A review of style migration based on GAN networks

Lingyu Yang
yang.lingyu@northeastern.edu

Jiayun Xin
xin.jiayun@northeastern.edu

Yihao Huang
huang.yihao@northeastern.edu

Abstract

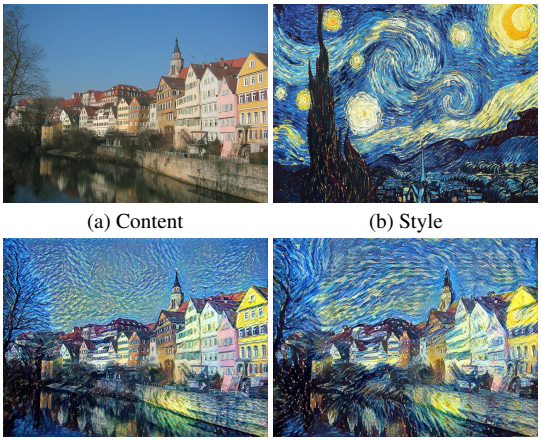
In this article, we aim to review a basic model and 4 image style transfer techniques based on GAN networks including supervised and unsupervised models. We first introduce the GAN basic model, a deep learning model proposed by Goodfellow et al in 2014, which is trained in an adversarial setting deep neural network. Since then, the process of using GAN to render a style-transferred image has replaced the traditional CNN method and has become a popular model both in academic literature and industrial applications. Pix2Pix and StarGAN are two supervised image-to-image translation models where Pix2Pix trains a generator and a discriminator in a given direction while StarGAN trains the single generator to map multiple domains. Then, we describe CycleGAN and MUNIT which are two unsupervised models avoiding paired dataset input to carry out the image-style transfer. CycleGAN proposed cycle consistency loss to preserve the key image content with a fixed transfer direction, while MUNIT achieves image-to-image translation into multimodal domains. Finally, a discussion of various applications of GANs in recent years and open problems for future research is concluded.

1. Introduction

Painting is one of the art types. It means that only human beings could do this. Additionally, different artists generate their own unique styles, especially, since some of their paintings are so extraordinary that they gradually become very famous, such as Van Gogh, Da Vinci, etc. However, with the development of computer vision, scientists have begun to study how to use computer technology to automatically turn ordinary images into artistic paintings. At that moment, the concept of image style transfer has arisen.

Image style transfer is the process of rendering an image without changing the content of the image but adjusting its texture, color, and so on. Earlier style transfer includes Non-Photorealistic Rendering(NPR) [6], and Texture Transfer. Non-photorealistic rendering is the means of generating imaging that does not aspire to realism. Also, NPR is able to simulate digitally different kinds of specific artistic styles. However, NPR transfer models could only generate one image and transfer it according to one style.

Style transfer is usually studied as a problem of texture that is homogeneous and consists of repeated elements, often subject to some randomization in their location, size, color, orientation, etc. In other words, Texture can be described by a statistical model of the local features of the image. Julesz pioneered the statistical characterization of textures by hypothesizing that the Nth-order joint empirical densities of image pixels (for some unspecified N), could be used to partition textures into classes that are preattentive indistinguishable from a human observer [24].



(a) Content (b) Style

(c) Small Stroke (d) Large Stroke

Figure 2. Control the brush stroke size in NST [11]

Hertzmann [7] further proposes a framework named image analogies to perform a generalized style transfer by learning the analogous transformation from the provided example pairs of unstyled and stylized images. Nevertheless, the common limitation of these methods is that they only consider low-level detailed features, instead of ex-

