IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 27, 2020, accepted March 11, 2020, date of publication March 20, 2020, date of current version April 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982224

A State-of-the-Art Review on Image Synthesis With Generative Adversarial Networks

LEI WANG[1,2], WEI CHEN[1,2,3], (Member, IEEE), WENJIA YANG[1,2], FANGMING BI[1,2], AND FEI RICHARD YU[4], (Fellow, IEEE)

[1]School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China
[2]Mine Digitization Engineering Research Center of the Ministry of Education, China University of Mining and Technology, Xuzhou 221116, China
[3]Information Engineering College, Beijing Institute of Petrochemical Technology, Beijing 102617, China
[4]School of Information Technology, Carleton University, Ottawa, ON K1S 5B6, Canada

Corresponding author: Wei Chen (chenwdavior@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51874300 and Grant 51874299, in part by the National Natural Science Foundation of China and Shanxi Provincial People's Government Jointly Funded Project of China for Coal Base and Low Carbon under Grant U1510115, and in part by the Open Research Fund of Key Laboratory of Wireless Sensor Network and Communication, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, under Grant 20190902 and Grant 20190913.

ABSTRACT Generative Adversarial Networks (GANs) have achieved impressive results in various image synthesis tasks, and are becoming a hot topic in computer vision research because of the impressive performance they achieved in various applications. In this paper, we introduce the recent research on GANs in the field of image processing, including image synthesis, image generation, image semantic editing, image-to-image translation, image super-resolution, image inpainting, and cartoon generation. We analyze and summarize the methods used in these applications which have improved the generated results. Then, we discuss the challenges faced by GANs and introduce some methods to deal with these problems. We also preview some likely future research directions in the field of GANs, such as video generation, facial animation synthesis and 3D face reconstruction. The purpose of this review is to provide insights into the research on GANs and to present the various applications based on GANs in different scenarios.

INDEX TERMS Generative adversarial networks, image synthesis, image-to-image translation, image editing, cartoon generation.

I. INTRODUCTION

Artificial intelligence (AI) has aroused widespread interest in both the press and on social media. Especially with the rapid development of deep learning, image processing has made great progress. An enormous amount of images are applied in social media, which make the generative models became a hot topic in deep learning research.

Some generative models are promising unsupervised learning techniques with powerful semantic information representation capabilities, and are attracting more and more attention. Among them, the Variational Auto-encoder (VAE) [1] cannot generate clear enough images. The Glow [2] is a flow-based generation model, which has not been widely used so far. The Generative Adversarial Networks (GANs) have achieved impressive results in image processing, and are attracting growing interests in the academic and industrial fields.

The associate editor coordinating the review of this manuscript and approving it for publication was Hongjun Su.

Nowadays, GANs are applied to various research and applications, such as image generation [3], image inpainting [4], text generation [5], medical image processing [6]–[13], semantic segmentation [14]–[17], image colorization [18], [19], image-to-image translation [20], and art generation [21]. Besides, GANs are widely used in face synthesis and face editing, such as face age [22]–[24] and gender translation [25].

The research of GANs is divided into two directions: 1) Theoretical research on GANs based on information theory or energy-based models, and focus on the unsolved problems of GANs during training, such as mode collapse, unstable training and hard to evaluate. We will briefly discuss this aspect of problems and the challenges of GANs in Section IX. 2) The applications of GANs in various computer vision tasks. Although there are still some unresolved problems, various GAN-variants have improved the performance of GANs by numerous research studies. In this work, we mainly focus on the second aspect of current research on GANs.

Although there have been some surveys on GANs so far, like [26], in the field of deep learning, especially GANs are developing fast. This paper focuses on the recent research of GANs in image synthesis. It provides a comparison and analysis in terms of the pros and cons of these applications based on GANs. Besides, we analyze and summarize the methods that have been used in these applications to improve the generated images. Meanwhile, we discuss the challenges faced by GANs in terms of training and evaluating of GANs. Some methods for stable training and evaluation of GANs are provided. Then, we discuss the likely future research directions, such as video generation, facial animation synthesis, and 3D face reconstruction. The rest of the paper is organized as follows: Section II gives a brief introduction of GANs. Section III introduces some applications of image synthesis based on GANs. Section IV focuses on the supervised and unsupervised methods for image-to-image translation. Section V discusses several methods in the application of image editing. Section VI describes several methods of cartoon generation. Section VII reviews the current challenges and limitations of GAN-based methods, as well as previews likely future research work in the area of GANs. Conclusions are given in Section VIII.

II. GENERATIVE ADVERSARIAL NETWORKS

GANs are especially successful in image tasks due to the great potential in image processing. They are considered to be the most effective method in the task of image generation and play an important role in various applications.

The Generative Adversarial Network (GAN) is a model that has been prevailing since Goodfellow *et al.* [27] proposed it in 2014. GAN consists of a generator G and a discriminator D, the general structure of a Generative Adversarial Network is illustrated in Fig. 1.

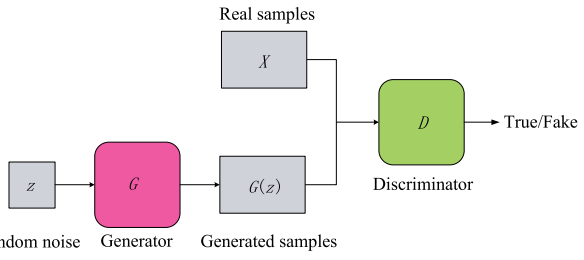

FIGURE 1. The general structure of a generative adversarial network.

The generator G is used to generate realistic samples from random noise and tries to fool the discriminator D. The discriminator D is used to identify whether the sample is real or generated by the generator G. The generator and the discriminator are competing with each other until the discriminator cannot distinguish between real and fake generated images. The whole process can be regarded as a two-player minimax game where the main aim of GAN training is to achieve the Nash equilibrium [28]. The loss function of the GAN is formulated as follows:

$$\min_G \max_D V(D, G) = \mathrm{E}_{x\sim Pdata(x)}[\log D(x)] + \mathrm{E}_{z\sim Pz(z)}[\log(1 - D(G(z)))] \quad (1)$$

where Pdata(x) denotes the true data distribution, Pz(z) denote the noise distribution.

Due to the special network structure and the generation performance of GANs, extensive research has produced numerous applications based on GANs, as shown in Fig. 2.

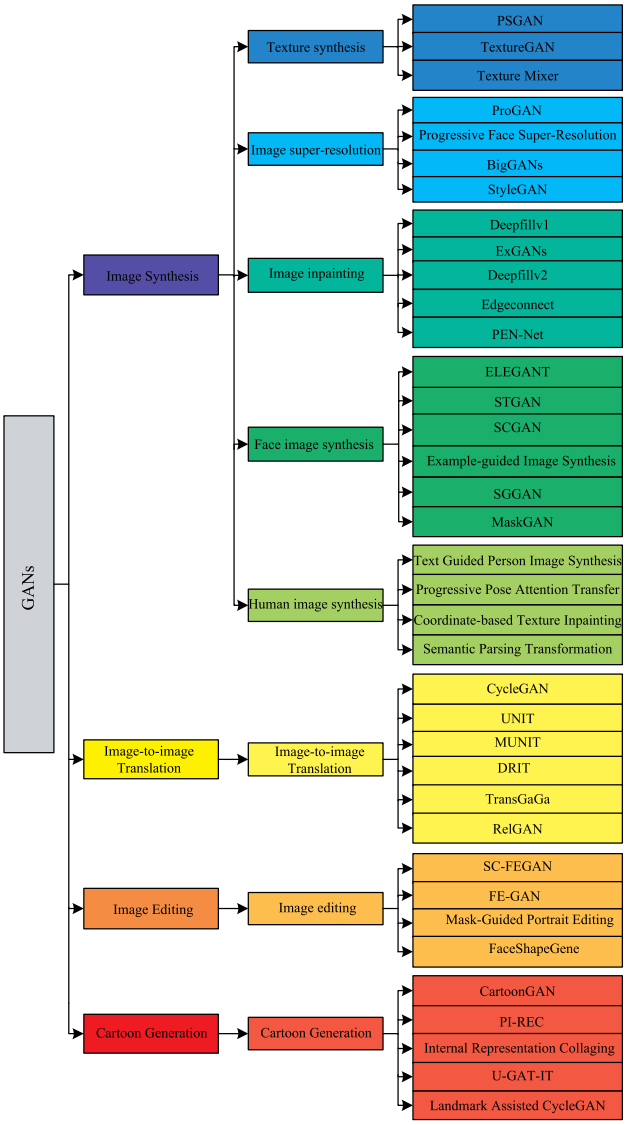

FIGURE 2. Taxonomy of GANs.

III. IMAGE SYNTHESIS

Image synthesis has attracted people's attention because of its wide application in social media. The GANs have achieved excellent results in the field of image synthesis, such as GauGAN [29]. A variety of image synthesis methods have emerged so far.

A. TEXTURE SYNTHESIS

Image synthesis can be divided into fine-grained texture synthesis and coarse-grained texture synthesis. The coarse-grained texture synthesis pays attention to the similarity between the input image and the output image while the fine-grained texture synthesis pursues whether the synthetic texture is similar to the ground truth.

1) PSGAN

Bergmann *et al.* [30] proposed a new method of texture synthesis based on the Generative Adversarial Network called Periodic Spatial GAN (PSGAN). The model of PSGAN is illustrated in Fig. 3.

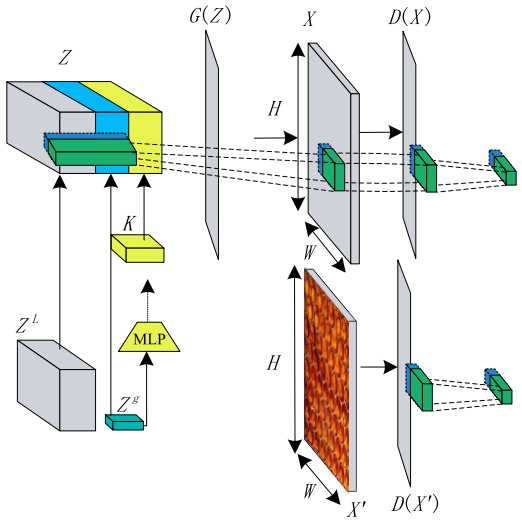

FIGURE 3. Illustration of the PSGAN model [30].

The loss function of the PSGAN is defined as:

$$\min_G \max_D V(D,G) = \frac{1}{LM}\sum_{\lambda=1}^{L}\sum_{\mu=1}^{M} \mathrm{E}_{Z\sim Pz(Z)}[\log(1-D_{\lambda\mu}(G(Z)))] + \frac{1}{LM}\sum_{\lambda=1}^{L}\sum_{\mu=1}^{M} \mathrm{E}_{X'\sim Pdata(X)}[\log D_{\lambda\mu}(X')] \quad (2)$$

PSGAN can learn multiple textures from one or more complex datasets of large images. The method can not only smoothly interpolate between samples in a structured noise space and generate novel samples that are perceptually located between the textures of the original dataset, but also accurately learn periodic textures. PSGAN has the flexibility to handle a wide range of textures and image data sources. It is a method of highly scalable and can produce output images of any size.

2) TextureGAN

Xian *et al.* [31] proposed a texture synthesis method called TextureGAN, which combines sketch, color, and texture to synthesize images that people expect. The training process is shown in Fig. 4 and Fig. 5.

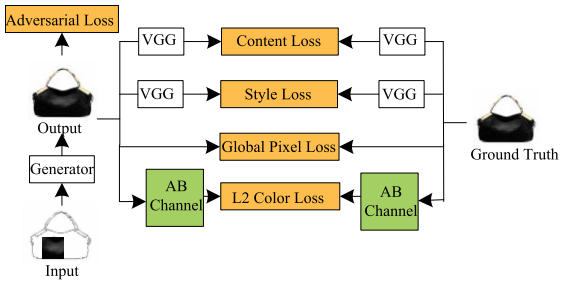

FIGURE 4. TextureGAN pipeline for the ground-truth pre-training [31].

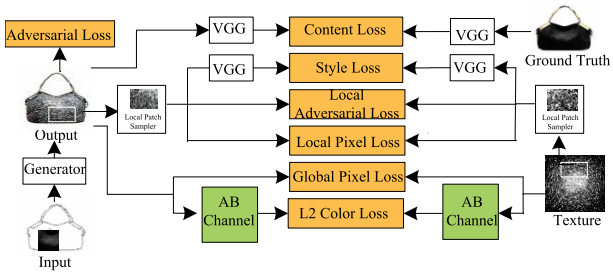

FIGURE 5. TextureGAN pipeline for the external texture fine-tuning [31].

The objective function of ground-truth pre-training is defined as:

$$L = L_F + W_{ADV}L_{ADV} + W_S L_S + W_P L_P + W_C L_C \quad (3)$$

The objective function of external texture fine-tuning is defined as:

$$L = L_F + W_{ADV}L_{ADV} + W_P L'_P + W_C L'_C + L_t \quad (4)$$

TextureGAN is an image synthesis method that can control the texture of generated images. It allows the users to place a texture patch anywhere on the sketch and at any scale to control the desired output texture. Besides, it can not only process various texture inputs and generate texture compositions that follow sketch outlines, but also achieve good results in the sketch and texture-based image synthesis.

3) TEXTURE MIXER

Yu *et al.* [32] proposed a new method that can control texture interpolation called Texture Mixer. The structure of Texture Mixer is shown in Fig. 6.

The training objective is:

$$\min_{E^\ell,E^g,G}\max_{D^{rec},D^{itp}} \mathrm{E}_{S_1,S_2\sim S}(\lambda_1 L^{rec}_{pix} + \lambda_2 L^{rec}_{Gram} + \lambda_3 L^{rec}_{adv} + \lambda_4 L^{itp}_{Gram} + \lambda_5 L^{itp}_{adv}) \quad (5)$$

The method utilizes deep learning and GAN to realize controllable interpolation of textures, which combines two different types of texture patterns and makes the transition natural. It proposes a neural network trained with a reconstruction task and a generation task to project the texture of the sample onto a latent space and project linear interpolation onto the image domain to ensure the quality of the intuitive control and realistic generated results. Furthermore, it is superior to

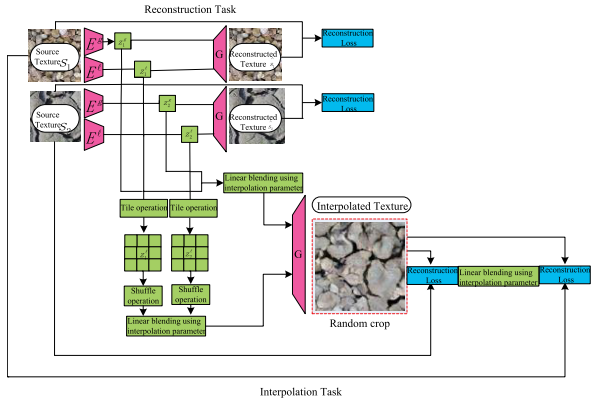

FIGURE 6. A diagram of the texture mixer [32].

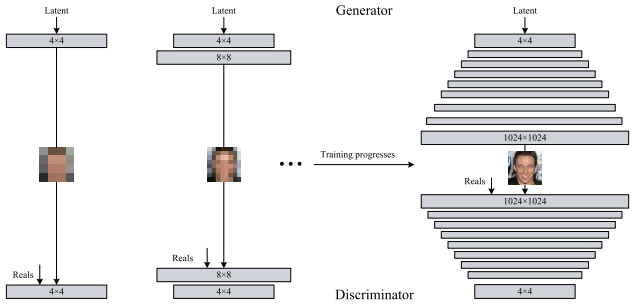

FIGURE 7. The structure of ProGAN [36].

many baseline methods and has a good performance in texture synthesis in the dimensions of controllability, smoothness, and realism.

