Anime-like Character Face Generation: A Survey

Xuanfei Li *

Department of Architecture, Shanghai University, Shanghai, China

* Corresponding author: fujiwarakd@gmail.com

Abstract. Recently, anime-style modeling has become increasingly popular. Automatic anime character generation can reduce costs and save manpower. There have been numerous related studies that need to be summarized and analyzed. This paper presents an overview of anime character face generation. Firstly, it summarizes the types and methods of 2D face generation for anime characters. Secondly, we introduce methods for anime character face modeling, including deep learning-based approaches and other approaches. Finally, we summarize the challenges of anime character face generation and evaluate their performance.

Keywords: deep learning, generative adversarial networks, style transfer, 3D face generation.

1. Introduction

In recent years, with the development of computer animation technology, 3D games, animation, film industry has been paid more and more attention. Three-dimensional character modeling holds significant importance as a research direction. At the same time, with the spread of Japanese animation culture in the world, anime-like character is gradually favored by people. "Anime-like" is a term that refers to a style that is not realistic and has the characteristics of Japanese anime. This style is different from verisimilitude and has its own unique charm. Anime-like 3d character models are often used in games, animations, and also as virtual avatars, such as Virtual youtubers, who are active on video platforms in recent years. And face plays a very important role in character modeling.

However, manually building 3D face models is a tedious and time-consuming task, so automatic generation of 3D models is an important research direction. Deep learning models have been used in industries such as image and video compositing, as well as animation and film production. In recent years, generative adversarial networks (GANs) have been extensively researched and used to solve a number of problems in the field of computer vision. Anime-like 3D face models generation consists of 2D face generation and 3D models generation. GAN is very powerful in image translation and 3D face generation.

The rest of the paper is organized as follows: Section 2 discusses 2D face generation, encompassing topics such as image-to-image generation and stylized 2D face generation. Section 3 focuses on 3D face generation, exploring deep learning-based approaches and other approaches. Section 4 provides a detailed discussion, while Section 5 concludes the paper.

2. 2D Face Generation

The 2D face generation methods consist of image hyper-resolution, image composition and image to image translation. The image super-resolution method includes generating high-level face images. Image composition consists of face image generation that refers to the synthesis of face images from different angles based on a face image. And image-to-image translation is the focus of this section.

2.1. Image-to-image Translation

Image-to-image translation involves taking a face image from one domain and transforming it into the image style of another domain.

Pix2pix (2018) [1] has studied the solution of conditional adversarial networks to the image translation problem. CGANs are well-suited for image-to-image translation tasks due to their ability to learn conditional generative models of data. StarGAN (2018) [2] is a generative adversarial

network that uses only a single generator and a discriminator to learn the mappings among multiple domains, efficiently trained from images of all domains. It has the capability to perform facial attribute transfer and facial expression synthesis tasks. CycleGAN (2020) [3] is presented to capture the particular features of one image collection and to find out how these features translate to another collection of images, all without any paired training examples. Kim et al. proposed U-GAT-IT (2020) [4], which combines an attention module and a normalization function that allows for end-to-end learning. It has superiority on style transfer and object transfiguration. PSGAN (2020) [5] can transfer makeup to and from real-world apps, and it consists of a makeup distillation network, a fine-grained makeup deformation module, and a makeup application network.

2.2. Stylized 2D Face Generation

Stylization is a different art form from realism, and anime style is one of them. There are many papers on cartoon style face translation. CartoonGAN (2018) [6] is a GAN-based methodology designed for the task of photo cartoonization. It employs unpaired sets of images during the training process, enabling the network to learn a mapping from real-world photographs to corresponding cartoon-style images. Compared to CycleGAN, CartoonGAN takes about half the training time. Zheng et al. (2019) [7] proposes a method to translate natural human faces from the source photo domain to the target cartoon domain based on GAN. Their method shows performance gains over state-of-the-art image-to-image translation methods, particularly in terms of exaggeration and abstraction. But their models also fail in some cases and it is still difficult to generate small details, so that the generated cartoons are not detailed enough. WarpGAN (2019) [8] solved the problem of style conversion and face distortion in a joint learning framework. The caricatures synthesized by WarpGAN are enjoyable to look at and also photorealistic. StyleCariGAN (2021) [9] is a comic generation framework based on shape and style manipulation to generate realistic and detailed comics. The key is the shape exaggeration block, which is used to modulate StyleGAN's coarse layer feature mapping to produce the desired comic shape exaggeration. CariGAN (2022) [10] model can address a weakly paired image-to-image translation task based on conditional generative adversarial networks.