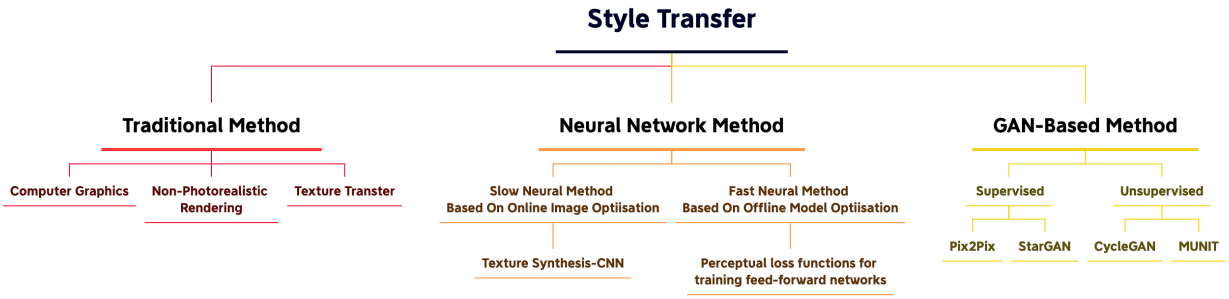


Figure 1. Style Transfer

tracting high-level semantic representation and often fail to transfer style corresponding to semantic content [11]. Aiming to resolve these weaknesses, neural style transfer has arisen.

Gatys. etc [4] found that deep neural networks could simultaneously extract image texture information and semantic information and pioneer to employ of VGG networks to effectively realize style transfer. This model accepts two inputs: one image is used to provide content, and another is used to offer style. Then, compute the loss between the image of generative content and the image of style. The goal of learning is a minimum loss function and the outcome is a new picture. In 2016, Johnson et al. [12] first proposed an image style transfer method for iteratively optimized generative models, that is, fast style transfer. After this, multiple sorts of generative adversarial networks, such as Coupled Generative Adversarial Networks [19] written by Ming-Yu Liu and Perceptual Generative Adversarial Networks [27] proposed by Bingzhe Wu have arisen and propagated, which make image-to-image style transfer improved extremely. It is different from learning only one image style that GANs are more able to learn the style from a cluster of images and capture analogous features.

With the rapid development of GANs in computer vision, the method of style transfer has been mature, such as Pix2Pix [10], CycleGAN [32], StarGAN [2], MUNIT [8] algorithms. These approaches are able to perform very well in many datasets. So we decide to overview the existing image style transfer algorithms based on GANs.The contributions of this review are as follows:

Firstly, we make a statement of our topic and generalize early methods and development of style transfer. Then, we introduce generative adversarial networks(GANs) that were rather important to the advancement of style transfer.

Secondly, we detailedly introduce the classic model of GAN from several aspects: the principle, learning method, and evaluation methods. In addition, we presented other GAN-based style migration models regarding supervised and unsupervised methods.

At last, we introduced the applicational scenarios of different models and analysis of model advantages and disadvantages.

2. GAN

A generative adversarial network (GAN) [9] is a deep generative model proposed by Goodfellow et al. in 2014, which learns to capture real data distributions through an adversarial process. GAN is structurally inspired by the two-person zero-sum game in game theory [20] and usually consists of a generator G and a discriminator D, as shown in Figure 1.

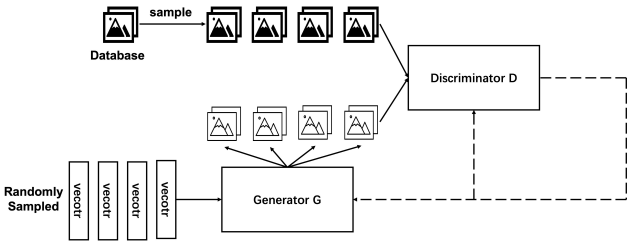


Figure 3. GAN algorithm

2.1. GAN Basic Principle

The generator captures the potential distribution of real data samples and generates new data samples; the discriminator is a binary classifier that discriminates whether the input is real data or generated samples. Both the generator and the discriminator can be used in deep neural networks, which are currently a hot research activity.

Any differentiable function can be used to represent the generator and discriminator of GAN, thus, we use the differentiable functions D and G to represent the discriminator and generator, respectively, whose inputs are the real data x and the random variable z.

$G(z)$ is a sample generated by G that obeys the true data distribution p_{data} as much as possible. If the input to the discriminator is from the real data, it is labeled as 1. If the input sample is $G(z)$, it is labeled as 0.

The goal of D is to achieve binary discrimination between the sources of the data: true (from the distribution of the real data x) or pseudo (from the pseudo data $G(z)$ of the generator), while the goal of G is to make the performance $D(G(z))$ of its generated pseudo data $G(z)$ on D the same as the performance $D(x)$ of the real data x on D.

These two processes confront each other and iterate When the discriminative power of D is finally improved to a certain level and the data source cannot be correctly discriminated, the generator G is considered to have learned the distribution of the real data.