4) OTHER METHODS

Li and Wand [33] proposed an efficient texture synthesis method called Markovian Generative Adversarial Networks (MGANs). It can not only directly decode brown noise to realistic texture but also decode the photo to the painting, which improves the quality of texture synthesis. Jetchev *et al.* [34] proposed an architecture called spatial GAN (SGAN) which is well-suited for texture synthesis. It is a method that can synthesize texture images with high quality and can fuse multiple different source images in complex textures.

The texture synthesis based on GANs adopts the method of interpolation can produce realistic details of texture and realize the natural transition of texture synthesis. Interpolation and extrapolation are two approaches to enforce constraints for GANs. The incorporation of constraints is built into the training of the GAN while the constraints are enforced after each step through projection on the space of constraints for extrapolation [35]. However, sometimes the texture synthesis model is difficult to converge during training and it can suffer from "mode dropping".

B. IMAGE SUPER-RESOLUTION

The image generation model is designed to explore how to generate a desired image, while producing high-quality large images has always been a challenging task. The ability to produce high-quality and high-resolution images is an important advantage of GANs, and significant progress has been made in generating high-quality and visually realistic images. A series of models based on GANs are emerging in the purpose of producing higher-resolution images.

1) ProGAN

Karras *et al.* [36] proposed an image generation method called ProGAN. The structure of ProGAN is shown in Fig. 7.

The key idea of this approach is to gradually increase the generator and discriminator, which starts from a low resolution and adds new layers as the training progresses to make the model increase fine details. It is a method which can not only speed the training up but also greatly stabilize it. Compared with the earlier works on GANs, the quality of the results using this method is generally high, and the training is stable in high resolution. However, there are some shortcomings in this method. For example, semantic sensitivity and understanding depend on the constraints of the dataset.

2) PROGRESSIVE FACE SUPER-RESOLUTION

Kim *et al.* [37] proposed a novel face super-resolution (SR) method which can generate photo-realistic face images with fully retained facial details. The network architecture is shown in Fig. 8.

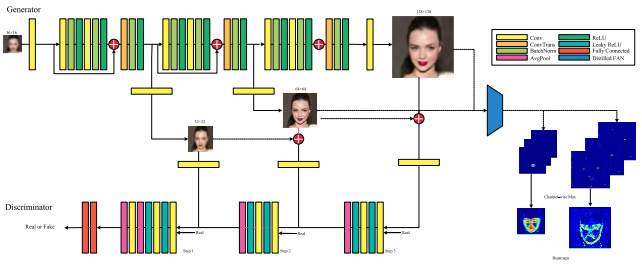

FIGURE 8. The network architecture of [37].

The loss term is shown as:

$$L_{Ours} = \alpha L_{pixel} + \beta L_{feat} + \gamma L_{WANG} \tag{6}$$

$$L_{Ours} = \alpha L_{pixel} + \beta L_{feat} + \gamma L_{WANG} + \lambda L_{heatmap} + \eta L_{attention} \tag{7}$$

The authors use a progressive training approach that allows stable training by dividing the network into successive steps, each step producing an output of progressively higher resolution. A novel facial attention loss has also been proposed and applied at each step to focus on restoring facial attributes in more detail by multiplying pixel differences and heatmap values. They also proposed a compressed version of the face alignment network (FAN) to extract suitable landmark heatmaps for face super-resolution (SR), and the overall

training time can also be reduced. Furthermore, it can learn the restoration of facial details and generate super-resolution facial images that are similar to real ones. The results are superior to the earlier methods in terms of qualitative and quantitative measurements, especially in perceptual quality.

3) BigGANs

Brock *et al.* [38] proposed models called BigGANs, which realized the work of generating high-resolution and diverse images from complex datasets. A typical network architecture of BigGANs is shown in Fig. 9.

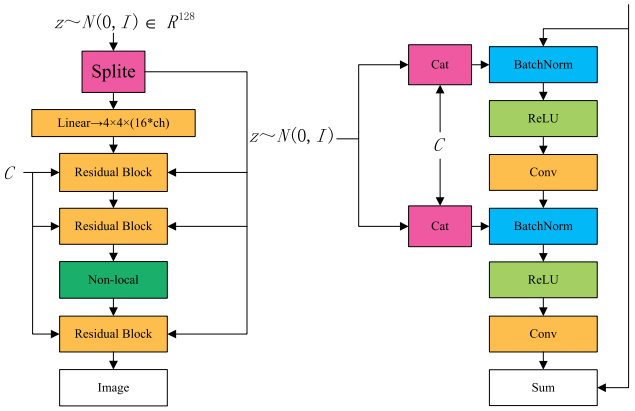

FIGURE 9. A typical network architecture of BigGANs [38].

This method achieves the goal of generating high-resolution and diverse samples from the complex dataset ImageNet successfully. It is the largest scale of Generative Adversarial Networks that have been trained so far and can generate images of unprecedented quality. It is far superior to the earlier methods in terms of the realism of the generated image. The authors applied orthogonal regularization to the generator to handle the specific instability of such scale and truncated the latent space to control the fidelity and variety of generated images.

4) StyleGAN

Karras *et al.* [39] proposed an alternative generator architecture called StyleGAN. The network architecture of StyleGAN is shown in Fig. 10.

The authors redesigned the generator architecture which can adjust its image style based on the latent code in each convolutional layer. It is able to control the entire image synthesis process which starts with very low resolution and generates high-resolution artificial images step by step. Besides, it controls the visual features by modifying the input of each level in the network separately, from coarse features to fine details. The breakthrough of StyleGAN is that it not only produces high-quality and realistic images but also provides better control and understanding of the generated images. The method implements automatic learning, unsupervised high-level attribute separation, and stochastic variation of generated images, which enables intuitive, scale-specific

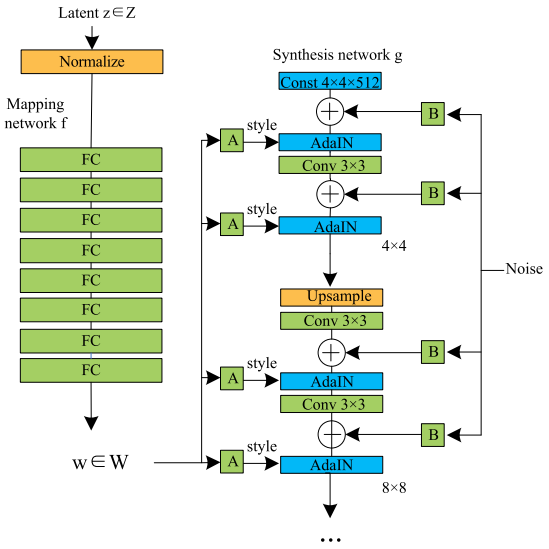

FIGURE 10. The network architecture of StyleGAN [39].

control synthesis of the composition. The method is superior to the traditional GAN generator architecture and can generate a high-resolution image that looks more realistic.

5) OTHER METHODS

Ledig *et al.* [40] proposed a generative adversarial network for image super-resolution (SR) called SRGAN, and can significantly improve the perceptual quality. Wang *et al.* [41] improved SRGAN to derive an Enhanced SRGAN (ESRGAN) which not only improved the problem of artifacts in SRGAN and the visual quality of generated images but also obtained more realistic and natural textures. Wang *et al.* [42] proposed a method to recover natural and realistic texture called SFTGAN, which is equipped with a novel Spatial Feature Transform (SFT) layer and can generate more realistic and visually pleasing textures.

The image super-resolution method which gets the best results trains generators and discriminators from low-resolution images, and adds a higher-resolution network layer each time to generate artificial images step by step. It can generate high-resolution and diverse images with high-quality. However, the training time is long and more GPUs are required.

C. IMAGE INPAINTING

In the past few years, deep learning technology has made significant progress in the image inpainting. Image inpainting refers to the technique of restoring and reconstructing images based on background information. The generated images are expected to look very natural and difficult to distinguish from the ground truth. High-quality image inpainting not only requires the semantics of the generated content to be reasonable but also requires that the texture of generated image clear and realistic enough. Recently, image inpainting

methods based on deep learning have achieved promising results, especially based on GANs.

1) Deepfillv1

Yu *et al.* [43] proposed a deep generative model-based image inpainting approach called Deepfillv1. The framework of Deepfillv1 is summarized in Fig. 11.

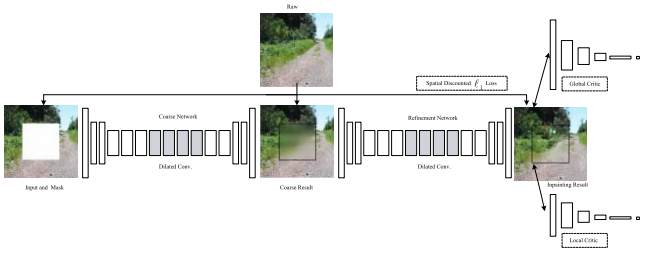

FIGURE 11. Overview of the improved generative inpainting framework [43].

Deepfillv1 combines the solution of deep learning algorithms concerning the advantages of traditional algorithms. It further improves the generation network and can automatically repair a picture with multiple holes or large holes, which can produce images of higher quality than earlier methods. The method can synthesize novel image structures and makes use of surrounding image features to make better predictions. The authors utilized a feedforward and fully convolutional neural network to process images with multiple holes during the test time. This method is a coarse-to-fine generative image inpainting framework with a novel contextual attention module that can improve the image inpainting results by learning the feature representations for explicit matching and attending to relevant background patches.

2) ExGANs

Dolhansky *et al.* [44] proposed a novel in-painting approach called Exemplar GANs (ExGANs). The architecture of ExGANs is shown in Fig 12.

The learning objective of reference image inpainting is defined as:

$$\min_G \max_D V(D,G) = \mathrm{E}_{x_i,r_i\sim Pdata(x,r)}[\log D(x_i,r_i)] + \mathrm{E}_{r_i\sim p_c,G(\cdot)\sim Pz}[\log 1 - D(G(z_i,r_i))] + \|G(z_i,r_i) - x_i\|_1 \quad (8)$$

The adversarial objective of code inpainting is defined as:

$$\min_G \max_D V(D,G) = \mathrm{E}_{x_i,c_i\sim Pdata(x,c)}[\log D(x_i,c_i)] + \mathrm{E}_{c_i\sim p_c,G(\cdot)\sim Pz}[\log 1 - D(G(z_i,c_i))] + \|G(z_i,c_i) - x_i\|_1 + \|C(G(z_i,c_i) - c_i\|_2 \quad (9)$$

The authors use exemplar information as a reference image of the region to inpaint a person with closed eyes in a natural picture which can produce high-quality and personalized inpainting results. It can also describe the object with

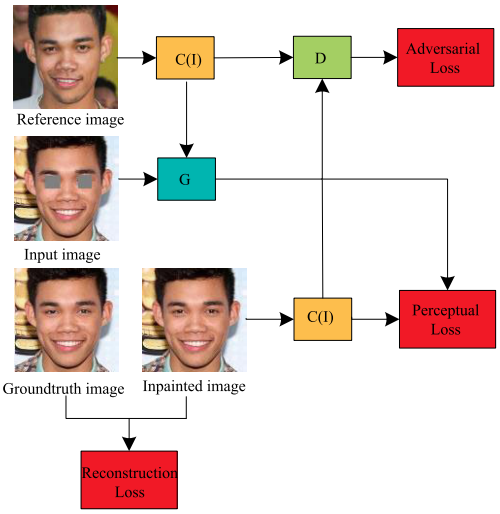

FIGURE 12. General architecture of an exemplar GAN [44].

a perceptual code in the task of a closed-to-open eye to produce a photo-realistic and personalized image in terms of perception and semantics. ExGANs are a type of conditional GAN that can increase the descriptive power by inserting at multiple points within the adversarial network with the extra information. It is a useful method for image generation or inpainting that use reference images or perceptual codes as identifying information which has superior perceptual results.

3) Deepfillv2

Yu *et al.* [45] proposed a novel image inpainting system based on deep learning which uses free-form masks and inputs to complete images called Deepfillv2. The architecture of Deepfillv2 is shown in Fig. 13.

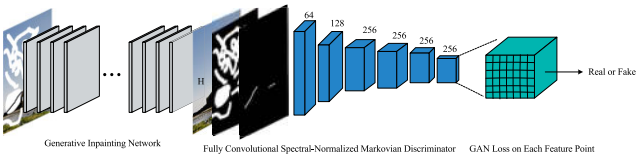

FIGURE 13. The architecture of Deepfillv2 [45].

The objective function is:

$$L_{D^{sn}} = \mathrm{E}_{X\sim Pdata(X)}[ReLU(1 - D^{sn}(x))] + \mathrm{E}_{Z\sim Pz(Z)}[ReLU(1 + D^{sn}(G(z)))] \quad (10)$$

$$L_G = -\mathrm{E}_{Z\sim Pz(Z)}[D^{sn}(G(z))] \quad (11)$$

This method is based on gated convolutions and can handle images with free-form masks anywhere or any shapes. The authors proposed a GAN loss called SN-PatchGAN which makes the training fast and stable. It is superior to the previous methods and can produce more flexible results with higher-quality. Furthermore, it can be used to remove distracting objects, clear watermarks, edit faces and fill in missing regions. Moreover, the image inpainting system which is

based on an end-to-end generative network is useful to improve inpainting results with user guidance input.

4) EdgeConnect

Nazeri *et al.* [46] proposed a two-stage adversarial model called EdgeConnect, a novel approach for image inpainting. The structure of the EdgeConnect is shown in Fig. 14.

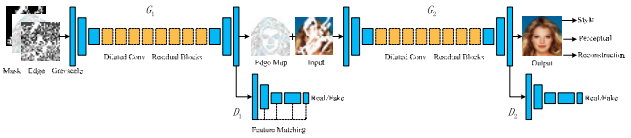

FIGURE 14. The structure of EdgeConnect [46].

The training objective of the edge generator network is:

$$\min_{G_1} \max_{D_1} L_{G_1} = \min_{G_1}(\lambda_{adv,1} \max_{D_1}(L_{adv,1}) + \lambda_{FM} L_{FM}) \quad (12)$$

The loss function of the image completion network is:

$$L_{G_2} = \lambda_{\ell_1} L_{\ell_1} + \lambda_{adv,2} L_{adv,2} + \lambda_p L_{perc} + \lambda_s L_{style} \quad (13)$$

EdgeConnect is an image completion network that uses hallucinated edges as a priori to fill in the missing regions. It consists of an edge generator and an image completion network which can reproduce filled regions exhibiting fine details. The edge generator is used to get edges of the missing region of the image which can be regular or irregular, and the image completion network is used to fill in the missing regions. The authors proposed a new image painting method based on deep learning that can be used for image inpainting task and reconstruct reasonable structures of the missing regions. Furthermore, it does a good job of dealing with images that have multiple or irregular shapes of missing regions. It can be used for removing unwanted objects from the images or as an interactive image editing tool and get a good result in terms of quantitative and qualitative measurements. However, the current problem is that the edge generating model sometimes fails to depict the edges accurately when a large part of the image is missing or in highly textured regions.

5) PEN-Net

Zeng *et al.* [47] proposed an image inpainting method based on deep generative models called Pyramid-context ENcoder Network (PEN-Net). The structure of the PEN-Net is shown in Fig. 15.