Figure 1. StyleCariGAN [9]

But there is a considerable difference between the manga style and the anime style, although both are special effects styles. Traditional image-to-image translation methods can also be used for anime style translations, but the results are often unsatisfactory. In recent years, with the global spread of anime, there have been a number of papers on the translation of anime styles. Yamakawa et al. (2018) [11] explored a method that permits users to deform the projected image of a 3D facial model using intuitive feature space axis parameters that represent the style used to draw the face for animation. AnimeGAN (2020) [12] is a lightweight generative adversarial network to achieve fast animation style transfer. AnimeGAN (2020) [12] is a lightweight generative adversarial network (GAN) developed specifically for fast animation style transfer. Its efficient design allows for the rapid

generation of high-quality anime-style images. DENG et al. proposed an asymmetric structure CycleGAN (2022) [13] that employs an attention mechanism and semantic-level information enhancement to solve the low-resource image-to-image translation problem. Their model performs well in face style translation. AniGAN (2022) [14] is a GAN-based model that combines color/texture style transfer and the transformation of local facial shape from a source photo face to an anime-like shape, drawing inspiration from a reference animated face's style. It carefully preserves the overall structure of the source photo face. Moreover, the model incorporates a dedicated discriminator that enhances the generation of top-notch animated faces by learning domain-specific distributions. This discriminator also plays a crucial role in mitigating significant distortions in the generated faces by learning shared distributions across both animated and photo face domains.



Figure 2. AniGAN [14]

In addition to this, there are a number of papers on research into the automatic generation of anime-like faces. JIN et al. (2017) [15] explored the automatic generation of anime character avatars. They built a DRAGAN-based model and collected a dataset. The model was able to generate anime character avatars automatically. Style-gender based anime GAN (SGA-GAN) (2019) [16] is proposed to generate diversified anime faces, in which gender is taken as the direct condition to ensure gender differentiation, and style characteristics as the condition to ensure style diversity. They also estimate image quality by fine-tuning a network of style features to prevent unclear or distorted face images.



Figure 3. Towards the Automatic Anime Characters Creation with Generative Adversarial Networks [15]

Besides, some paper is about anime line art colorization. Zhang et al. (2017) [17] proposed a technique that facilitates the transfer of an illustration's color style to a sketch. They accomplished this by combining the residual U-net architecture with an auxiliary classifier generative adversarial network (AC-GAN). This combined approach allows for the successful transfer of the color style from an illustration to grayscale sketches. Ci et al. (2018) [18] introduced a deep conditional adversarial architecture specifically designed for coloring anime line art using a graffiti-based approach. Their methodology involved the integration of a conditional framework with the WGAN-GP standard, enhancing it further with perceptual loss to achieve more natural and realistic synthetic images.



Figure 4. Style Transfer for Anime Sketches with Enhanced Residual U-net and Auxiliary Classifier GAN [17]

3. 3D Face Generation

3.1. Deep Learning-based Approaches

CNN and GAN are both deep learning models. Convolutional neural networks (CNNs) have powerful performance in image classification, target detection and image generation. Face generation has witnessed remarkable advancements with the recent application of GANs and adversarial training

processes. There have been a number of studies on the reconstruction of 3D object. Reconstructing a 3D face from a single 2D image is a challenging task. GANFIT (2019) [19] is a method based on GAN and DCNNs for reconstructing facial volumes as well as textures from a single image. MeshGAN (2019) [20] has the capability to produce diverse 3D face meshes exhibiting distinct identities and expressions. Compared to methods based on volume or point cloud representation, MeshGAN is capable of generating meshes with a higher level of details. 3DFaceGAN (2020) [21] is an adversarial network for 3D face generation and translation. They introduce facial data and are able to make 3D facial models more realistic in detail. Lattas et al. (2020) [22] introduced a method to generate high-quality renderable facial reconstructions from arbitrary natural images. They compiled a substantial dataset of facial shapes and utilized these reconstructions to create cutting-edge 3D textures. To achieve realistic rendering of human skin, they simulated diffuse reflectance and highlight albedo for the desired geometry as well as normals. Their approach enables the production of realistic facial renderings with accurate lighting and surface properties.