2.2. GAN Learning Method

The core idea of GAN is based on the Nash equilibrium of game theory. The goal of the generator is to learn the real data distribution as much as possible, while the goal of the discriminator is to correctly discriminate whether the input data comes from the real data or the generator.

To win the game, the two players need to continuously optimize, each improving their own generative and discriminative abilities, and this learning optimization process is to find a Nash equilibrium between the two.

Therefore, the optimization problem of GAN is a minimal-maximization problem, and the objective function of GAN can be described as follows:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(x)}[\log(1-D(G(z)))] \quad (1)$$

For the GAN learning process, we need to train model D to maximize the accuracy of discriminating data from real data or pseudo-data distribution $G(z)$, and we need to train model G to minimize $log(1-D(G(z)))$.

Here we can use an alternating optimization method: first fix the generator G and optimize the discriminator D to maximize the discriminative accuracy of D; then fix the discriminator D and optimize the generator G to minimize the discriminative accuracy of D.

The global optimal solution is reached when and only when $p_{data} = p_g$. When training the GAN, the parameters of D are updated k times and the parameters of G are updated 1 time in the same round.The specific training process is shown in Algorithm 1.

2.3. GAN Evaluation

In the wave of artificial intelligence frenzy, GAN has been proposed to meet the research and application needs of many fields and to inject new momentum into these fields. In an interview, the famous scholar Yann LeCun mentioned GAN as "the most exciting idea in machine learning in the last decade".

Currently, the field of image and vision is one of the most widely researched and applied fields of GAN, which can generate digital and human face objects, form various realistic indoor and outdoor scenes, recover original images from segmented images, colorize black and white images, recover object images from object outlines, and generate high-resolution images from low-resolution images.

Although GANs is a powerful deep generative model, there are several problems with the training of GANs, such as pattern collapse and training instability, as well as low resolution and quality of the generated images.

Next, we will briefly describe several classical models based on GAN model improvements for the image style migration problem.

Algorithm 1 GAN algorithm [9]

for number of training iterations **do**
  for k steps **do**
    •Sample minibatch of m noise samples $z^{(1)},...,z^{(m)}$ from noise prior $p_g(z)$.
    •Sample minibatch of m examples $x^{(1)},...,x^{(m)}$ from data generating distribution $p_{data}(x)$.
    •Update the discriminator by ascending its stochastic gradient:
    $\bigtriangledown_{\theta_d}\frac{1}{m}\sum_{i=1}^{m}[\log D(x^{(i)}) + \log(1-D(G(z^{(i)})))]$
  end for
  •Sample minibatch of m noise samples $z^{(1)},...,z^{(m)}$ from noise prior $p_g(z)$.
  •Update the generator by descending its stochastic gradient:
  $\bigtriangledown_{\theta_g}\frac{1}{m}\sum_{i=1}^{m}\log(1-D(G(z^{(i)})))$
end for

3. GAN-based classical model

Researchers have been exploring and eventually proposing many GAN-based variants to address the problems of difficult training and lack of constraints in GAN.

3.1. Supervised GAN

Since the GAN model has no control over the generated data, especially in the case of data with multiple tag classes. Therefore, additional information is needed to guide the direction of the distribution to direct the generated results to more than one tag class.

3.1.1 Pix2Pix

To control the data generation process by supervision, Mirza et al. proposed the conditional GAN (CGAN) [21] model. The data generation process is guided by adding constraints to the original GAN to make the network generate samples in a given direction.

Pix2Pix [10] is a supervised image-to-image translation method based on conditional generative adversarial net-