The adversarial loss for the discriminator is denoted as:

$$L_D = \mathrm{E}_{X \sim Pdata(X)}[\max(0, 1 - D(x))] + \mathrm{E}_{Z \sim P_z}[\max(0, 1 + D(z))] \quad (14)$$

The adversarial loss for the generator is denoted as:

$$L_G = -\mathrm{E}_{Z \sim P_z}[D(z)] \quad (15)$$

The PEN-Net is a method which is proposed for high-quality image inpainting and is used to fill in the missing

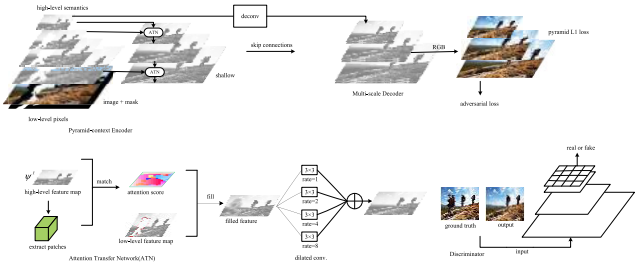

FIGURE 15. The structure of PEN-Net [47].

regions of the image with plausible content. The authors put forward the idea of a pyramid-context encoder which uses a high-level semantic feature map as a guide and transfer the learned attention to the previous low-level feature map so that the network can learn the region affinity progressively. It can be used to fill in the regions in a damaged image and get a result both visually and semantically plausible. Moreover, the main idea of the method is to encode the contextual semantics learned from the full resolution input, and restore an image by decoding the semantic features back. Both visual and semantic coherence of the generated content can be ensured with the attention transferred from deep to shallow in a pyramid fashion. At the same time, the authors proposed a new loss function to make the training converge fast and generate more realistic results. The network is superior to the previous method and can generate semantically-reasonable and visually-realistic images.

6) OTHER METHODS

Yang *et al.* [48] proposed a multi-scale neural patch synthesis method based on deep learning which uses image content and texture constraints to optimize the task of image inpainting. The method can not only restore images with semantically plausible contents but also preserve the high-frequency details. Yeh *et al.* [49] proposed a new approach for the semantic image inpainting, which can achieve pixel-level photorealism and generate satisfactory results. Li *et al.* [50] proposed an effective face completion method based on a deep generative model, and it can restore images with a large area of missing pixels and achieve a realistic face completion result.

Image inpainting methods based on GANs nowadays can achieve more reasonable and semantically consistent results than traditional methods. Currently, some methods use gated convolutions to restore images with free-form masks, which can restore images with multiple holes or fill in missing areas with irregular shapes. However, the quality of the image inpainting is sensitive to the position and size of the masks.

D. FACE IMAGE SYNTHESIS

In recent years, face image synthesis is a hot topic in photo processing because of the heavy use of pictures on social media. Due to the performance improvement of GANs, facial image processing has made great progress. A series of

methods have emerged to improve the quality of face image generation.

1) ELEGANT

Xiao *et al.* [51] proposed a model for transferring multiple face attributes called ELEGANT. The framework of ELEGANT is shown in Fig. 16.

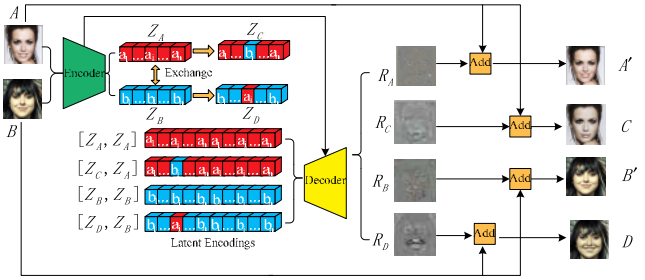

FIGURE 16. The framework of ELEGANT [51].

The loss of the discriminator is:

$$L_D = L_{D_1} + L_{D_2} \tag{16}$$

The loss of the generator is:

$$L_G = L_{reconstruction} + L_{adv} \tag{17}$$

ELEGANT is an effective method for face attributes transferring. It receives two images of opposite attributes as inputs and can produce high-quality images with finer details. Furthermore, it exchanges a certain part of the encodings to transfer the same type of attributes from one image to another. This method can manipulate several attributes simultaneously by encoding different attributes into disentangled parts in the latent space. The model is based on a U-Net [52] structure and is trained with multi-scale discriminators which can help to improve the quality of the generated images. Besides, it can generate higher resolution images with the help of residual learning to facilitate training.

2) STGAN

Liu *et al.* [53] proposed an arbitrary facial attribute editing model called STGAN, which achieves high-quality editing results. The structure of STGAN is shown in Fig. 17.

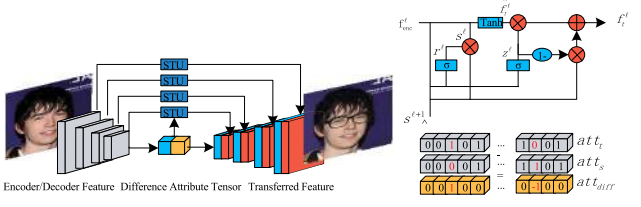

FIGURE 17. The structure of STGAN [53].

The objective function of discriminator D is formulated as:

$$\min_D L_D = -L_{D_{adv}} + \lambda_{L_1} L_{D_{att}} \tag{18}$$

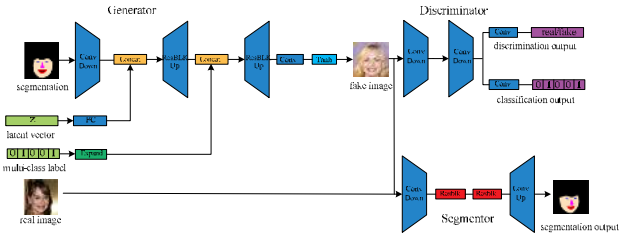

FIGURE 18. The framework of SCGAN [56].

The objective function of generator G is formulated as:

$$\min_G L_G = -L_{G_{adv}} + \lambda_2 L_{G_{att}} + \lambda_3 L_{rec} \tag{19}$$

This method solves the fine-grained control on the label of the face attribute and realizes multi-attribute transformation. The model takes a difference attribute vector as input to change the related attributes instead of all target attributes in specific editing tasks. STGAN is a high-precision attribute editing model based on AttGAN [54] and StarGAN [55]. It helps to improve the generated image quality and to get a clear editing result. Besides, it can not only improve the ability of face attribute manipulation but also can be used for season translation. The authors proposed selective transfer units (STUs) to enhance attribute editing which can improve the accuracy of attribute manipulation and improve perception quality. STGAN can improve the quality of generated images and realize flexible translation of attributes by focusing on the editing attributes to be changed.

3) SCGAN

Jiang *et al.* [56] proposed a novel image generation model called Spatially Constrained Generative Adversarial Network (SCGAN). The framework of SCGAN is shown in Fig. 18.

The objective function of SCGAN is represented as:

$$L_S = L_{seg}^{real} \tag{20}$$

$$L_D = -L_{adv} + \lambda_{cls} L_{cls}^{real} \tag{21}$$

$$L_G = L_{adv} + \lambda_{cls} L_{cla}^{fake} + \lambda_{seg} L_{seg}^{fake} \tag{22}$$

This method can generate images with clear edge details and can preserve spatial information. It makes the spatial constraints feasible as additional controllable signals which are decoupled from the latent vector. Moreover, the authors designed a generator network that takes a semantic segmentation, a latent vector and an attribute-level label as inputs to enhance the spatial controllability step by step. Meanwhile, the authors proposed a segmentor network to impose spatial constraints on the generator which can accelerate and stabilize the model convergence. SCGAN is an effective method that can control the spatial contents and can generate high-quality images. It can not only solve the foreground-background mismatch problem but it is also easy and fast to train. Besides, SCGAN is very effective at controlling spatial contents which can specify attributes and help to improve general visual quality and get quantitative results.

4) EXAMPLE-GUIDED IMAGE SYNTHESIS

Wang *et al.* [57] proposed an example-guided image synthesis solution by using a semantic label map and an exemplary image, and its framework is summarized as shown in Fig. 19.

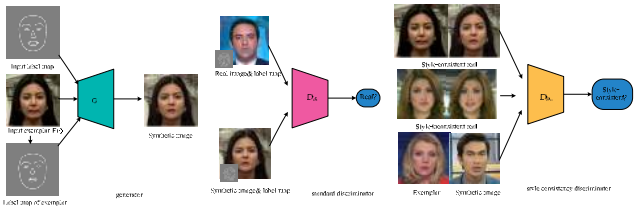

FIGURE 19. Overview of the framework [57].

The objective function is formulated as:

$$G^* = \arg\min_G \max_{D_R, D_{SC}} L(G, D_R, D_{SC}) \tag{23}$$

This method is based on conditional generative adversarial networks aim to synthesize images from semantic label maps and using an exemplary image to indicate facial expression or full body poses. The authors proposed a novel style consistency discriminator and an adaptive semantic consistency loss to make sure that the synthesized image is consistent in style with the exemplar. Furthermore, a training data sampling strategy is also used to synthesize style-consistent results. It is an effective method that can be used on the face or street view synthesis tasks which can produce qualitative and quantitative results. Moreover, it can generate realistic and style-consistent images with the help of style consistency discriminator.

5) SGGAN

Jiang *et al.* [58] proposed a novel multi-domain face image translation method called Segmentation Guided Generative Adversarial Networks (SGGAN). The framework of SGGAN is shown in Fig. 20.

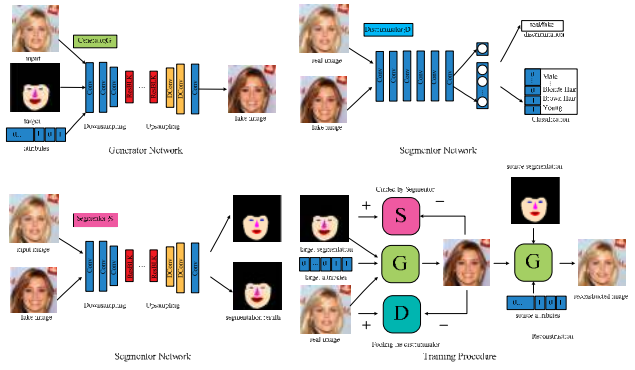

FIGURE 20. Illustration of SGGAN [58].

The objective function of the SGGAN network is summarized as:

$$L_S = L_{seg}^{real} \tag{24}$$

$$L_D = -L_{adv} + \lambda_1 L_{cls}^{real} \tag{25}$$

$$L_G = L_{adv} + \lambda_1 L_{cls}^{fake} + \lambda_2 L_{seg}^{fake} + \lambda_3 L_{rec} \tag{26}$$

The method is based on a deep generative model that pays attention to higher-level and instance-specific information and can generate realistic images of high quality. It has spatial controllability in the image translation process by utilizing semantic segmentation to improve the performance of image generation and provides spatial mapping. The authors proposed a segmentor network to provide the generated images with semantic information. Besides, it can improve the quality of image generation with the ability of spatial modification. The method uses the segmentation information to guide the generation of images which can make the details clear. SGGAN can be used for face image translation by providing strong regulations during the training process.

6) MaskGAN

Lee *et al.* [59] proposed a geometry-oriented face manipulation framework called MaskGAN. The pipeline of MaskGAN is shown in Fig. 21.

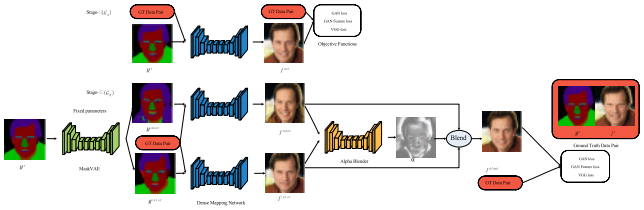

FIGURE 21. The pipeline of MaskGAN [59].

The objective loss function is:

$$L_{G_A,G_B} = L_{adv}(G, D_{1,2}) + \lambda_{feat} L_{feat}(G, D_{1,2}) + \lambda_{percept} L_{percept}(G) \tag{27}$$

The method overcomes the shortcomings of operating on a predefined set of face attributes. It makes the users manipulate images with more freedom by using semantic masks as an intermediate representation, which enables diverse and interactive face manipulation. MaskGAN can achieve diverse generation results by using dense mapping networks to learn style mapping between the free-form user modified mask and the target image. Furthermore, it makes the framework more robust to manipulate by using editing behavior simulated training which models users editing behavior on the source mask. MaskGAN can be used to manipulate face image flexibly and preserve the fidelity.

7) OTHER METHODS

Lin *et al.* [60] proposed an unpaired image-to-image translation method called Domain-supervised GAN (DosGAN), and it uses domain information as explicit supervision and achieves conditional translation with face images in CelebA. Mokady *et al.* [61] proposed a novel mask-based method which uses the masks to reconstruct the face images and enables high quality and various content translations. Yin *et al.* [62] proposed an instance-level facial attribute transfer method which uses the geometry-aware flow as a

representation for transferring the images with instance-level facial attributes.

The Face image synthesis method uses encoder-decoder and generative adversarial networks to solve the problem of arbitrary attribute editing. High-quality images with fine detail can be generated by this architecture which makes high-precision face attribute editing come true. However, there may have some mode collapse problems.

E. HUMAN IMAGE SYNTHESIS

Human image synthesis aims to manipulate the visual appearance of the character images by transferring the pose of a character to the target pose, which can be calculated from other characters.

1) TEXT GUIDED PERSON IMAGE SYNTHESIS

Zhou *et al.* [63] proposed an approach which can manipulate the pose and attribute of generated person images according to a specific text description. The structure is shown in Fig. 22 and Fig. 23.

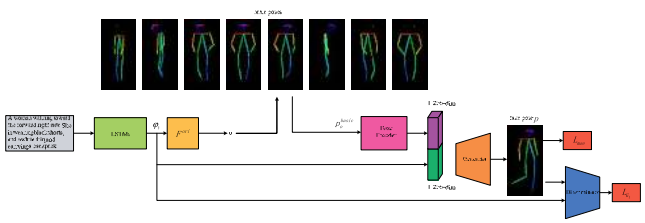

FIGURE 22. Text guided pose generator [63].

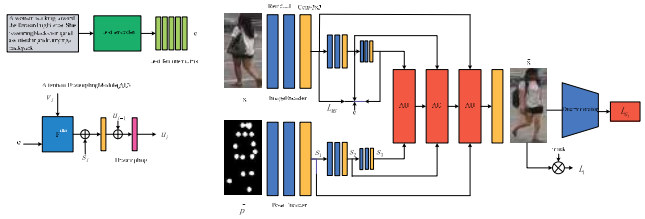

FIGURE 23. Pose and attribute transferred person image generator [63].

The objective function of text guided pose generator is formulated as:

$$L_{Stage-\mathrm{I}} = L_{G_1} + \lambda_1 L_{mse} + \lambda_2 L_{cls} \tag{28}$$

The objective function of the multi-task person image generator is defined as:

$$L_{Stage-\mathrm{II}} = L_{G_2} + \lambda_1 L_1 + \lambda_2 L_{MS} \tag{29}$$

This method consists of text guided pose generation in the first stage and visual appearance transferred image synthesis in the second stage. The method can generate and edit images according to text description by establishing a mapping between image space and language space which extracts information from the text. The authors proposed a new image processing method based on natural language descriptions and a human pose inference network based on GAN. It uses the Visual Question Answering (VQA) perceptual score to assess the correctness of the change in attributes corresponding to a particular body part. The method first learns to infer a reasonable target human body posture according to the description and then synthesizes the appearance transferred character image based on the text and the target posture. It is an effective method that can manipulate the visual appearance by editing the generated person images based on natural language descriptions.

