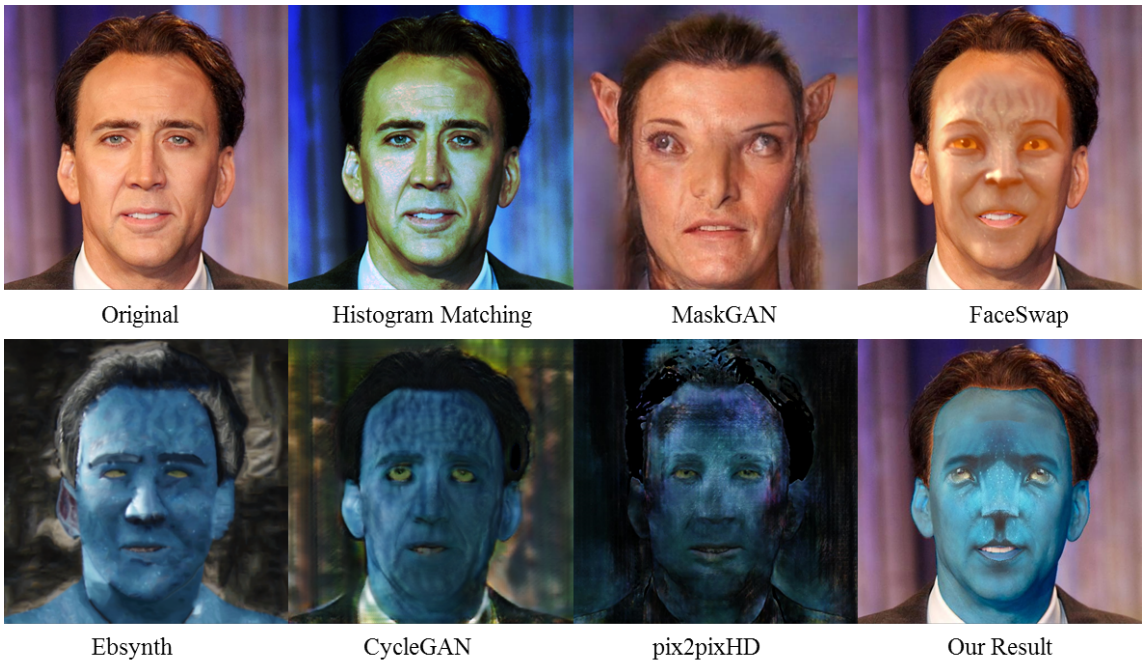

Figure 5: Comparison with the state-of-the-art methods

CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (2014DP173025), Guangdong-Hong Kong-Macao Joint Laboratory of Human-Machine Intelligence-Synergy Systems (2019B121205007).

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Advances in Neural Information Processing Systems 27, 2014, pp. 2672–2680.
[2] T. C. Wang, M. Y. Liu, J. Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018, pp. 8798–8807.
[3] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in 2017 IEEE International

Table 1: FID Comparison between Different Methods (lower is better)

Methods	FID (Entire Image)	FID (Face Image)
Histogram Matching [31]	400.69	329.01
MaskGAN [4]	251.48	228.35
FaceSwap [16]	207.20	209.14
Ebsynth [32]	173.02	183.49
CycleGAN [3]	183.88	180.48
pix2pixHD [2]	252.75	211.59
Ours	200.44	183.06