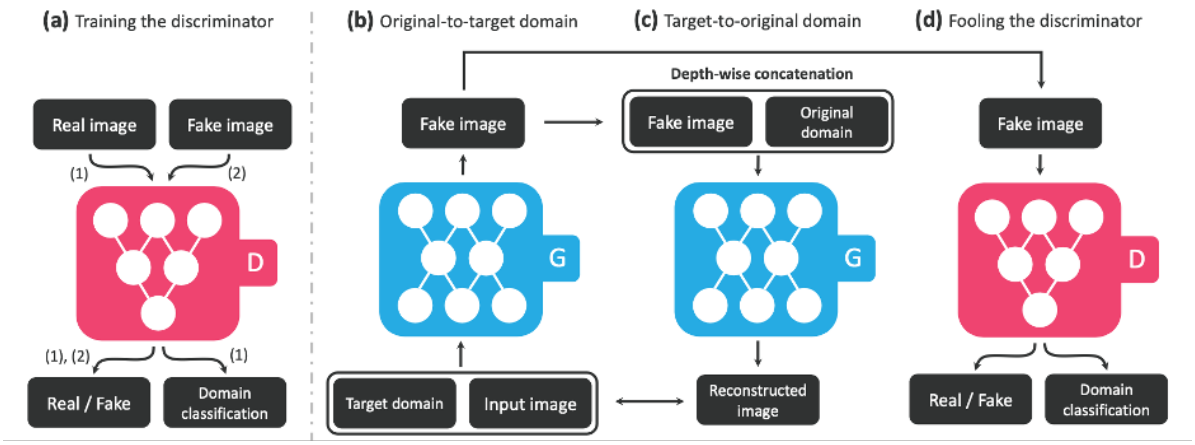


Figure 4. StarGAN algorithm [2]

works. Pix2Pix requires pairs of images to learn a one-to-one mapping and uses two datasets: one dataset as input and the other as conditional input.

The generator uses a U-Net [25] based architecture, which relies on skip connections at each layer. In contrast, the discriminator uses a convolution-based PatchGAN as a classifier. The objective function of Pix2Pix uses a cGAN:

$$\begin{aligned}\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_y[\log D(y)] \\ + \mathbb{E}_{x,z}[\log(1 - D(G(x,z))\end{aligned} \tag{2}$$

with an $\mathcal{L}_1$ criterion ($\mathcal{L}_1(G) = E_{x,y,z}[\|y - G(x,z)\|_1]$) instead of an $\mathcal{L}_2$ [23] criterion, which results in less ambiguity. The final objective function is as follows:

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda\mathcal{L}_1(G) \tag{3}$$

Pix2Pix works by using paired training data with different configurations of the GAN model. It enables the implementation of image transformation operations such as semantic map to street view, image coloring, sketch to real image, etc.

3.1.2 StarGAN

Models such as Pix2Pix [10] only address image transformation between two domains, and it has very limited scalability and robustness when dealing with more than two domains. Therefore Choi et al. proposed the StarGAN model in 2018 to solve the training problem across domains and multiple datasets. The contributions conclude the novel GAN [9] network, the use of the mask vector method, and superior qualitative and quantitative results on facial attribute transfer and facial expression synthesis tasks.

In StarGAN [2], the traditional fixed translation is quit, but the domain information, and pictures are input together for training The mask vector is added to the domain label, which is convenient for different training sets to do joint training. As a solution to solve traditional cross-domain models, the StarGAN model takes multiple domains within a single generator with a label to represent domain information.

Choi et al proposed a novel framework to train a single generator to map multiple domain image-to-image translation. As shown in figure, there are two modules in total. The discriminator D is trained by inputting real and fake images to classify them into the corresponding domain. The Generator G is trained by adding the target domain to construct image fooling the discriminator. To achieve an ideal generator and discriminator, this StarGAN paper introduced three loss functions. Adversarial Loss is the same as the loss function in cGAN [21].

$$\mathcal{L}_{cls}^r = \mathbb{E}_{x,c'}\left[-\log D_{cls}\left(c' \mid x\right)\right] \tag{4}$$

$$\mathcal{L}_{cls}^f = \mathbb{E}_{x,c}\left[-\log D_{cls}(c \mid G(x,c))\right] \tag{5}$$

The domain classification loss function is divided into two parts as shown above. When training the discriminant network D, use the supervision signal of the real picture in the original field for training. When training the generation network G, use the supervision of the generated picture in the target field.

$$\mathcal{L}_{rec} = \mathbb{E}_{x,c,c'}\left[\|x - G\left(G(x,c),c'\right)\|_1\right] \tag{6}$$

The reconstruction loss function is shown above and designed to preserve the content of the input images during style transfer because minimizing the adversarial loss and domain classification loss cannot guarantee content integrity. Finally, the formula combines all the loss functions as a full objective.