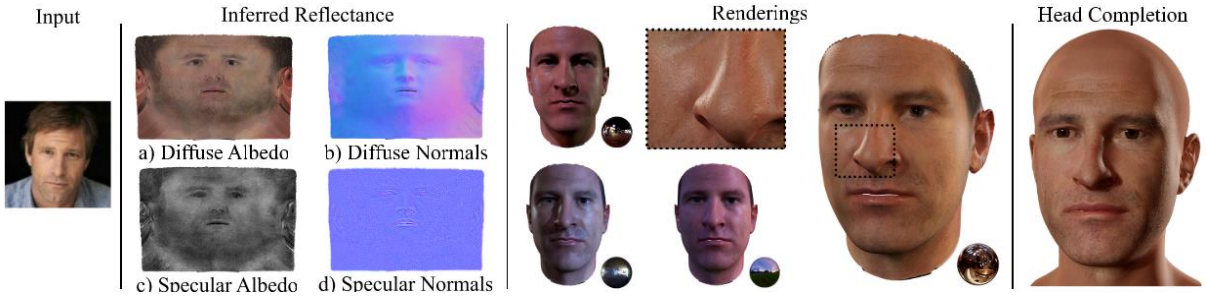


Figure 5. AvatarMe [22]

Generating stylized face models using deep learning models is also an important task. FaceWarehouse (2014) [23] is a comprehensive database of 3D facial expressions specifically designed for visual computing applications. The database comprises facial expression data from 150 individuals, with each person having 47 distinct facial expressions. This resource provides a valuable collection for researchers and developers working on various tasks related to facial expression analysis, computer vision, and related fields. DeepSketch2Face (2017) [24] is a system that performs 3D face as well as caricature face generation. They explored the utilization of a deep regression network based on Convolutional Neural Networks (CNN) to infer 3D face models from 2D sketches. MMfitting introduces errors that are 3 times larger than those of this method. Jackson et al. (2017) [25] proposed a new volumetric representation using 3D facial geometry and a suitable convolutional neural network (CNN) architecture approach. The CNNs are trained to generate the corresponding 3D bodies directly from 2D facial images. The technique aims to take advantage of the volumetric representation and deep learning methods to improve the accuracy and efficiency of 3D face reconstruction. Wu et al. (2018) [26] introduced an algorithm specifically designed for the generation of 3D comic models from 2D comic images. Their algorithm enables the production of 3D comics that exhibit similar exaggeration and stylization as their 2D counterparts.

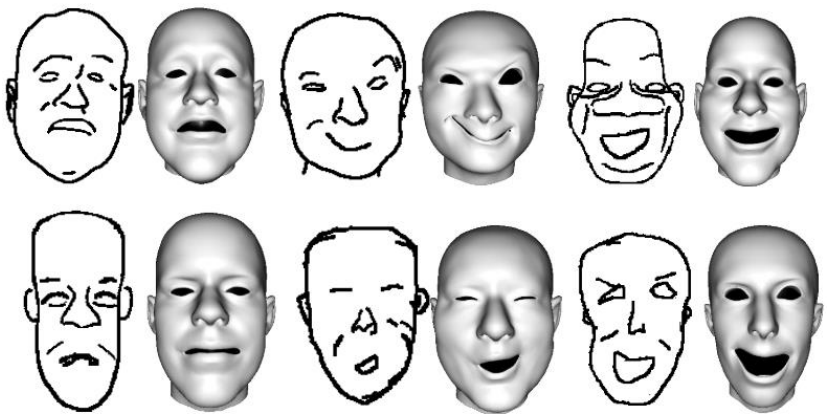


Figure 6. DeepSketch2Face [24]

Deep learning models can also be used for the automatic generation of game-related models. Chai et al. (2016) [27] presented a fully automated method for modelling single-image 3D hair based on deep convolutional neural networks. Their method is capable of generating high-quality hair models in a moderate amount of time and is suitable for processing web images. Face-to-Parameter (F2P) (2019) [28] is a method specifically developed for the automatic creation of in-game characters using an input face photo. It operates by translating the facial features from the photo into parameterized representations that can be used to generate character models. F2P aims to streamline the process of character creation by automating the transformation from a real-world face to a parameterized representation suitable for in-game use. Li et al. (2020) [29] proposed a method that allows automatic generation of natural NPCs by combining character customization with neural networks.