2) PROGRESSIVE POSE ATTENTION TRANSFER

Zhu *et al.* [64] proposed a new pose transfer method based on a generative adversarial network. Its architecture is shown in Fig. 24.

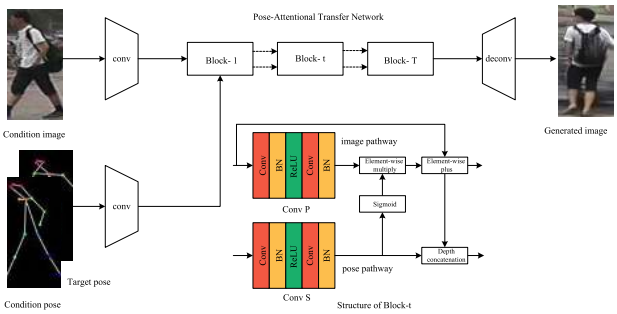

FIGURE 24. Generator architecture of the proposed method in [64].

The loss function is denoted as:

$$L_{full} = \arg\min_{G}\max_{D} \alpha L_{GAN} + L_{combL1} \tag{30}$$

This method can generate person images by using Pose Attentional Transfer Blocks (PATBs) to transfer certain regions in the generator. It can generate more realistic person images that are consistent with the input images in terms of appearance and shape. Furthermore, it uses the attention mechanism to guide the deformable transfer process of the appearance and pose progressively. It can not only improve computational efficiency but also reduce the model complexity. The method uses an appearance discriminator and a shape discriminator to determine whether the appearance and pose generated by the generator are true and produces more natural results than the previous method. The network is more interpretable by its attention masks which make the progressive pose-attentional transfer process visible. Moreover, it is capable of generating realistic images in both qualitative and quantitative measurements.

3) SEMANTIC PARSING TRANSFORMATION

Song *et al.* [65] proposed an unsupervised person image generation approach. Its framework is shown in Fig. 25.

The loss function of the semantic generative network is denoted as follows:

$$L_S^{total} = L_S^{adv} + \lambda^{ce} L_S^{ce} \tag{31}$$

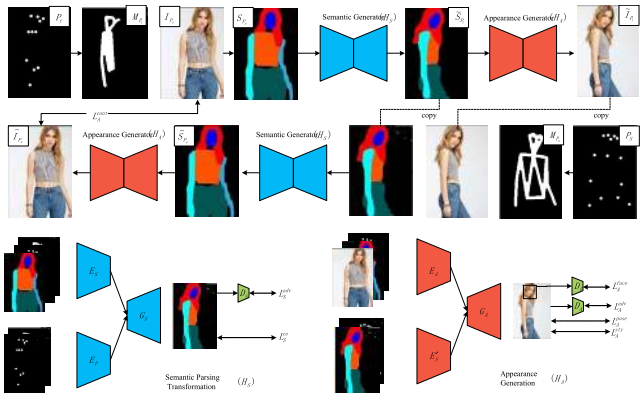

FIGURE 25. The framework for unsupervised person image generation [65].

The loss function of the appearance generative network is denoted as follows:

$$L_A^{total} = L_A^{adv} + \lambda^{pose} L_A^{pose} + \lambda_A^{cont} L_A^{cont} + \lambda^{sty} L_A^{sty} + L_A^{face} \quad (32)$$

The approach is divided into two subtasks which reduce the complexity of learning a direct mapping between human bodies with different poses. The semantic parsing transformation task is based on a semantic generative network that can transform between semantic parsing maps and simplify the non-rigid deformation learning. The appearance generation task is based on an appearance generative network that can synthesize semantic-aware textures. It is an unsupervised pose-guided person image generation method which can keep the clothing attributes and better body shapes. Moreover, it can be used to transfer clothing texture or control image manipulation. However, the problem is that the model would fail if there is an error in the conditional semantic map.

4) COORDINATE-BASED TEXTURE INPAINTING

Grigorev *et al.* [66] proposed a pose-guided human image generation approach based on deep learning. Its framework is shown in Fig. 26 and Fig. 27.

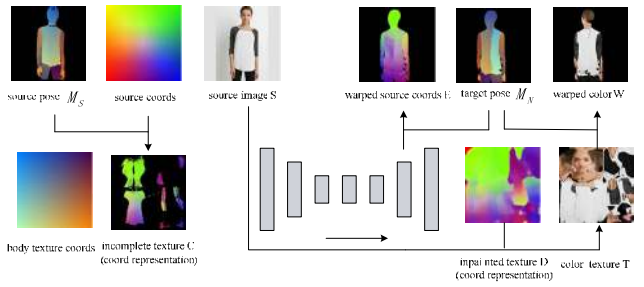

FIGURE 26. The coordinate-based texture inpainting [66].

The main idea of the method is to complete the texture of the human body by using a new inpainting method which estimates the appropriate source location for each part of the body surface. It establishes the correspondence between source and target view by warping the correspondence field between input image and texture into the target image coordinate frame according to the desired pose. The method uses the estimated correspondence field to guide the deformable skip connections in a fully-convolutional architecture which helps to synthesize the output image. It is a new method based on coordinate-based texture inpainting which can produce more texture details. Moreover, it works by estimating the texture of the human body based on a single photograph that can be used for garment transfer or pose-guided face resynthesis.

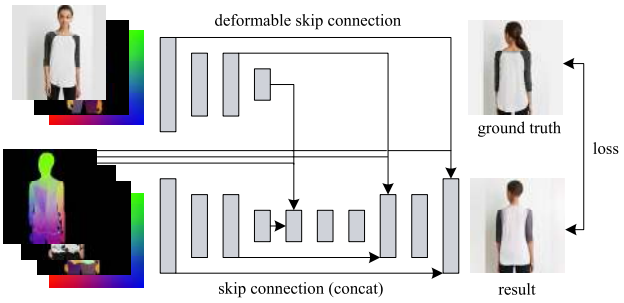

FIGURE 27. The final resynthesis [66].

5) OTHER METHODS

Tang *et al.* [67] proposed a keypoint-guided image generation method called Cycle In Cycle Generative Adversarial Network, which can generate photo-realistic person pose images. Ma *et al.* [68] proposed a person image generation method called Pose Guided Person Generation Network, and it can synthesize high-quality person images with arbitrary poses based on a person image and a pose. Ma *et al.* [69] proposed a person image generation approach, which can not only learn a disentangled representation of the image factors but also generate realistic person images based on a two-stage reconstruction pipeline.

Human image synthesis methods are usually based on a person image and an arbitrary pose to manipulate the visual appearance of a person image. It is possible to reconstruct detail-rich textures for pose-guided human image generation. However, sometimes the texture and the generated images are blurred.

IV. IMAGE-TO-IMAGE TRANSLATION

Recently, image-to-image translation has made great progress. The goal of image translation is to learn the mapping from the source image domain to the target image domain, which changes the style or some other properties of the source domain to the target domain while keeps the image content unchanged.

A. IMAGE-TO-IMAGE TRANSLATION

Image-to-image translation using generative adversarial networks has drawn great attention in both supervised learning and unsupervised learning research. Noise-to-image GANs generate realistic images from random noise samples while image-to-image GANs generate diverse images from images.

Many GAN-variants have been proposed, which achieved good results in image-to-image translation tasks.

1) CycleGAN

Zhu *et al.* [70] presented an unpaired image-to-image translation approach called CycleGAN. The model of CycleGAN is shown in Fig. 28.

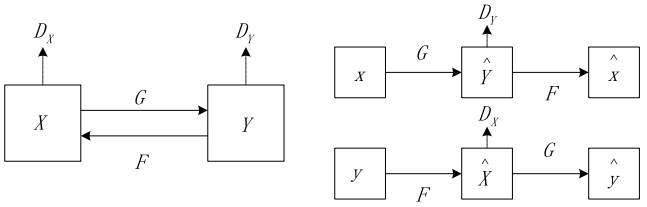

FIGURE 28. The model of CycleGAN [70].

The objective is:

$$\begin{aligned} G^*, F^* &= \arg\min_{G,F} \max_{D_X,D_Y} L(G, F, D_X, D_Y) \\ &= L_{GAN}(G, D_Y, X, Y) \\ &\quad + L_{GAN}(F, D_X, Y, X) \\ &\quad + \lambda L_{cyc}(G, F) \end{aligned} \tag{33}$$

CycleGAN is an innovation of method in the field of unsupervised image translation research. Based on CycleGAN, various unsupervised image translation studies have emerged. It proposed the cycle consistency loss which can learn the mapping without a training set of aligned image pairs. The method achieves good results on many translation tasks involve color and texture changes, such as collection style transfer, object transfiguration, season transfer. However, it fails when it requires geometric changes.

2) UNIT

Liu *et al.* [71] proposed an unsupervised image-to-image translation framework called UNsupervised Image-to-image. Translation (UNIT) based on Coupled GANs [72]. The framework of UNIT is shown in Fig. 29.

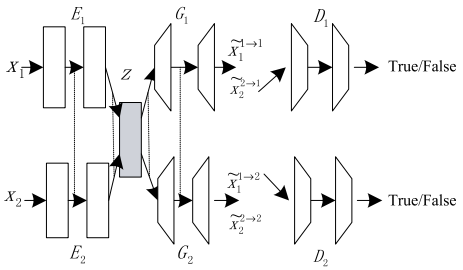

FIGURE 29. The framework of UNIT [71].

The objective function is defined as:

$$\begin{aligned} \min_{E_1,E_2,G_1,G_2} \max_{D_1,D_2} & L_{VAE_1}(E_1, G_1) + L_{GAN_1}(E_2, G_1, D_1) \\ & + L_{CC_1}(E_1, G_1, E_2, G_2) \\ & \times L_{VAE_2}(E_2, G_2) + L_{GAN_2}(E_1, G_2, D_2) \\ & + L_{CC_2}(E_2, G_2, E_1, G_1) \end{aligned} \tag{34}$$

The method is based on generative adversarial networks as well as variational autoencoders. It generates corresponding images in two domains by adversarial training objective interacts with a weight-sharing constraint to enforce a shared-latent space. Besides, it relates the translated images with the input images in the respective domains by using variational autoencoders. UNIT is a method which can not only present image translation results with high quality but also can be used for various unsupervised image translation tasks, such as street scene image translation, or face image translation. However, there are two limitations to this framework. On the one hand, as a result of the Gaussian latent space assumption, the translation model is unimodal. On the other hand, the saddle point searching problem may cause the training unstable.

3) MUNIT

Xun Huang *et al.* [73] proposed an unsupervised image-to-image translation framework called Multimodal Unsupervised Image-to-image Translation (MUNIT). The architecture of MUNIT is shown in Fig. 30.

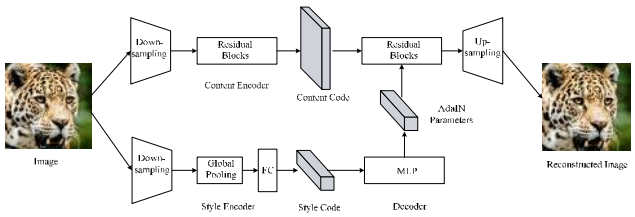

FIGURE 30. The method overview of DRIT [74].

The objective function is defined as:

$$\begin{aligned} \min_{E_1,E_2,G_1,G_2} \max_{D_1,D_2} & L(E_1, E_2, G_1, G_2, D_1, D_2) \\ &= L^{x_1}_{GAN} + L^{x_2}_{GAN} + \lambda_x(L^{x_1}_{recon} + L^{x_2}_{recon}) \\ &\quad + \lambda_c(L^{c_1}_{recon} + L^{c_2}_{recon}) + \lambda_s(L^{s_1}_{recon} + L^{s_2}_{recon}) \end{aligned} \tag{35}$$

MUNIT can generate diverse results from the source domains which are multimodal conditional distribution. It trains two auto-encoders, one encodes the content of the image, and the other encodes the style, which enables the generation of multimodal images. Furthermore, it decomposes the image representation into a content code that is domain-invariant, and a style code to captures domain-specific properties. The method recombines the content code with a random style code sampled from the style space of the target domain to translate the image to another domain. Moreover, two domains that share the same content distribution with different style distributions. It can control the style of translation results according to an example style image that achieves high quality and diversity.

4) DRIT

Hsin-Ying Lee *et al.* [74] proposed an image-to-image translation approach termed DRIT. The method of DRIT is shown in Fig. 31.

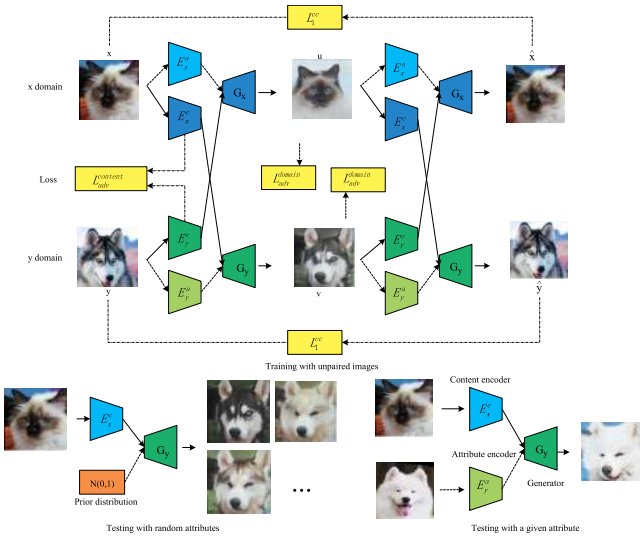

FIGURE 31. The architecture of MUNIT [73].

The objective function of the network is:

$$\min_{G,E^c,E^a} \max_{D,D^c} \lambda_{adv}^{content} L_{adv}^{c} + \lambda_1^{cc} L_1^{cc} + \lambda_{adv}^{domain} L_{adv}^{domain} + \lambda_1^{recon} L_1^{recon} + \lambda_1^{latent} L_1^{latent} + \lambda_{KL} L_{KL} \quad (36)$$

DRIT is a method that is capable of generating realistic and diverse results without aligned training pairs based on disentangled representation. The generator for each domain in DRIT consists of two encoders, one encodes the content of the image and the other encodes the style of the image, which makes a domain-invariant content space to capture the shared information across domains as well as a domain-specific attribute space. It can generate diverse results with the encoded content features from an image and attribute vectors from the attribute space. Furthermore, it facilitates the factorization of the domain-invariant content space along with domain-specific attribute space by using a content discriminator and trains the model with paired images by using a cross-cycle consistency loss according to disentangled representations. Moreover, it can produce qualitative and quantitative outputs on a wide range of tasks in the absence of paired data. Meanwhile, the approach called DRIT ++ [75] seeks regularization term to alleviate the mode collapse problem in DRIT, especially in shape-variation translation.

5) TransGaGa

Wu *et al.* [76] proposed a geometry-aware disentangle-and-translate framework which can be used for unsupervised image-to-image translation called TransGaGa. The architecture of TransGaGa is shown in Fig. 32.