Conference on Computer Vision (ICCV), Oct 2017, pp. 2242–2251.
[4] C. H. Lee, Z. Liu, L. Wu, and P. Luo, "MaskGAN: Towards diverse and interactive facial image manipulation," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
[5] M. Johnson, G. Brostow, J. Shotton, O. Arandjelovic, V. Kwatra, and R. Cipolla., "Semantic photo synthesis," Computer Graphics Forum, vol. 25, no. 3, pp. 407–413, 2006.
[6] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2photo: Internet image montage," ACM Transactions on Graphics, vol. 28, no. 5, p. 110, Dec. 2009.
[7] M. Elad and P. Milanfar, "Style transfer via texture synthesis," IEEE Transactions on Image Processing, vol. 26, no. 5, pp. 2338–2351, 2017.
[8] M. Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in Proceedings of the 31th International Conference on Neural Information Processing Systems, ser. NIPS'17, 2017, pp. 700–708.
[9] K. Li, T. Zhang, and J. Malik, "Diverse image synthesis from semantic layouts via conditional imle," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 4219–4228.
[10] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5967– 5976.
[11] T. Shi, Y. Yuan, C. Fan, Z. Zou, Z. Shi, and Y. Liu, "Face-to-parameter translation for game character auto-creation," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 161–170.
[12] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-toimage translation," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018, pp. 8789–8797.
[13] T. Li, R. Qian, C. Dong, S. Liu, Q. Yan, W. Zhu, and L. Lin, "BeautyGAN: Instance-level facial makeup transfer with deep generative adversarial network," in Proceedings of the 26th ACM International Conference on Multimedia, ser. MM '18, New York, NY, USA, 2018, pp. 645–653.
[14] W. Jiang, S. Liu, C. Gao, J. Cao, R. He, J. Feng, and S. Yan, "PSGAN: Pose and expression robust spatial-aware gan for customizable makeup transfer," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020, pp. 5193-5201.
[15] P. Korshunov and S. Marcel, "Deepfakes: a new threat to face recognition? assessment and detection," ArXiv, vol. abs/1812.08685, 2018.
[16] M. Kowalski. Faceswap. [Online]. Available: https://github.com/MarekKowalski/FaceSwap
[17] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, ser. SIGGRAPH '99, 1999, pp. 187–194.
[18] P. Zhu, R. Abdal, Y. Qin, and P. Wonka, "Sean: Image synthesis with semantic region-adaptive normalization," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020, pp. 5103-5112.
[19] T. Beeler, F. Hahn, D. Bradley, B. Bickel, P. Beardsley, C. Gotsman, R. W. Sumner, and M. Gross, "High-quality passive facial performance capture using anchor frames," ACM Transactions on Graphics, vol. 30, no. 4, 2011.
[20] T. Beeler, B. Bickel, G. Noris, S. Marschner, P. Beardsley, R. W. Sumner, and M. Gross, "Coupled 3d reconstruction of sparse facial hair and skin," ACM Transactions on Graphics, vol. 31, no. 4, 2012.
[21] E. Richardson, M. Sela, R. Or-El, and R. Kimmel, "Learning detailed face reconstruction from a single image," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5553–5562.
[22] A. Chen, Z. Chen, G. Zhang, K. Mitchell, and J. Yu, "Photo-realistic facial details synthesis from single image," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 9428–9438.
[23] F. Liu, R. Zhu, D. Zeng, Q. Zhao, and X. Liu, "Disentangling features in 3d face shapes for joint face reconstruction and recognition," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 5216–5225.
[24] Y. Guo, juyong Zhang, J. Cai, B. Jiang, and J. Zheng, "Cnn-based real-time dense face reconstruction with inverse-rendered photo-realistic face images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 6, pp. 1294–1307, 2019.
[25] B. Gecer, S. Ploumpis, I. Kotsia, and S. Zafeiriou, "Ganfit: Generative adversarial network fitting for high fidelity 3d face reconstruction," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 1155–1164.
[26] A. Tewari, M. Zollhfer, F. Bernard, P. Garrido, H. Kim, P. Perez, and C. Theobalt, "High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 2, pp. 357–370, 2020.
[27] C. Cao, H. Wu, Y. Weng, T. Shao, and K. Zhou, "Real-time facial animation with image-based dynamic avatars," ACM Transactions on Graphics, vol. 35, no. 4, p. 110, Dec. 2016.
[28] L. Hu, S. Saito, L. Wei, K. Nagano, J. Seo, J. Fursund, I. Sadeghi, C. Sun, Y.-C. Chen, and H. Li, "Avatar digitization from a single image for real-time rendering," ACM Transactions on Graphics, vol. 36, no. 6, Nov. 2017.
[29] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention (MICCAI), vol. 9351, 2015, pp. 234–241.
[30] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," IEEE Computer Graphics and Applications, vol. 21, no. 5, pp. 34–41, 2001.
[31] D. Coltuc, P. Bolon, and J.-M. Chassery, "Exact histogram specification," IEEE Transactions on Image Processing, vol. 15, no. 5, pp. 1143–1152, May 2006.
[32] O. Jamriska, "Ebsynth: Fast example-based image synthesis and style transfer," https://github.com/jamriska/ebsynth, 2018.
[33] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 2017, pp. 6629–6640.