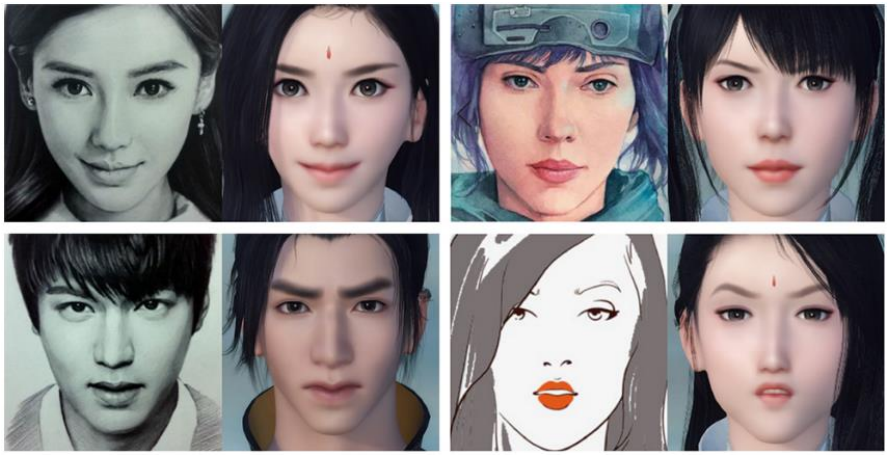


Figure 7. Face-to-Parameter [28]

3.2. Other Approaches

Kondo et al. (1997) [30] introduced a method for generating facial caricatures based on a straightforward approach using corresponding 3D facial images. Fujiwara et al. (2001) [31,32] proposed a method for preparing an average face by automatically extracting feature points of 3D face image, generation of triangular blocks and caricature deformation. It models a 3D face by measuring the color texture and surface shape information of the face simultaneously from a 2D image of the face captured by a surrounding 2D camera. Clarke et al. (2011) [33] presented an approach for the automated generation of digital caricatures from facial photographs, while retaining the characteristic artistic distortion style observed in hand-drawn caricatures. Their approach involves the use of a deformable head model, enabling the conversion of facial photographs into computer-generated caricatures. It allows for the conversion of facial photographs into computer-generated caricatures through a deformable head model. Hara et al. (2012) [34] proposed a fast automatic 3D face reconstruction method. By structuring-motion and deforming the facial model, using their method, a 3D facial model can be created in 5.8 seconds by simply shaking the head freely in front of a single camera. Chalás et al. (2017) [35] proposed a method for automatic face generation, which is synthesized after scanning real human faces. Their method takes as input a grid of real people's faces stored in the fidentis database or a model created in a certain format, and then segments the model and uses its individual facial parts to create artificial facial models automatically or semi-automatically. Niki et al. (2019) [36] presented a method for generating anime-like 3D face models from a single input image. Their approach involved applying Principal Component Analysis (PCA) to an existing anime-style 3D model, which allowed them to derive a natural 3D base model. This base model served as the foundation for further modifications to achieve a final 3D model that closely resembled the source image. By leveraging the principles of PCA and custom adjustments, their method enabled the creation of anime-like 3D face models that closely matched the appearance of the input image. Chen et al. (2020) [37] proposed an automatic stylized modeling system based on

actual human faces. The 3D face is first scanned and the system can automatically generate a stylized face model.

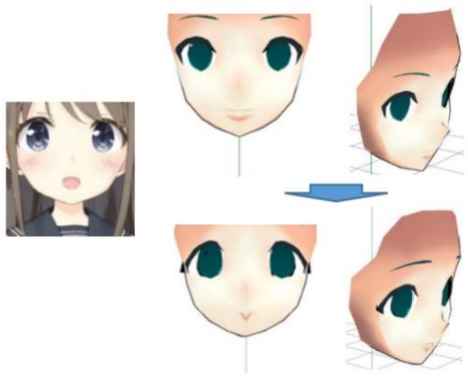


Figure 8. Semi-Automatic Creation of an Anime-Like 3D Face Model from a Single Illustration [36]