The loss function of the method is:

$$L_{total} = L_{CVAE} + L_{prior} + L_{con}^{s} + L_{cyc}^{s} + L_{cyc}^{g} + L_{cyc}^{pix} + L_{adv}^{a} + L_{adv}^{g} + L_{adv}^{pix} \quad (37)$$

This method can learn a mapping between two visual domains as well as the translation across large geometry

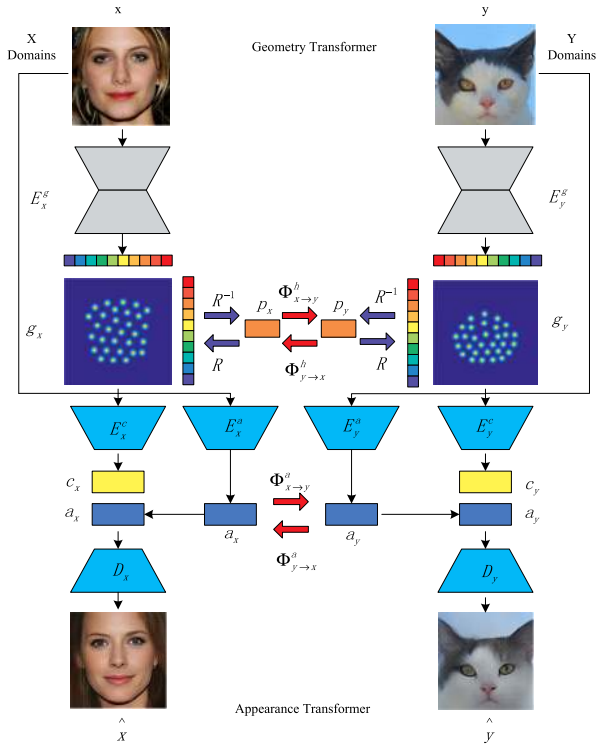

FIGURE 32. The architecture of TransGaGa [76].

variations. It learns a translation which is built on appearance and geometry space separately by disentangling the image space into an appearance space and a geometry latent space to decompose image-to-image translation into two separate problems. Furthermore, it proposed a geometry prior loss and a conditional VAE loss that can learn independent but complementary representations. TransGaGa is capable of dealing with complex objects image-to-image translation tasks such as near-rigid or non-rigid objects translation. Besides, it supports multimodal translation and achieves qualitative and quantitative results.

6) RelGAN

Lin *et al.* [77] proposed a multi-domain image-to-image translation method based on relative attributes called RelGAN. The model of RelGAN is shown in Fig. 33.

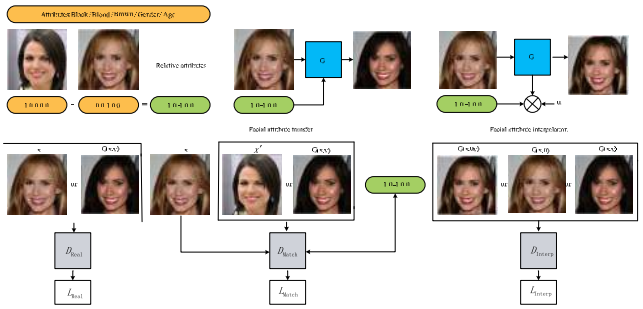

FIGURE 33. The model of RelGAN [77].

The loss function of the method is formulated as:

$$\min_D L^D = -L_{Real} + \lambda_1 L^D_{Match} + \lambda_2 L^D_{Interp} \tag{38}$$

$$\min_G L^G = L_{Real} + \lambda_1 L^G_{Match} + \lambda_2 L^G_{Interp} + \lambda_3 L_{Cycle} + \lambda_4 L_{Self} + \lambda_5 L_{Ortho} \tag{39}$$

The method takes relative attributes as input to describe the selected attributes that need to be changed. It is able to produce images by changing specific properties of interest in a continuous manner and keep the other attributes unchanged. RelGAN helps to improve interpolation quality by training on real-valued relative attributes instead of binary-valued attributes with additional discriminators. It is can be used for facial attribute transfer and interpolation. Furthermore, it achieves quantitative and qualitative results in multi-domain image-to-image translation tasks.

7) OTHER METHODS

Li *et al.* [78] proposed an Attribute Guided UIT (Unpaired Image-to-Image Translation) approach termed AGUIT, which can perform image translation tasks by adopting a novel semi-supervised learning process and decomposing the image representation into domain-invariant content code and domain-specific style code. Chang *et al.* [79] proposed an image-to-image translation approach called Sym-parameterized Generative Network (SGN), and it focuses on the loss area and infers translations of images in mixed domains by learning the combined characteristics of each domain. Tomei *et al.* [80] proposed a semantic-aware approach that can reduce the gap between visual features of artistic and realistic data by translating artworks to photo-realistic visualizations. Mo *et al.* [81] proposed an unsupervised image-to-image translation approach called instance-aware GAN (InstaGAN), which can not only incorporate the instance information but also improve the multi-instance transfiguration.

The style transfer method widely adopts an encoder-decoder-discriminator (EDD) architecture and can produce diverse outputs. However, it may generate images with artifacts sometimes. Besides, the training may be unstable and there may have mode collapse problems.

V. IMAGE EDITING

A. IMAGE EDITING

The image editing is an interesting but challenging task in computer vision. It mainly manipulates images through color and geometric interactions to complete tasks such as image deformation and blending. Recently, image editing based on deep learning has received more and more attention, especially with the development of GANs. Image editing using GANs has made great progress and becomes a highly recognized subject in computer vision. A series of image editing methods have appeared.

1) SC-FEGAN

Jo and Park [82] proposed a face editing system called SC-FEGAN based on Generative Adversarial Network, and it can synthesize images with high quality by using intuitive user inputs. The architecture of SC-FEGAN is shown in Fig. 34.

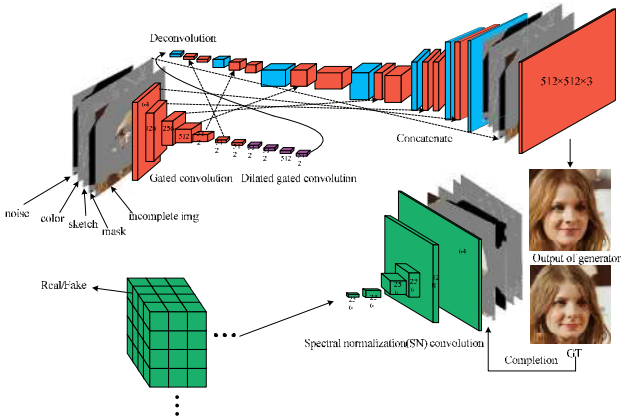

FIGURE 34. The network architecture of SC-FEGAN [82].

The loss functions are shown below:

$$L_{G\_SN} = -IE[D(I_{comp})] \tag{40}$$

$$L_G = L_{per-pixel} + \alpha L_{percept} + \beta L_{G\_SN} + \gamma(L_{style}(I_{gen}) + L_{style}(I_{comp})) + \nu L_{tv} + \varepsilon IE[D(I_{gt})^2] \tag{41}$$

$$L_D = IE[1 - D(I_{gt})] + IE[1 + D(I_{comp})] + \theta L_{GP} \tag{42}$$

SC-FEGAN is a face editing method that uses a free-form mask, sketch and color as an input. The method can restore the area of any shape and reconstruct detail-rich textures of large regions. It generates images guided with sketches and color by using an end-to-end trainable convolutional network and free-form user input with color and shape. In addition, the method is able to generate realistic results by training an additional style loss. It can generate high quality and realistic results with the proposed network architecture and loss functions.

2) FE-GAN

Dong *et al.* [83] proposed an image editing approach called Fashion Editing Generative Adversarial Network (FE-GAN) by using a multi-scale attention normalization. The architecture of FE-GAN is shown in Fig. 35.

The objective function of the free-form parsing network is formulated as:

$$L_{free-form-parser} = \gamma_1 L_{parsing} + \gamma_2 L_{feat} + \gamma_3 L_{adv} \tag{43}$$

The objective function of the parsing-aware inpainting network is formulated as:

$$L_{inpainter} = \lambda_1 L_{mask} + \lambda_2 L_{foreground} + \lambda_3 L_{face} + \lambda_4 L_{faceTV} + \lambda_5 L_{perceptual} + \lambda_6 L_{style} + \lambda_7 L_{adv} \tag{44}$$

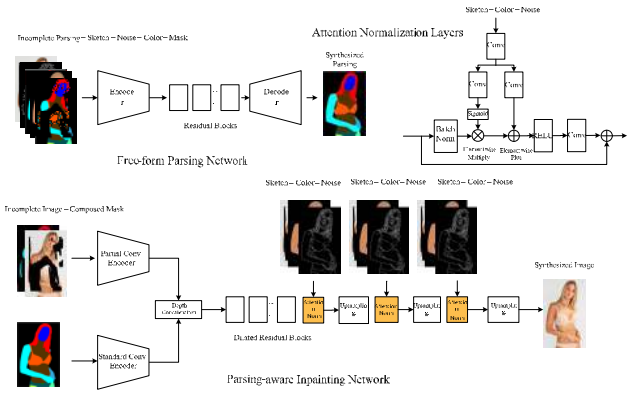

FIGURE 35. The network architecture of FE-GAN [83].

FE-GAN uses sketches and color strokes to manipulate and edit fashion images. It is able to leverage the semantic structural information to edit fashion images by free-forms sketches and sparse color strokes. The method controls the human parsing generation with the sketch and color by using a free-form parsing network. It renders detailed textures with semantic guidance from the human parsing map by using a parsing-aware inpainting network. Furthermore, it improves the quality of the generated images by using a new attention normalization layer in the decoder of the inpainting network. The method can generate high-quality images with convincing details by using a foreground-based partial convolutional encoder.

3) MASK-GUIDED PORTRAIT EDITING

Gu *et al.* [84] proposed a portrait editing framework based on mask-guided conditional GANs, and it uses the face masks to guide the image generation. Its framework is shown in Fig. 36.

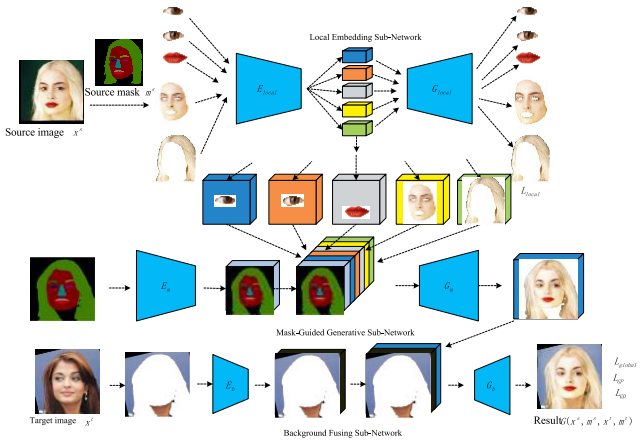

FIGURE 36. The framework for mask-guided portrait editing in [84].

The loss function is:

$$L_G = \lambda_{local} L_{local} + \lambda_{global} L_{global} + \lambda_{GD} L_{GD} + \lambda_{GP} L_{GP} \quad (45)$$

The method is guided by face masks which can generate diverse images with high-quality. It controls the synthesis and editing of facial images by learning feature embeddings for each face component separately. It also helps to improve the performance of image translation and local face editing. This method can edit face components in the generated images with the help of changeable input facial masks and the source image. Moreover, it leverages the input masks to synthesize facial data which can be used for the face paring model. The method can produce realistic outputs and realize face editing.

4) FaceShapeGene

Xu *et al.* [85] proposed a face image editing approach termed FaceShapeGene, and it can compute a disentangled shape representation for face images. The pipeline of FaceShapeGene is shown in Fig. 37.

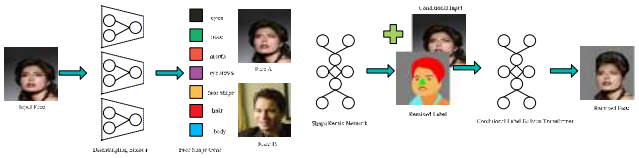

FIGURE 37. The pipeline of FaceShapeGene [85].

The objective function is:

$$L_{total} = L_{cycle,I} + \lambda_{CL} L_{cyc,L} + \lambda_{GI} L_{GAN,I} + \lambda_{GL} L_{GAN,L} + \lambda_{ID} L_{ID} \quad (46)$$

The FaceShapeGene realizes face editing tasks by encoding the shape information of each semantic facial part into a 1D latent vector separately. The authors proposed a shape-remix network to recombine the part-wise latent vectors from different individuals which produces remixed face shape in the form of a label map. They also use a conditional label-to-face transformer to perform part-wise face editing while preserving the identity of the subject. Furthermore, it trains the system in an unsupervised manner by using a cyclic training strategy. The method can disentangle the shape features of different semantic parts correctly and achieve partial editing with realistic results.

5) OTHER METHODS

Shen *et al.* [86] proposed a semantic face editing approach termed InterFaceGAN, which can synthesize high-fidelity image by semantic face editing in latent space. Portenier *et al.* [87] proposed a face image editing approach called FaceShop, and it can produce high quality and semantically consistent results.

In recent years, image editing using GANs has made great progress and achieve good results. Mask-guided image editing method is widely used and it can synthesize realistic with high quality. However, most of the methods proposed at present can only be used for the task of face portrait editing, and there will be artifacts and blurred results when performing whole-body editing.

VI. CARTOON GENERATION

A. CARTOON GENERATION

The cartoon is popular with young people because of its interesting story. GANs have also attracted the interest of researchers in the field of cartoon generation, and they proposed a series of fresh and interesting cartoon generation methods.

1) CartoonGAN

Chen *et al.* [88] proposed a photo cartoonization solution called CartoonGAN, and it can transform a photo of a real-world scene into a cartoon style image. The architecture of CartoonGAN is shown in Fig. 38.

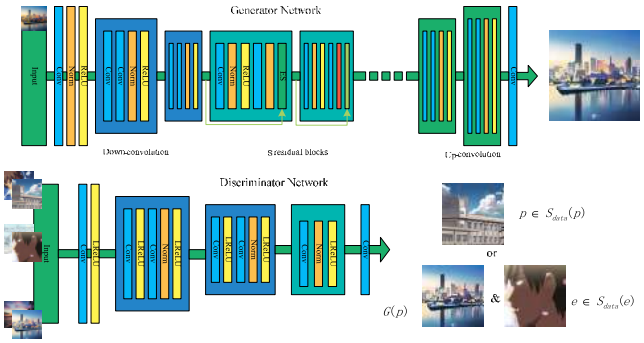

FIGURE 38. The architecture of the CartoonGAN [88].

The objective loss function is:

$$(G^*, D^*) = \arg\min_G \max_D L(G, D) = L_{adv}(G, D) + \omega L_{con}(G, D) \quad (47)$$

CartoonGAN is a cartoon generation method based on a generative adversarial network, which is easy to use by training with unpaired photos and cartoon images. The method is capable of generating high-quality cartoon images with clear edges and smooth color shading from real-world photos, following the style of specific artists. It copes with the substantial style variation between photos and cartoons by proposing a semantic content loss, which is formulated as a sparse regularization of high-level feature maps in the VGG network. CartoonGAN is able to preserve clear edges by proposing an edge-promoting adversarial loss. In addition, it is capable of improving the convergence of the network to the target manifold by introducing an initialization phase. The method can produce cartoon images with high-quality from real-world photos.

2) PI-REC

You *et al.* [89] proposed an image reconstruction approach called PI-REC, which can generate images from binary sparse edge and flat color domain. The architecture of PI-REC is shown in Fig. 39.