The StarGAN achieves image-to-image translation across multiple domains and successfully produces high-

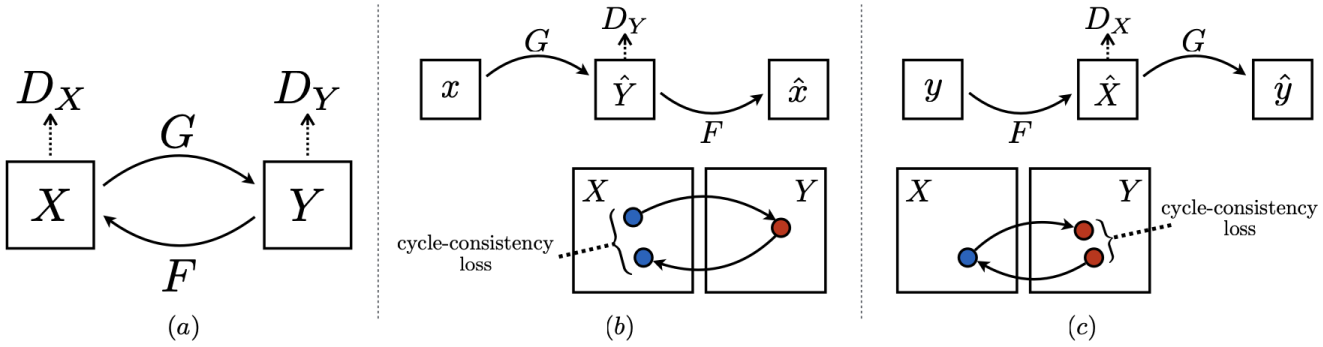


Figure 5. CycleGAN algorithm [32]

visual-quality images. Besides, the use of mask vectors allows StarGAN to combine multiple datasets with different sets of domain labels.

3.2. Unsupervised GAN

Since this supervised learning algorithm requires training on pre-processed paired datasets, in many cases, such perfect paired datasets are not available. Therefore, the study of unsupervised learning-based style migration GAN continues in depth.

3.2.1 CycleGAN

Cycle-consistent Generative Adversarial Network (CycleGAN) [32] was proposed by Zhu et al. in 2017. CycleGAN uses two unidirectional GANs to form a loop GAN structure, and uses cycle consistency loss to preserve the key information of input and transformed images, which solves the problem of needing paired training data for image style migration. The algorithm network structure is shown in Figure 2.

Two generators (G, F) are used to transform each other between two domains (X, Y) and two discriminators (Dx, Dy) are used to discriminate the authenticity of the images in the two domains.

The objective function of CycleGAN contains two terms: adversarial loss and cycle-consistency loss.

The adversarial loss [9] is designed to match the distribution of the generated images with the distribution of the target images, and is formulated as follows:

$$\begin{aligned}\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y\sim p_{data}(y)}[\log D_Y(y)] \\ + \mathbb{E}_{x\sim p_{data}(x)}[\log(1 - D_Y(G(x))]\end{aligned} \tag{7}$$

The cycle-consistency loss consists of forward and backward cyclic consistency terms, which aim to prevent the learned mappings G and F from contradicting each other. The formula is as follows:

$$\begin{aligned}\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x\sim p_{data}(x)}[\|F(G(x)) - x\|_1] \\ + \mathbb{E}_{y\sim p_{data}(y)}[\|G(F(y)) - y\|_1]\end{aligned} \tag{8}$$

With its special network setup, CycleGAN achieves the result of producing amazing style migration without training data pairing. It has a similarity with DiscoGAN [15] and DualGAN [30], which were published at the same time.

3.2.2 MUNIT

Although CycleGAN, DiscoGAN, and other popular computer vision models are designed to solve transfer problems from one domain to another, such as super-resolution [3], coloring [31], picture-in-picture [23], attribute transfer [16], and style transfer [5]. However, many cross-domain interest mappings are multimodal. MUNIT [8] research focus on study whether unsupervised image-to-image transformations can be applied to multimodal domains.

To address this limitation, Huang et al proposed Multimodal Unsupervised Image-to-image Translation (MUNIT) framework. Their primary assumptions are that an image can be decomposed into a content space and a style space and different domains share the same content space. For the conversion of a source domain image to the target domain, what MUNIT does is recombine the content code and style code. The MUINT model is shown below.