4. Discussion

Table 1. studies related to 2D face generation

REFERENCE	FOCUS	ADVANTAGES
Style Transfer for Anime Sketches with Enhanced Residual U-net and Auxiliary Classifier GAN	Apply an illustration's color style to a sketch	Integrated residual U-net with AC-GAN
User-Guided Deep Anime Line Art Colorization with Conditional Adversarial Networks	A deep conditional adversarial architecture for scribble-based anime line art colorization	Integrate the conditional framework with WGAN-GP criteria
Image-to-image translation with conditional adversarial networks	Conditional adversarial networks for the image-to-image translation	Can generate a corresponding output image by conditioning an input image
Generating Anime-like Face Images from Projected 3D Model	Deform 3D facial model projection images with intuitive parameters	Allow users to deform 3D facial model projection images with intuitive parameters
Unpaired photo-to-caricature translation on faces in the wild	Translate faces in the wild from the source photo domain to the target caricature domain	Show performance gains over state-of-the-art image-to-image translation methods
StyleCariGAN	a caricature generation framework based on shape and style manipulation	Can generate realistically detailed comics
WarpGAN	Addresses both style transfer and face deformation	The caricatures synthesized by WarpGAN are enjoyable and photorealistic
CartoonGAN	A GAN-based approach to photo cartoonization	CartoonGAN takes about half the training time compared to CycleGAN
AnimeGAN	A lightweight generative adversarial network to achieve fast animation style transfer	Can quickly generate higher-quality anime style images
CycleGAN	Solve the low-resource image-to-image translation problem	Perform well in face style translation
U-GAT-IT	Integrate a new attention module and a new learnable normalization function	It has superiority on style transfer and object transfiguration
StarGAN	Perform facial attribute transfer and facial expression synthesis tasks.	Uses only a single generator and a discriminator
CariGAN	Can address a weakly paired image-to-image translation task	Produce the best all-around perceptual performance in terms of user rating scores
AniGAN	transfer color/texture styles and transform the local facial shapes	Can synthesize high-quality anime-faces and avoid noticeable distortions

This article presents a review of methods for generating anime-like character faces. Table 1 summarizes the studies related to 2D face generation in this survey. We discuss the focus as well as the advantages of face generation related research. Due to the superiority of GAN in the field of image synthesis and image translation, recent research on 2D face generation has been based on GAN. where Pix2Pix is almost the most significant part of the solution to the image translation problem. The remaining methods all have different focuses, for example, StarGAN focuses on facial expression transfer, PSGAN focuses on makeup transfer, etc. Based on these methods, there are also many studies on stylized face generation. methods such as WarpGAN, StyleCariGAN, and CariGAN are for comic style face generation, while methods such as AnimeGAN and AniGAN focus on anime style face generation. These targeted methods have stronger performance compared to traditional image translation methods.

Figure 9 summarizes the methods of 3D face generation in a timeline manner. Traditional methods generally use facial scanning and facial feature point methods for model generation. Deep learning-based methods generate more detailed face models and take less time compared to traditional methods. However, for the time being, the traditional methods may be able to have higher reproduction for the task of generating 3D face models from anime style images. Deep learning-based methods are currently under-researched for anime style 3D model generation.