The loss function is calculated as below:

$$L_{G_1} = \alpha L_{per-pixel} + \beta L_{GAN-G} + \gamma L_{feature} + \delta L_{style} \quad (48)$$

$$L_{D_1} = L_{GAN-D} \quad (49)$$

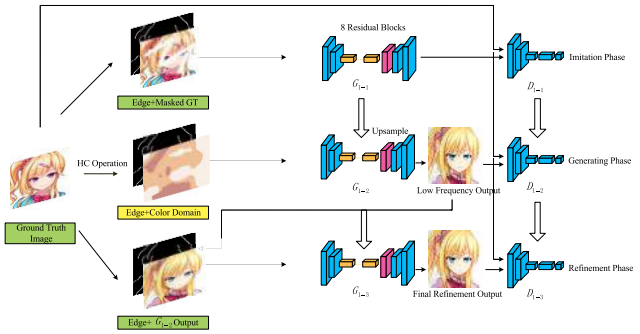

FIGURE 39. The network architecture of PI-REC [89].

PI-REC is able to reconstruct images by inputting binary sparse edge and flat color domain, which can not only control the content and style of generated images freely and accurately but also produce refined reconstruction results with high-quality. The method consists of three phases: Imitation Phase to initialize the networks, Generating Phase aims at reconstructing preliminary images, and Refinement Phase to fine-tune preliminary images and produce outputs with details. Besides, it can be used for hand-drawn draft translation tasks by utilizing parameter confusion operation, which obtains remarkable results. Furthermore, it is able to create anime characters by feeding the well-trained model with edge and color domain extracted from realistic photos, which improves the controllability and interpretability and generates abundant high-frequency details.

3) INTERNAL REPRESENTATION COLLAGING

Suzuki *et al.* [90] proposed an image synthesis strategy based on CNN, and it can manipulate the feature-space representation of the image in a trained GAN model to change the semantic information of an image over an arbitrary region. The algorithm of applying the feature-space collaging is shown in Fig. 40.

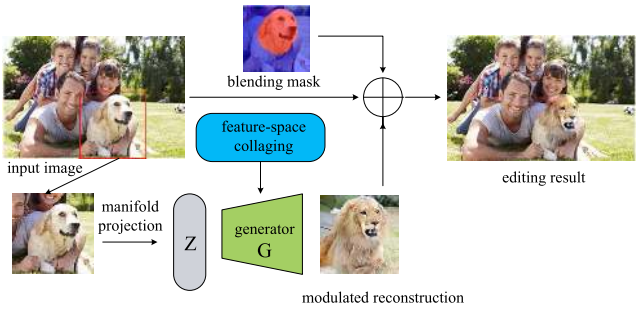

FIGURE 40. The algorithm of applying the feature-space collaging [90].

The method can be used to edit artificial images or real images by using spatial conditional batch normalization (sCBN), which is a type of conditional batch normalization with user-specifiable spatial weight maps. It can modify the intermediate features directly and by using feature-blending in any GAN with conditional normalization layers. Besides, it is able to be used to edit anime face which can synthesize realistic results. However, the problem is that it may

perform poorly in the transformation of a specific individual, and some information is bound to be lost in the process of projecting the target images to the restricted space of images.

4) U-GAT-IT

Kim *et al.* [91] proposed an image-to-image translation approach termed U-GAT-IT, and it can generate realistic anime images by using adaptive layer-instance normalization. The architecture of U-GAT-IT is shown in Fig. 41.

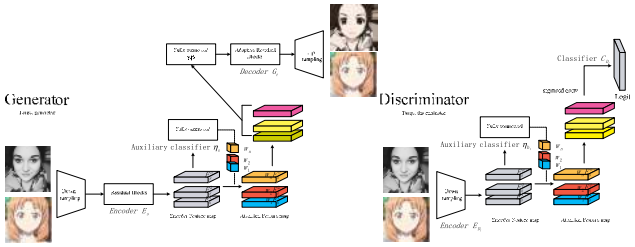

FIGURE 41. The architecture of U-GAT-IT [91].

The objective function is:

$$\min_{G_{s\to t},G_{t\to s},\eta_s,\eta_t}\ \max_{D_s,D_t,\eta D_s,\eta D_t} \lambda_1 L_{gan} + \lambda_2 L_{cycle} + \lambda_3 L_{identity} + \lambda_4 L_{cam} \quad (50)$$

The U-GAT-IT is an unsupervised image-to-image translation approach, which can translate images with holistic or large shape changes by handling the geometric changes between domains. It incorporates a learnable normalization function and an attention module to distinguish between the source and target domains. The attention module is based on the attention map and obtained by the auxiliary classifier to guide the model to focus on more important regions. Furthermore, the attention-guided model is able to control the amount of change in shape and texture flexibly with the proposed Adaptive Layer-Instance Normalization (AdaLIN). Moreover, it is a method that can produce anime face with more visually pleasing results based on the attention module and AdaLIN.

5) LANDMARK ASSISTED CycleGAN

Wu *et al.* [92] proposed a cartoon face generation approach based on CycleGAN, and it trains with unpaired data between real faces and cartoon ones, the architecture is shown in Fig. 42.

The landmark consistency loss is:

$$L_c(G_{(X,L)\to Y}) = \left\| R_Y(G_{(X,L)\to Y}(x,\ell)) - \ell \right\|_2 \quad (51)$$

The method is able to generate a cartoon face with high quality which captures the essential facial features of a person by proposing a landmark consistency loss and training a local discriminator in CycleGAN. It can produce cartoon faces with high-quality based on the conditional generator and discriminator, which enforces structural consistency in landmarks. Besides, it is a method guided by facial landmarks that can constrain the facial structure between two domains and

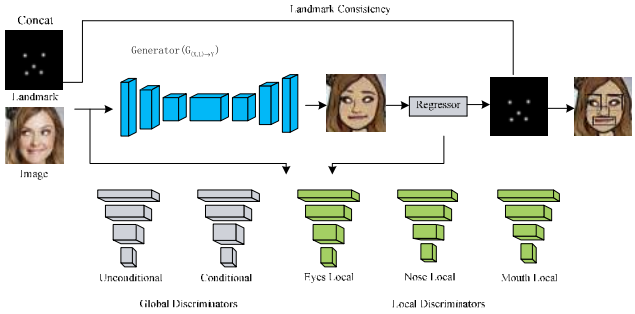

FIGURE 42. The architecture of the cartoon-face landmark-assisted CycleGAN [92].

can generate impressive high-quality cartoon faces according to the input human faces.

6) OTHER METHODS

Taigman *et al.* [93] proposed an image generation approach called Domain Transfer Network (DTN), which can transfer from face photos to emoji. Jin *et al.* [94] proposed a method, which can generate facial images of anime characters with a promising result. Li [95] proposed an image translation method called TwinGAN, and it achieves unsupervised image translation from a human to anime characters. Ma *et al.* [96] proposed an instance-level image translation approach called DA-GAN, which can synthesize animation face from a human face. Hamada *et al.* [97] proposed an anime generation method called Progressive Structure-conditional Generative Adversarial Networks (PSGAN), and it is able to generate character images with full-body and high-resolution based on structural information. Gokaslan *et al.* [98] proposed an image-to-image translation approach termed the GAN-imorph, which can translate human faces to anime faces. Cao *et al.* [99] proposed a photo-to-caricature translation approach, and it is able to transfer from face photos to caricatures. Xiang and Li [100] proposed a Generative Adversarial Disentanglement Network, which can generate high-fidelity anime portraits.

Cartoon generation based on GANs mostly uses the normalization method. It can produce high-quality cartoon images from real-world photos. However, it sometimes generates images with artifacts. Besides, it may not perform well in the image translation of a particular individual.

VII. DISCUSSION

So far, we have discussed some of the applications of GANs in the field of image synthesis. Table 1 gives a summary and comparison between different methods based on GANs in terms of pros and cons.

A. OPEN QUESTIONS

GANs can not only learn a highly nonlinear mapping from latent space to data space but also can utilize a large amount of unlabeled image data for deep representation learning. Compared with other frameworks, GANs tend to produce

TABLE 1. A brief summary of different GANs.

Application	Year	Pros & Cons	Comparison
PSGAN[30]	2017	1. Learn textures of great variability from large images. 2. Learn periodical textures. 3. Learn whole manifolds of textures and smoothly blend between their elements, thus creating novel textures. 4. Generate images of any desired size with a fast forward pass of a convolutional neural network. 5. Linear scalability in memory and increase the output image size. 6. Convergence can be sometimes tricky. 7. Suffer from "mode dropping". 8. Images that have larger structures or more general non-periodic features are not representable.	CIS 26.81
TextureGAN[31]	2018	1. Generate more realistic images. 2. Enable separate controls on color and content.	N/A
Texture Mixer[32]	2019	1. Controllability, smoothness, and realism.	CIS 26.68
ProGAN[36]	2018	1. Speed the training up. 2. Produce images of unprecedented quality. 3. The training is stable in large resolutions. 4. Semantic sensibility and understanding dataset-dependent constraints.	Resolution 1024x1024
Progressive Face Super-Resolution[37]	2019	1. Allow stable training. 2. Reduce overall training time. 3. Generate photo-realistic 8×super-resolved face images.	Resolution 128×128
BigGANs[38]	2018	1. Generate high-resolution, diverse samples from complex datasets.	Resolution 512×512
StyleGAN[39]	2019	1. Enable intuitive, scale-specific control of the synthesis.	Resolution 1024x1024
Deepfillv1[43]	2018	1. It can process images with multiple holes at arbitrary locations and with variable sizes.	L1 Loss 8.6%
ExGANs[44]	2018	1. Produce high-quality, personalized inpainting results.	L1 Loss 1.4%
Deepfillv2[45]	2018	1. Improve inpainting results with free-form masks and user guidance input. 1. Can deal with images with multiple, irregular shape missing regions. 2. The edge generating	L1 Loss 9.1%

TABLE 1. *(Continued.)* A brief summary of different GANs.

Edgeconnect[46]	2019	model sometimes fails to accurately depict the edges in highly textured areas, or when a large portion of the image is missing.	L1 Loss 3.86%
PEN-Net[47]	2019	1. Generate semantically-reasonable and visually-realistic results for image inpainting.	L1 Loss 9.94%
ELEGANT[51]	2018	1. Manipulate several attributes simultaneously. 2. Generate high-quality images with finer details and fewer artifacts.	FID 30.71
STGAN[53]	2019	1. Simultaneously improve attribute manipulation accuracy as well as perception quality. 2. Improve the image reconstruction quality and enhance the flexible translation of attributes.	N/A
SCGAN[56]	2019	1. It can control spatial contents, specify attributes and improve general visual quality.	N/A
Example-guided image synthesis[57]	2019	1. Produce realistic and style-consistent images. 2. The synthetic background in the face and dance image synthesis tasks may be blurry because the semantic labels do not specify any background scenes.	FID 31.26
SGGAN[58]	2019	1. Leverage semantic segmentation to further boost generation performance and provide the spatial mapping.	N/A
MaskGAN[59]	2019	1. Enable diverse and interactive face manipulation. 2. Enable diverse generation results. 3. More robust to various manipulated inputs.	FID 48.24
Text Guided Person Image Synthesis[63]	2019	1. It can interactively exert control over the process of person image generation by natural language descriptions.	SSIM 0.364
Progressive Pose Attention Transfer[64]	2019	1. Generated person images possess better appearance consistency and shape consistency with the input images. 2. Improve the computational efficiency and reduce the model complexity.	SSIM 0.311
Semantic Parsing Transformation[65]	2019	1. Enable a better semantic map prediction and further final results. 2. The model fails when errors exist in the condition semantic map.	SSIM 0.736
Coordinate-based Texture Inpainting[66]	2019	1. Allow reconstructing detail-rich textures.	SSIM 0.791

TABLE 1. *(Continued.)* **A brief summary of different GANs.**

CycleGAN[70]	2017	1. It can learn the mapping without a training set of aligned image pairs. 2. It fails when it requires geometric changes.	N/A
UNIT[71]	2017	1. Present high-quality image translation results on various challenging unsupervised image translation tasks. 2. The translation model is unimodal due to the Gaussian latent space assumption. 3. Training could be unstable due to the saddle point searching problem.	FID 66.84
MUNIT[73]	2018	1. Allow users to control the style of translation outputs by providing an example style image. 2. Achieve quality and diversity superior to existing unsupervised methods.	FID 27.60
DRIT[74]	2018	1. It can generate diverse and realistic images on a wide range of tasks without paired training data.	FID 31.04
TransGaGa[76]	2019	1. The translation is built on appearance and geometry space separately.	FID 23.23
RelGAN[77]	2019	1. It can modify images by changing particular attributes of interest in a continuous manner while preserving the other attributes.	FID 22.74
SC-FEGAN[82]	2019	1. Improve inpainting results. 2. Produce high quality and realistic results while requiring minimal efforts from the users.	N/A
FE-GAN[83]	2019	1. Enable users to manipulate the fashion image with an arbitrary sketch and a few sparse color strokes. 2. Achieve high-quality performance with convincing details.	FID 3.70
Mask-Guided Portrait Editing[84]	2019	1. It can synthesize diverse, high-quality, and controllable facial images from given masks.	FID 8.92
FaceShapeGene[85]	2019	1. Accomplish novel face editing tasks.	FID 14.22
CartoonGAN[88]	2018	1. Can generate high-quality cartoon images from real-world photos.	N/A
PI-REC[89]	2019	1. Generate abundant high-frequency details from sparse input information. 2. Guarantee users with free and accurate control over the content or style of images.	FID 0.015
Internal Representation Collaging [90]	2019	1. It allows the user to change the semantic information of an image over an arbitrary region by manipulating the feature-space representation of the image in a trained GAN model. 2. It cannot make expressions beyond the representation power of the used generator.	N/A
U-GAT-IT[91]	2019	1. It can translate both images requiring holistic changes and images requiring large shape changes. 2. Flexibly control the amount of change in shape and texture by learned parameters depending on datasets.	N/A
Landmark Assisted CycleGAN[92]	2019	1. Impressive high-quality cartoon faces and bitmojis are generated. 2. Generate high-quality images with relatively low resolution.	FID 1988.50

better results with realistic and clear images, which have attracted extensive attention. Due to the great potential and wide applicability of GANs, the researchers are constantly attracted to the research of GANs. However, there are still some problems that have not been completely solved in training and evaluating GANs, such as mode collapse, unstable training problem, and vanishing gradient problem. In addition, GANs are also faced with the problem of non-convergence and sensitivity to hyperparameters. At present, these problems remain to be solved, which need continuous research and efforts from the researchers, and many improved GAN-variants have emerged, including Least Square GAN (LSGAN) [101], Wasserstein GAN (WGAN) [102], WGAN-GP [103], and Spectral Normalized GANs (SNGAN) [104]. These models not only greatly improved the quality and the stability of GANs, but also make it easy to converge and aim to solve the problem of unstable training. However, the problem of collapse during training and the mode collapse have not been completely solved due to the high dimensional characteristics of image data. By using maximum likelihood pre-training, with the help of adversarial fine-tuning is now an effective solution to deal with mode collapse. Other techniques that can be used to stabilize and improve GANs training performance like large batch sizes, dense rewards and discriminator regularization [105]. Recently, Liu *et al.* [106] proposed an approach called WGAN-QC, which can stabilize and speed up the training process based on the quadratic transport cost. Petroski Such *et al.* [107] proposed an approach called Generative Teaching Networks (GTNs), and it can stabilize and prevent mode collapse of GAN training.