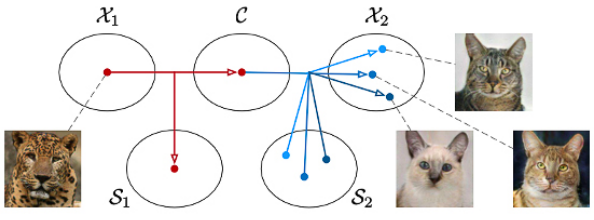


Figure 6. MUNIT algorithm [8]

The figure shows an abstract encoding part of MUNIT. X1 refers to all the input images, which are encoded into a domain-shared content space C and a domain-unique style space S. In the process of image conversion, the content space of the input image remains unchanged, and the style space represents the main change information of the image.

By combining with different style space, the model can generate a variety of images in which the style space can be given randomly, or it can be extracted from a picture in the target domain.

To train the MUNIT model, Huang et al proposed two loss functions, bidirectional reconstruction loss, and adversarial loss. The adversarial loss is the same as cGAN. The bidirectional reconstruction loss was designed to confirm the encoders and decoders are inverses, which means both image-to-latent-to-image and latent-to-image-to-latent directions can be constructed. The formula is shown as follows:

$$\mathcal{L}_{recon}^{c_1} = \mathbb{E}_{c_1 \sim p(c_1), s_2 \sim q(s_2)}[\|\mathbb{E}_2^c(G_2(c_1, s_2)) - c_1\|_1] \tag{9}$$

$$\mathcal{L}_{recon}^{s_2} = \mathbb{E}_{c_1 \sim p(c_1), s_2 \sim q(s_2)}[\|\mathbb{E}_2^s(G_2(c_1, s_2)) - s_2\|_1] \tag{10}$$

where $q(s_2)$ is the prior $N(0, 1), p(c_1)$ is given by $c_1 = E_1^c(x_1)$ and $x_1 \sim p(x_1)$.

The MUNIT model achieves an unsupervised image-to-image translation among multiple domains and produces qualified and diverse images. It addresses the limitations of image-to-image transfer like CycleGAN.

In recent years, the style transfer of GANs has been improved extremely in terms of image qualities and the efficiency of generating images. Simultaneously, the advancement of the industrial assembly line based on GANs is superior to trivial artificial productions. Therefore, style transfer has emerged in plenty of commercial applications, such as photo and video editors, text style transfer, video gaming and virtual reality. The following is a detailed introduction to the application of style transfer in these several aspects.

4. Application

In recent years, the style transfer of GANs has been improved extremely in terms of image qualities and the efficiency of generating images. Simultaneously, the advancement of the industrial assembly line based on GANs is superior to trivial artificial productions. Therefore, style transfer has emerged in plenty of commercial applications, such as photo and video editors, text style transfer, video gaming and virtual reality. The following is a detailed introduction to the application of style transfer in these several aspects.

4.1. Photo and video editors

On the basics of deep learning neural networks, style transfer algorithms are not restricted to adding a filter simply, instead employing neural networks to learn artistic style, texture, and other features in order to make photo software become outstanding artists. For example, Prisma, Depart, and other applications are able to transfer normal photos taken by users into a master's specific style. In other words, one of the clearest applications of style transfer is in a photo and video editing software. From sharing stylized selfies to augmenting user-generated music videos, and beyond, the ability to add famous art styles to images and video clips promises to add unprecedented power to these kinds of creativity tools.

And given the flexibility and performance of current deep-learning approaches, style transfer models can easily be embedded on edge devices - for instance, mobile phones-allowing for applications that can process and transform images and video in real-time. This means that professional-quality photo and video editing tools can become more widely accessible and easier to use than ever before. There are a number of incredible GAN models used to edit photos or videos. Furthermore, the researchers apply their algorithms to amounts of aspects such as animation, oil paintings, and face synthesis.

4.1.1 Photo-to-Animation Translation

Anime is a film and television style that contemporary young people like very much. However, a large number of manual hand-painted images and video frames are required, which is not only time-consuming but also high-cost. Accordingly, if we can utilize style transfer to substitute hand painting, it will substantially improve the anime production efficiency. For example, In 2018, Wenbin, etc. proposed CariGAN [18] which generates caricature through weakly paired adversarial learning. Also, CariGAN can systematically analyze face photos and exaggerate certain features. In 2020, As a novel approach, AnimeGAN [1] is used to transfer photos of real-world scenes into anime-style images. It realizes scene anime effectively.