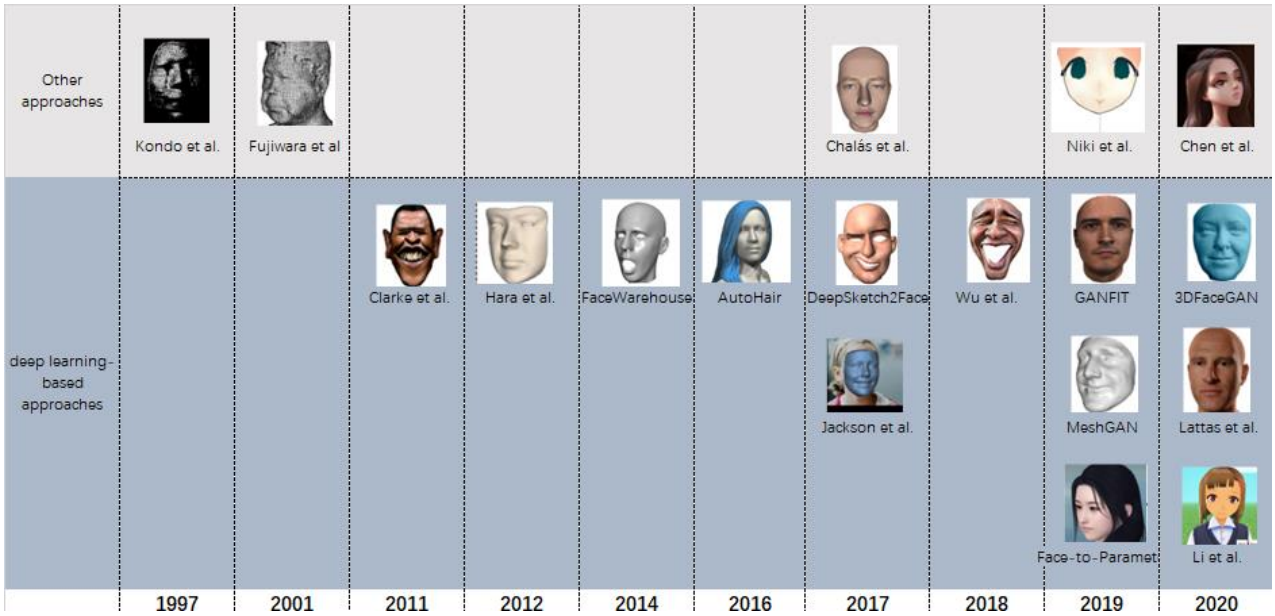


Figure 9. studies related to D face generation

Although a number of anime-based character generation methods have been proposed, there are still a lot of challenges. Firstly, the quality of the generated anime style face images is inconsistent and the quality may be difficult to achieve as required by the industry. Secondly, it remains a challenge to generate anime characters with specific attributes or elements. In addition to this, despite a number of studies on cartoon style 3d character models generation, there are still very few methods on anime style 3d character models generation. In the future, a database of anime style characters may be needed to help the research of anime style character generation.

5. Conclusion

This paper has provided an overview of the current methodologies employed in the generation of anime-like characters. Due to the progress of deep learning models, face images and 3D face models generation have become easier. And the popularity of Japanese anime has brought the anime style into the limelight. But there are still a number of challenges. Future research could place the technology into industries such as computer games, VR and AR. Specifically, the technology can be

used to generate anime style face models for use in these areas, or to generate exclusive anime style avatars for users.

References

[1] Isola, P., et al., Image-to-Image Translation with Conditional Adversarial Networks. arXiv pre-print server, 2018.

[2] Choi, Y., et al., StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. arXiv pre-print server, 2018.

[3] Zhu, J.-Y., et al., Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. arXiv pre-print server, 2020.

[4] Kim, J., et al., U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation. arXiv pre-print server, 2020.

[5] Jiang, W., et al. PSGAN: Pose and Expression Robust Spatial-Aware GAN for Customizable Makeup Transfer. in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.

[6] Chen, Y., Y.K. Lai, and Y.J. Liu. CartoonGAN: Generative Adversarial Networks for Photo Cartoonization. in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018.

[7] Zheng, Z., et al., Unpaired photo-to-caricature translation on faces in the wild. Neurocomputing, 2019. 355: p. 71 - 81.

[8] Shi, Y., D. Deb, and A.K. Jain. WarpGAN: Automatic Caricature Generation. in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019.

[9] Jang, W., et al., StyleCariGAN. ACM Transactions on Graphics, 2021. 40 (4): p. 1 - 16.

[10] Li, W., et al., CariGAN: Caricature generation through weakly paired adversarial learning. Neural Netw, 2020. 132: p. 66-74.

[11] Yamakawa, K. and S. Saito, Generating anime-like face images from projected 3D models, in SIGGRAPH Asia 2018 Posters. 2018, Association for Computing Machinery: Tokyo, Japan. p. Article 6.

[12] Chen, J., G. Liu, and X. Chen, AnimeGAN: A Novel Lightweight GAN for Photo Animation. 2020, Springer Singapore. p. 242 - 256.

[13] Deng, S., K. Uchida, and Z. Yin, Cross-modal and Semantics-Augmented Asymmetric CycleGAN for Data-Imbalanced Anime Style Face Translation, in Proceedings of the 2021 3rd International Conference on Video, Signal and Image Processing. 2022, Association for Computing Machinery: Wuhan, China. p. 43 – 51.