On the other hand, how to evaluate the quality of generated images still lacks effective means. Generative Adversarial

Networks are the most popular image generation methods today, but the way to evaluate and compare images produced by GANs is still an extremely challenging task. Many earlier studies on image synthesis based on GANs only used subjective visual assessments. Although it is very hard to quantify the quality of generated images, some studies of evaluating the GANs have begun to appear. For example, the Inception score (IS) [108] and Fréchet Inception Distance (FID) [109] are the most widely adopted evaluation metrics for quantitatively evaluating generated images. Besides, Bau *et al.* [110] proposed an approach to visualize and understand GANs at the scene-level. Zhou *et al.* [111] proposed a method called HUMAN EYE PERCEPTUAL EVALUATION (HYPE) to establish a gold standard human benchmark for generative realism. Grnarova *et al.* [112] propose an evaluation measure to monitor the training progress, which is able to detect failure modes, like unstable mode collapse and divergence. Other evaluation methods are studied as [113], [114].

With the successive development of methods for training and evaluation of GANs and the great progress that has been done on the GANs, the generative adversarial networks will be more and more widely used in various applications.

B. FUTURE OPPORTUNITIES

With the impetus of GANs, the research in the field of computer vision has been greatly developed in recent years, and various applications have emerged. Most of these applications involve image processing. Although there have been some studies involving video processing, such as video generation [115], video colorization [116], [117], video inpainting [118], motion transfer [119], and facial animation synthesis [120]–[123], the research on video using GANs is limited. In addition, although GANs have been applied to the generation and synthesis of 3D models, such as 3D colorization [124], 3D face reconstruction [125], [126], 3D character animation [127], and 3D textured object generation [128], the results are far from perfect. At present, GANs are still based on large amounts of training data. It is an inevitable trend to reduce the use of data in the future. Although there is already some weakly supervised learning research [129], [130], these are still very limited, and the results are far from optimal.

Besides, GANs have great potential in data augmentation, due to its ability to synthesize high-quality images, especially in areas with data paucity, such as medical image analysis [131]. Frid-Adar *et al.* [132] presented a GAN-based method to generate synthetic medical images for data augmentation, which can improve the performance of medical problems with limited data. Han *et al.* [133] proposed a two-step GAN-based data augmentation method to minimize the number of annotated images required for the medical imaging tasks. Sandfort *et al.* [134] used a GAN-based method for data augmentation to improve the performance of tasks in medical imaging. Han *et al.* [135] proposed a data augmentation approach called Conditional Progressive Growing of GANs (CPGGANs) to minimize expert physicians' annotation in medical applications. Schlegl *et al.* [136] presented an approach called fast AnoGAN (f-AnoGAN), which can identify anomalous images on a variety of biomedical data. Han *et al.* [137] proposed a data augmentation method called 3D Multi-Conditional GAN (MCGAN), and it can help to overcome medical data paucity.

Other directions that are equally noteworthy, such as in the modular [138] and game areas [139], have rarely been studied. Recently, Lin *et al.* [140] proposed a novel method called Conditional Coordinate GAN (COCO-GAN), which uses the spatial coordinates as the condition to generate images by parts, and it achieves a high generation quality. Particularly, this approach can generate images larger than any training sample and can be used for large field-of-view generation. We conclude that there are opportunities for future research and application on GANs, especially in these areas.

VIII. CONCLUSION

In this paper, we reviewed some basics of GANs and described some applications in the field of image synthesis based on GANs. The pros and cons of these GANs applications are also provided. Besides, we summarized the methods used in GANs applications which improved the performance of generated images. Although the research on GANs is becoming more and more mature, GANs are still faced with some challenges, such as unstable training and hard to evaluate, for which we introduced some methods for training and evaluating of GANs. We think there are some likely future research directions, such as video generation, facial animation synthesis, and 3D face reconstruction. The performance of GANs will continue to improve as various GAN-variants are proposed and GANs applications still need exploring. We expect more interesting applications based on GANs to appear in the future.

REFERENCES

[1] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–14. [Online]. Available: https://arxiv.org/abs/1312.6114

[2] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible 1×1 convolutions," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2018, pp. 236–245.

[3] D. Joo, D. Kim, and J. Kim, "Generating a fusion image: One's identity and Another's shape," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1635–1643.

[4] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–14, Jul. 2017.

[5] W. Nie, N. Narodytska, and A. Patel, "Relgan: Relational generative adversarial networks for text generation," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Aug. 2019, pp. 1–20. [Online]. Available: https://openreview.net/forum?id=rJedV3R5tm

[6] M. Majurski, P. Manescu, S. Padi, N. J. Schaub, N. A. Hotaling, G. C. Simon, and P. Bajcsy, "Cell image segmentation using generative adversarial networks, transfer learning, and augmentations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 1–22.

[7] R. Zhang, T. Pfister, and J. Li, "Harmonic unpaired Image-to-image translation," 2019, *arXiv:1902.09727*. [Online]. Available: http://arxiv.org/abs/1902.09727

[8] C. F. Baumgartner, L. M. Koch, K. C. Tezcan, and J. X. Ang, "Visual feature attribution using wasserstein GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8309–8319.

[9] G. Kwon, C. Han, and D.-S. Kim, "Generation of 3D Brain MRI Using Auto-Encoding Generative Adversarial Networks," in *Proc. Med. Image Comput. Comput. Assist. Intervent (MICCAI)*, 2019, pp. 118–126.
[10] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, and Y. Zheng, "X2CT-GAN: Reconstructing CT from biplanar X-Rays with generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 10611–10620.
[11] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.
[12] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin, "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1310–1321, Jun. 2018.
[13] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "SegAN: Adversarial network with multi-scale l1 loss for medical image segmentation," *Neuroinformatics*, vol. 16, nos. 3–4, pp. 383–392, Oct. 2018.
[14] X. Zhu, X. Zhang, X.-Y. Zhang, Z. Xue, and L. Wang, "A novel framework for semantic segmentation with generative adversarial network," *J. Vis. Commun. Image Represent.*, vol. 58, pp. 532–543, Jan. 2019.
[15] N. Souly, C. Spampinato, and M. Shah, "Semi supervised semantic segmentation using generative adversarial network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5688–5696.
[16] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," 2016, *arXiv:1611.08408*. [Online]. Available: http://arxiv.org/abs/1611.08408
[17] W. Zhu, X. Xiang, T. D. Tran, and X. Xie, "Adversarial deep structural networks for mammographic mass segmentation," 2016, *arXiv:1612.05970*. [Online]. Available: http://arxiv.org/abs/1612.05970
[18] K. Nazeri, E. Ng, and M. Ebrahimi, "Image colorization with generative adversarial networks," 2018, *arXiv:1803.05400*. [Online]. Available: http://arxiv.org/abs/1803.05400
[19] Y. Liu, Z. Qin, T. Wan, and Z. Luo, "Auto-painter: Cartoon image generation from sketch by using conditional wasserstein generative adversarial networks," *Neurocomputing*, vol. 311, pp. 78–87, Oct. 2018.
[20] S. Azadi, M. Fisher, V. Kim, Z. Wang, E. Shechtman, and T. Darrell, "Multi-content GAN for few-shot font style transfer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7564–7573.
[21] A. Elgammal, B. Liu, M. Elhoseiny, and M. Mazzone, "CAN: Creative adversarial networks, generating 'Art' by learning about styles and deviating from style norms," 2017, *arXiv:1706.07068*. [Online]. Available: https://arxiv.org/abs/1706.07068
[22] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning face age progression: A pyramid architecture of GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 31–39.
[23] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning continuous face age progression: A pyramid of GANs," 2019, *arXiv:1901.07528*. [Online]. Available: http://arxiv.org/abs/1901.07528
[24] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4352–4360.
[25] Y. Wang, A. Gonzalez-Garcia, J. van de Weijer, and L. Herranz, "SDIT: Scalable and diverse cross-domain image translation," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 1267–1276.
[26] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018, doi: 10.1109/MSP.2017.2765202.
[27] I. Goodfellow, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 2014, pp. 2672–2680 [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf
[28] L. J. Ratliff, S. A. Burden, and S. S. Sastry, "Characterization and computation of local Nash equilibria in continuous games," in *Proc. 51st Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Monticello, IL, USA, Oct. 2013, pp. 917–924.
[29] T. Park, M. Liu, T. Wang, and J. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 2337–2346.
[30] U. Bergmann, N. Jetchev, and R. Vollgraf, "Learning texture manifolds with the periodic spatial GAN," in *Proc. 34th Int. Conf. Mach. Learn.*, Aug. 2017, pp. 469–477. [Online]. Available: http://proceedings.mlr.press/v70/bergmann17a.html
[31] W. Xian, P. Sangkloy, V. Agrawal, A. Raj, J. Lu, C. Fang, F. Yu, and J. Hays, "TextureGAN: Controlling deep image synthesis with texture patches," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 8456–8465.
[32] N. Yu, C. Barnes, E. Shechtman, S. Amirghodsi, and M. Lukac, "Texture mixer: A network for controllable synthesis and interpolation of texture," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 12156–12165.
[33] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 702–716.
[34] N. Jetchev, U. Bergmann, and R. Vollgraf, "Texture synthesis with spatial generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 1–11.
[35] P. Stinis, T. Hagge, A. M. Tartakovsky, and E. Yeung, "Enforcing constraints for interpolation and extrapolation in generative adversarial networks," *J. Comput. Phys.*, vol. 397, Nov. 2019, Art. no. 08844, doi: 10.1016/j.jcp.2019.07.042.
[36] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*. [Online]. Available: https://arxiv.org/abs/1710.10196
[37] D. Kim, M. Kim, G. Kwon, and D.-S. Kim, "Progressive face super-resolution via attention to facial landmark," 2019, *arXiv:1908.08239*. [Online]. Available: https://arxiv.org/abs/1908.08239
[38] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2019, pp. 1–35. [Online]. Available: https://openreview.net/forum?id=B1xsqj09Fm
[39] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4401–4410.
[40] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 105–114.
[41] X. Wang, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshops (ECCVW)*, Sep. 2018, pp. 1–16.
[42] X. Wang, K. Yu, C. Dong, and C. Change Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 606–615.
[43] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA Jun. 2018, pp. 5505–5514.
[44] B. Dolhansky and C. C. Ferrer, "Eye in-painting with exemplar generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7902–7911.
[45] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Free-form image inpainting with gated convolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4471–4480.
[46] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, "EdgeConnect: Generative image inpainting with adversarial edge learning," 2019, *arXiv:1901.00212*. [Online]. Available: http://arxiv.org/abs/1901.00212
[47] Y. Zeng, J. Fu, H. Chao, and B. Guo, "Learning pyramid-context encoder network for high-quality image inpainting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1486–1494.
[48] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, "High-resolution image inpainting using multi-scale neural patch synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6721–6729.
[49] Y. R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 6882–6890.
[50] Y. Li, S. Liu, J. Yang, and M. Yang, "Generative face completion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 5892–5900.

[51] T. Xiao, J. Hong, and J. Ma, "ELEGANT: Exchanging latent encodings with GAN for transferring multiple face attributes," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 168–184.

[52] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.

[53] M. Liu, Y. Ding, M. Xia, X. Liu, E. Ding, W. Zuo, and S. Wen, "STGAN: A unified selective transfer network for arbitrary image attribute editing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Beach, CA, USA, Jun. 2019, pp. 3673–3682.

[54] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, "AttGAN: Facial attribute editing by only changing what you want," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5464–5478, Nov. 2019.

[55] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain Image-to-Image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8789–8797.

[56] S. Jiang, H. Liu, Y. Wu, and Y. Fu, "Spatially constrained generative adversarial networks for conditional image generation," (2019), *arXiv:1905.02320*. [Online]. Available: https://arxiv.org/abs/1905.02320

[57] M. Wang, G.-Y. Yang, R. Li, R.-Z. Liang, S.-H. Zhang, P. M. Hall, and S.-M. Hu, "Example-guided style-consistent image synthesis from semantic labeling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 1495–1504.

[58] S. Jiang, Z. Tao, and Y. Fu, "Segmentation guided Image-to-Image translation with adversarial networks," 2019, *arXiv:1901.01569*. [Online]. Available: http://arxiv.org/abs/1901.01569

[59] C.-H. Lee, Z. Liu, L. Wu, and P. Luo, "MaskGAN: Towards diverse and interactive facial image manipulation," 2019, *arXiv:1907.11922*. [Online]. Available: https://arxiv.org/abs/1907.11922

[60] J. Lin, Z. Chen, Y. Xia, S. Liu, T. Qin, and J. Luo, "Exploring explicit domain supervision for latent space disentanglement in unpaired image-to-image translation," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Oct. 29, 2019, doi: 10.1109/TPAMI.2019.2950198.

[61] R. Mokady, S. Benaim, L. Wolf, and A. Bermano, "Mask based unsupervised content transfer," 2019, *arXiv:1906.06558*. [Online]. Available: http://arxiv.org/abs/1906.06558

[62] W. Yin, Z. Liu, and C. Change Loy, "Instance-level facial attributes transfer with geometry-aware flow," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, Jul. 2019, pp. 9111–9118.

[63] X. Zhou, S. Huang, B. Li, Y. Li, J. Li, and Z. Zhang, "Text guided person image synthesis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3663–3672.

[64] Z. Zhu, T. Huang, B. Shi, M. Yu, B. Wang, and X. Bai, "Progressive pose attention transfer for person image generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 2342–2351.

[65] S. Song, W. Zhang, J. Liu, and T. Mei, "Unsupervised person image generation with semantic parsing transformation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 2352–2361.

[66] A. Grigorev, A. Sevastopolsky, A. Vakhitov, and V. Lempitsky, "Coordinate-based texture inpainting for pose-guided human image generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 12127–12136.

[67] H. Tang, D. Xu, G. Liu, W. Wang, N. Sebe, and Y. Yan, "Cycle in cycle generative adversarial networks for keypoint-guided image generation," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2052–2060.

[68] L. Ma, X. Jia, Q. Sun, B. Schiele, T. Tuytelaars, and L. V. Gool, "Pose guided person image generation," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 406–416.

[69] L. Ma, Q. Sun, S. Georgoulis, L. Van Gool, B. Schiele, and M. Fritz, "Disentangled person image generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 99–108.

[70] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[71] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 700–708.

[72] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2016, pp. 469–477.

[73] M.-Y. L. Xun Huang, J. Serge Belongie, and A. JanKautz, "Multimodal unsupervised image-to-image translation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 172–189.

[74] H.-Y. T. Hsin-Ying Lee, J.-B. Huang, M. Singh, and M.-H. Yang, "Diverse image-to-image translation via disentangled representations," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 35–51.

[75] H.-Y. Lee, H.-Y. Tseng, Q. Mao, J.-B. Huang, Y.-D. Lu, M. Singh, and M.-H. Yang, "DRIT++: Diverse Image-to-Image translation via disentangled representations," 2019, *arXiv:1905.01270*. [Online]. Available: http://arxiv.org/abs/1905.01270

[76] W. Wu, K. Cao, C. Li, C. Qian, and C. C. Loy, "TransGaGa: Geometry-aware unsupervised image-to-image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 8004–8013.

[77] Y.-J. Lin, P.-W. Wu, C.-H. Chang, E. Chang, and S.-W. Liao, "RelGAN: Multi-domain Image-to-Image translation via relative attributes," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul South Korea, Oct. 2019, pp. 5913–5921.

[78] X. Li, J. Hu, S. Zhang, X. Hong, Q. Ye, C. Wu, and R. Ji, "Attribute guided unpaired image-to-image translation with semi-supervised learning," 2019, *arXiv:1904.12428*. [Online]. Available: https://arxiv.org/abs/1904.1 2428

[79] S. Chang, S. Park, J. Yang, and N. Kwak, "SYM-parameterized dynamic inference for mixed-domain image translation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul South Korea, Oct. 2019, pp. 4803–4811.

[80] M. Tomei, M. Cornia, L. Baraldi, and R. Cucchiara, "Art2Real: Unfolding the reality of artworks via semantically-aware image-to-image translation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 5842–5852.

[81] S. Mo, M. Cho, and J. Shin, "InstaGAN: Instance-aware image-to-image translation," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–70.

[82] Y. Jo and J. Park, "SC-FEGAN: Face editing generative adversarial network with User's sketch and color," 2019, *arXiv:1902.06838*. [Online]. Available: http://arxiv.org/abs/1902.06838

[83] H. Dong, X. Liang, Y. Zhang, X. Zhang, Z. Xie, B. Wu, Z. Zhang, X. Shen, and J. Yin, "Fashion editing with adversarial parsing learning," 2019, *arXiv:1906.00884*. [Online]. Available: http://arxiv.org/abs/1906.00884

[84] S. Gu, J. Bao, H. Yang, D. Chen, F. Wen, and L. Yuan, "Mask-guided portrait editing with conditional GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 3436–3445.

[85] S.-Z. Xu, H.-Z. Huang, S.-M. Hu, and W. Liu, "Face shape gene: A disentangled shape representation for flexible face image editing," (2019), *arXiv:1905.01920*. [Online]. Available: https://arxiv.org/abs/1905.01920

[86] Y. Shen, J. Gu, X. Tang, and B. Zhou, "Interpreting the latent space of GANs for semantic face editing," 2019, *arXiv:1907.10786*. [Online]. Available: http://arxiv.org/abs/1907.10786

[87] T. Portenier, Q. Hu, A. Szabó, S. A. Bigdeli, P. Favaro, and M. Zwicker, "FaceShop: Deep Sketch-based Face Image Editing," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 99:1-99:13, Jul. 2018.

[88] Y. Chen, Y. Lai, and Y. Liu, "CartoonGAN: Generative adversarial networks for photo cartoonization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 9465–9474.

[89] S. You, N. You, and M. Pan, "PI-REC: Progressive image reconstruction network with edge and color domain," 2019, *arXiv:1903.10146*. [Online]. Available: http://arxiv.org/abs/1903.10146

[90] R. Suzuki, M. Koyama, T. Miyato, T. Yonetsuji, and H. Zhu, "Spatially controllable image synthesis with internal representation collaging," 2018, *arXiv:1811.10153*. [Online]. Available: http://arxiv.org/abs/1811.10153

[91] J. Kim, M. Kim, H. Kang, and K. Lee, "U-GAT-IT: Unsupervised generative attentional networks with adaptive layer-instance normalization for Image-to-Image translation," 2019, *arXiv:1907.10830*. [Online]. Available: http://arxiv.org/abs/1907.10830

[92] R. Wu, X. Gu, X. Tao, X. Shen, Y.-W. Tai, and J. iaya Jia, "Landmark assisted CycleGAN for cartoon face generation," 2019, *arXiv:1907.01424*. [Online]. Available: http://arxiv.org/abs/1907.01424

[93] Y. Taigman, A. Polyak, and L. Wolf, "Unsupervised cross-domain image generation," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–12.

[94] Y. Jin, J. Zhang, M. Li, Y. Tian, H. Zhu, and Z. Fang, "Towards the automatic Anime characters creation with generative adversarial networks," 2017, *arXiv:1708.05509*. [Online]. Available: http://arxiv.org/abs/1708.05509
[95] J. Li, "Twin-GAN - unpaired cross-domain image translation with weight-sharing GANs," 2018, *arXiv:1809.00946*. [Online]. Available: http://arxiv.org/abs/1809.00946
[96] S. Ma, J. Fu, C. W. Chen, and T. Mei, "DA-GAN: Instance-level image translation by deep attention generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 5657–5666.
[97] K. Hamada, K. Tachibana, T. Li, H. Honda, and Y. Uchida, "Full-body High-resolution Anime Generation with Progressive Structure-conditional Generative Adversarial Networks," in *Proc. Eur. Conf. Comput. Vis. Workshops (ECCVW)*, 2018, p. 1–4.
[98] A. Gokaslan, V. Ramanujan, D. Ritchie, K. I. Kim, and J. Tompkin, "Improving shape deformation in unsupervised image-to-image translation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 649–665.
[99] K. Cao, J. Liao, and L. Yuan, "CariGANs: Unpaired Photo-to-Caricature translation," 2018, *arXiv:1811.00222*. [Online]. Available: http://arxiv.org/abs/1811.00222
[100] S. Xiang and H. Li, "Disentangling style and content in Anime illustrations," 2019, *arXiv:1905.10742*. [Online]. Available: https://arxiv.org/abs/1905.10742
[101] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2794–2802.
[102] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," in *Proc. Int. Conf. Mach. Learn.*, vol. 70, Aug. 2017, pp. 214–223 [Online]. Available: http://proceedings.mlr.press/v70/arjovsky17a.html
[103] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 5767–5777. [Online]. Available: http://papers.nips.cc/paper/7159-improved-training-of-wassersteingans.pdf
[104] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–26. [Online]. Available: https://openreview.net/forum?id=B1QRgziT
[105] C. D. M. d'Autume, M. Rosca, W. Jack Rae, and S. Mohamed, "Training language GANs from scrat-ch," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2019, pp. 4302–4313. [Online]. Available: http://papers.nips.cc/paper/8682-training-language-gans-from-scratch.pdf
[106] H. Liu, X. Gu, and D. Samaras, "Wasserstein GAN with quadratic transport cost," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul South Korea, Oct. 2019, pp. 4831–4840.
[107] F. Petroski Such, A. Rawal, J. Lehman, K. O. Stanley, and J. Clune, "Generative teaching networks: Accelerating neural architecture search by learning to generate synthetic training data," 2019, *arXiv:1912.07768*. [Online]. Available: http://arxiv.org/abs/1912.07768
[108] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2016, pp. 2226–2234.
[109] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, G. Klambauer, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 6626–6637. [Online]. Available: http://papers.nips.cc/paper/7240-gans-trained-by-a-two-time-scale-update-rule-converge-to-a-local-nash-equilibrium.pdf
[110] D. Bau, J.-Y. Zhu, H. Strobelt, B. Zhou, B. Joshua Tenenbaum, T. William Freeman, and A. Torralba, "GAN dissection: Visualizing and understanding generative adversarial networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2019, pp. 1–18.
[111] S. Zhou, M. Gordon, R. Krishna, A. Narc-Omey, D. Morina, and M. S. Bernstein, "HYPE: A benchmark for human eYe perceptual evaluation of generati- ve models," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2019, pp. 3444–3456. [Online]. Available: http://papers.nips.cc/paper/8605-hype-a-benchmark-for-human-eye-perceptual-evaluation-of-generative-models.pdf
[112] P. Grnarova, K. Y. Levy, A. Lucchi, N. P. Erraudin, I. Goodfellow, T. Hofmann, and A. K. Rause, "A domain agnostic measure for monitoring and eva-luating GANs," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2019, pp. 12069–12079. [Online]. Available: http://papers.nips.cc/paper/9377-a-domain-agnostic-measure-for-monitoring-and-evaluating-gans.pdf
[113] S. Benaim, T. Galanti, and L. Wolf, "Estimating the success of unsupervised image to image translation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 218–233.
[114] Shmelkov, Konstantin, C. Schmid, and K. Alahari, "How good is my GAN?" in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 213–229.
[115] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, G. Liu, A. Tao, J. Kautz, and B. Catanzaro, "Video-to-video synthesis," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2018, pp. 1–14.
[116] P. Kouzouglidis, G. Sfikas, and C. Nikou, "Automatic video colorization using 3D conditional generative adversarial networks," 2019, *arXiv:1905.03023*. [Online]. Available: http://arxiv.org/abs/1905.03023
[117] H. Thasarathan, K. Nazeri, and M. Ebrahimi, "Automatic temporally coherent video colorization," 2019, *arXiv:1904.09527*. [Online]. Available: http://arxiv.org/abs/1904.09527
[118] Y. Chang, Z. Y. Liu, K. Lee, and W. Hsu, "Free-form video inpainting with 3D gated convolution and temporal patch GAN," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9066–9075.
[119] C. Chan, S. Ginosar, T. Zhou, and A. Efros, "Everybody dance now," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)* Seoul South Korea, Nov. 2019, pp. 5932–5941.
[120] E. Zakharov, A. Shysheya, E. Burkov, and V. Lempitsky, "Few-shot adversarial learning of realistic neural talking head models," 2019, *arXiv:1905.08233*. [Online]. Available: http://arxiv.org/abs/1905.08233
[121] K. Vougioukas, S. Petridis, and M. Pantic, "End-to-end speech-driven realistic facial animation with temporal GANs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 37–40.
[122] H. Zhou, Y. Liu, Z. Liu, P. Luo, and X. Wang, "Talking face generation by adversarially disentangled audio-visual representation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, Jul. 2019, pp. 9299–9306.
[123] L. Chen, R. K. Maddox, Z. Duan, and C. Xu, "Hierarchical cross-modal talking face generation with dynamic pixel-wise loss," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7832–7841.
[124] Z. Yang, L. Liu, and Q.-X. Huang, "Learning generative neural networks for 3D colorization," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2018, pp. 2580–2587.
[125] B. Gecer, S. Ploumpis, I. Kotsia, and S. Zafeiriou, "GANFIT: Generative adversarial network fitting for high fidelity 3D face reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 1155–1164.
[126] A. Chen, Z. Chen, G. Zhang, K. Mitchell, and J. Yu, "Photo-realistic facial details synthesis from single image," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9429–9439.
[127] C. Weng, B. Curless, and I. Kemelmacher-Shlizerman, "Photo wake-up: 3D character animation from a single photo," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5908–5917.
[128] W. Chen, J. Gao, H. Ling, E. Smith, J. Lehtinen, A. Jacobson, and S. Fidler, "Learning to Pred-ict 3D objects with an interpolation-based differentiable Ren-Derer," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2019, pp. 9605–9616. [Online]. Available: http://papers.nips.cc/paper/9156-learning-to-predict-3d-objects-with-an-interpolation-based-differentiable-renderer.pdf
[129] M.-Y. Liu, X. Huang, A. Mallya, T. Karras, T. Aila, J. Lehtinen, and J. Kautz, "Few-shot unsupervised Image-to-Image translation," 2019, *arXiv:1905.01723*. [Online]. Available: http://arxiv.org/abs/1905.01723
[130] J. Lin, Y. Xia, S. Liu, T. Qin, and Z. Chen, "ZstGAN: An adversarial approach for unsupervised zero-shot Image-to-Image translation," 2019, *arXiv:1906.00184*. [Online]. Available: http://arxiv.org/abs/1906.00184
[131] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101552, doi: 10.1016/j.media.2019.101552.
[132] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018, doi: 10.1016/j.neucom.2018.09.013.
[133] C. Han, L. Rundo, R. Araki, Y. Nagano, Y. Furukawa, G. Mauri, H. Nakayama, and H. Hayashi, "Combining noise-to-image and image-to-image GANs: Brain MR image augmentation for tumor detection," *IEEE Access*, vol. 7, pp. 156966–156977, 2019, doi: 10.1109/ACCESS.2019.2947606.

[134] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, "Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, Art. no. 16884, doi: 10.1038/s41598-019-52737-x.
[135] C. Han, K. Murao, T. Noguchi, Y. Kawata, F. Uchiyama, L. Rundo, H. Nakayama, and S. Satoh, "Learning more with less: Conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2019, pp. 119–127.
[136] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "F-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Med. Image Anal.*, vol. 54, pp. 30–44, May 2019, doi: 10.1016/j.media.2019.01.010.
[137] C. Han, Y. Kitamura, A. Kudo, A. Ichinose, L. Rundo, Y. Furukawa, K. Umemoto, Y. Li, and H. Nakayama, "Synthesizing diverse lung nodules wherever massively: 3D multi-conditional GAN-based CT image augmentation for object detection," in *Proc. Int. Conf. 3D Vis. (3DV)*, Québec City, QC, Canada, Sep. 2019, pp. 729–737.
[138] B. Zhao, B. Chang, Z. Jie, and L. Sigal, "Modular generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 150–165.
[139] T. Shi, Y. Yuan, C. Fan, Z. Zou, Z. Shi, and Y. Liu, "Face-to-parameter translation for game character auto-creation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul South Korea, Oct. 2019, pp. 161–170.
[140] C. H. Lin, C.-C. Chang, Y.-S. Chen, D.-C. Juan, W. Wei, and H.-T. Chen, "COCO-GAN: Generation by parts via conditional coordinating," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4512–4521.

LEI WANG received the B.S. degree in network engineering from the Nanjing University of Information Science and Technology, Nanjing, China, in 2015. He is currently pursuing the M.S. degree in computer science and technology with the China University of Mining and Technology, Xuzhou, China. His research interests include computer vision, and machine learning.

WEI CHEN (Member, IEEE) received the B.Eng. degree in medical imaging and the M.S. degree in paleontology and stratigraphy from the China University of Mining and Technology, Xuzhou, China, in 2001 and 2005, respectively, and the Ph.D. degree in communications and information systems from the China University of Mining and Technology, Beijing, China, in 2008. In 2008, he joined the School of Computer Science and Technology, China University of Mining and Technology, where he is currently a Professor. His research interests include machine learning, image processing, and computer networks. He is a member of ACM and EAI.

WENJIA YANG received the B.Eng. and M.S. degrees in computer science from the China University of Mining and Technology, Xuzhou, China, in 2003 and 2007, respectively. He currently works with the School of Computer Science and Technology, China University of Mining and Technology, where he is currently an Assistant Professor. His research interests include machine learning, program analysis, and computer networks.

FANGMING BI received the B.S. degree in geophysical, the M.S. degree in computer science and technology, and the Ph.D. degree in cartography and geographic information system from the China University of Mining and Technology, Xuzhou, China, in 1997, 2002, and 2010, respectively. He joined the School of Computer Science and Technology, China University of Mining and Technology, where he is currently an Associate Professor. His research interests include intelligent information processing, spatial information security, and big data processing.

FEI RICHARD YU (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of British Columbia (UBC), in 2003. From 2002 to 2006, he was with Ericsson, Lund, Sweden, and a start-up in California, USA. He joined Carleton University, in 2007, where he is currently a Professor. He received the IEEE Outstanding Service Award in 2016, the IEEE Outstanding Leadership Award in 2013, the Carleton Research Achievement Award in 2012, the Ontario Early Researcher Award (formerly Premiers Research Excellence Award) in 2011, the Excellent Contribution Award at IEEE/IFIP TrustCom 2010, the Leadership Opportunity Fund Award from Canada Foundation of Innovation in 2009, and the Best Paper Awards at IEEE ICNC 2018, VTC 2017 Spring, ICC 2014, Globecom 2012, IEEE/IFIP TrustCom 2009 and the Int'l Conference on Networking 2005. His research interests include wireless cyber-physical systems, connected/autonomous vehicles, security, distributed ledger technology, and deep learning. He serves on the editorial boards of several journals, including co-editor-in-chief for *Ad Hoc and Sensor Wireless Networks*, lead series editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, and the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS. He has served as the Technical Program Committee (TPC) Co-Chair of numerous conferences. He is a registered Professional Engineer in the province of Ontario, Canada, and a Fellow of the Institution of Engineering and Technology (IET). He is a Distinguished Lecturer, the Vice President (Membership), and an elected member of the Board of Governors (BoG) of the IEEE Vehicular Technology Society.