[14] Li, B., et al., AniGAN: Style-Guided Generative Adversarial Networks for Unsupervised Anime Face Generation. IEEE Transactions on Multimedia, 2022. 24: p. 4077 - 4091.

[15] Jin, Y., et al., Towards the Automatic Anime Characters Creation with Generative Adversarial Networks. arXiv pre-print server, 2017.

[16] Li, H. and T. Han. Towards Diverse Anime Face Generation: Active Label Completion and Style Feature Network. in Eurographics. 2019.

[17] Zhang, L., et al., Style Transfer for Anime Sketches with Enhanced Residual U-net and Auxiliary Classifier GAN, in 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR). 2017. p. 506 - 511.

[18] Ci, Y., et al., User-Guided Deep Anime Line Art Colorization with Conditional Adversarial Networks, in Proceedings of the 26th ACM international conference on Multimedia. 2018, Association for Computing Machinery: Seoul, Republic of Korea. p. 1536 – 1544.

[19] Gecer, B., et al. GANFIT: Generative Adversarial Network Fitting for High Fidelity 3D Face Reconstruction. in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019.

[20] Cheng, S., et al., MeshGAN: Non-linear 3D Morphable Models of Faces. arXiv pre-print server, 2019.

[21] Moschoglou, S., et al., 3DFaceGAN: Adversarial Nets for 3D Face Representation, Generation, and Translation. International Journal of Computer Vision, 2020. 128 (10-11): p. 2534 - 2551.

[22] Lattas, A., et al., Avatar Me: Realistically Render able 3D Facial Reconstruction "in-the-wild". arXiv pre-print server, 2020.

[23] Cao, C., et al., Face Warehouse: A 3D Facial Expression Database for Visual Computing. IEEE Transactions on Visualization and Computer Graphics, 2014. 20 (3): p. 413 - 425.

[24] Han, X., C. Gao, and Y. Yu, DeepSketch2Face: a deep learning-based sketching system for 3D face and caricature modeling. ACM Trans. Graph., 2017. 36 (4): p. Article 126.

[25] Jackson, A.S., et al. Large Pose 3D Face Reconstruction from a Single Image via Direct Volumetric CNN Regression. in 2017 IEEE International Conference on Computer Vision (ICCV). 2017.

[26] Wu, Q., et al. Alive Caricature from 2D to 3D. in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018.

[27] Chai, M., et al., AutoHair: fully automatic hair modeling from a single image. ACM Trans. Graph., 2016. 35(4): p. Article 116.

[28] Tianyang, et al., Face-to-Parameter Translation for Game Character Auto-Creation. arXiv pre-print server, 2019.

[29] Li, R., M. Nakayama, and I. Fujishiro. Automatic Generation of 3D Natural Anime-like Non-Player Characters with Machine Learning. in 2020 International Conference on Cyberworlds (CW). 2020.

[30] Kondo, T., K. Murakami, and H. Koshimizu. From coarse to fine correspondence of 3-D facial images and its application to 3-D facial caricaturing. in Proceedings. International Conference on Recent Advances in 3-D Digital Imaging and Modeling (Cat. No.97TB100134). 1997.

[31] Fujiwara, T., et al. A method for 3D face modeling and caricatured figure generation. in Proceedings. IEEE International Conference on Multimedia and Expo. 2002.

[32] Fujiwara, T., et al. 3D modeling system of human face and full 3D facial caricaturing. in Proceedings Seventh International Conference on Virtual Systems and Multimedia. 2001.

[33] Clarke, L., M. Chen, and B. Mora, Automatic Generation of 3D Caricatures Based on Artistic Deformation Styles. IEEE Transactions on Visualization and Computer Graphics, 2011. 17(6): p. 808-821.

[34] Hara, T., et al., Fast-automatic 3D face generation using a single video camera, in ACM SIGGRAPH 2012 Posters. 2012, Association for Computing Machinery: Los Angeles, California. p. Article 91.

[35] Chalás, I., et al., Generating various composite human faces from real 3D facial images. The Visual Computer, 2017. 33 (4): p. 443 - 458.

[36] Niki, T. and T. Komuro. Semi-Automatic Creation of an Anime-Like 3D Face Model from a Single Illustration. in 2019 International Conference on Cyberworlds (CW). 2019.

[37] Chen, B. and T. Komma. Face Stylized Modeling for Virtual Character. in 2020 Nicograph International (NicoInt). 2020.