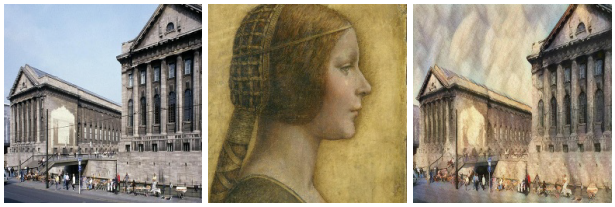
(a) Real Photo (b) Hayao Style

Figure 7. AnimeGAN [1]

4.1.2 Photo-to-Oil Paintings

Oil-painting style transfer is the shared structural similarity of synthetic content images, reflecting the style of artistic style. The artistic style here refers to the painter's

painting genre, while the artistic image refers to a group of images created by the same artist, each image has its own unique personality. To generate a new ornamental image, add an image's oil painting style information to any image while preserving the image's semantic content. Scientists proposed cycleGAN which supports to transfer of oil painting style to unpaired images. For example, cycleGAN could transfer Monet's painting style to any other photos. In 2021, Wenju et al. proposed a dynamic res-block generative adversarial Network (DRB-GAN) [28] for oil-painting style transfer to obtain high-resolution images.



(a) Input (b) DaVinci (c) Output

Figure 8. DRB-GAN [28]

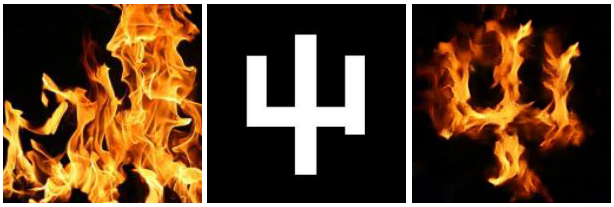
4.1.3 AI Generate Faces

With the development of style transfer, the images generated by GAN-based methods often show unprecedented prospects. Since ProGAN [13] has resolved the problem of generating high-resolution facial images in 2017, styleGAN [14] as a new network modified every layer of represented visual features in order to get better performance in 2018. In the same year, starGAN successfully realizes image transfer learning for multiple domains, which makes face synthesis improve greatly. Additionally, in 2020, Fania .etc proposed Cross-Domain Face Synthesis generating realistic synthetic face images that reflect capture conditions in the target domain [22].

Figure 9. AI Generate Faces [5, 9, 17, 26]

4.2. Text Style Transfer

With the advancement of style transfer, GAN-based style transfer methods are also used for natural language processing(NLP). Text style transfer is an important task, which aims to control certain attributes in the generated text, such as politeness, emotion, humor, and many others. For artificial intelligence systems to accurately understand and generate language, it is necessary to model language with style/attribute, which goes beyond merely verbalizing the semantics in a non-stylized way. The application of text style transfer can be beyond your imagination. Text stylization is a process of designing special effects for text and rendering it into a unique artistic word. For example, Shape-MatchingGAN [29] as a method of artistic text style transfer is the task of migrating the style from a source image to the target text to create artistic typography. Furthermore, some of the GAN-based models are able to transfer negative words into positive words.



(a) Input Style (b) Target Text (c) Result

Figure 10. Shape-MatchingGAN [29]

4.3. Video Gaming

At a press conference during the 2019 Game Developers Conference, Google introduced Stadia, it's cloud-powered video game streaming service. And one of the primary features included in that demo was an in-game style transfer feature that automatically recforposes the virtual world with textures and color palettes from a potentially limitless range of art styles. Google noted that efforts like these are intended to "empower the artist inside of every developer." This implies that gaming, much like other applications of style transfer, will broaden the accessibility of artistic creation to those who we might not traditionally see as artists.

4.4. Virtual Reality

Much like gaming, where immersive virtual worlds represent the anchor of the user experience, virtual reality has also had its share of interest in exploring what's possible with style transfer. While the applications with style transfer in VR are still largely in the research phase, the possibilities are exciting and promising. Facebook touts the potential of style transfer to radically alter the ways VR developers tell visual stories through their applications, games, films, and more. And there are some early demos that show how style transfer could help augment the immersiveness of worlds created for virtual reality.

Model	Unsupervised	Datasets	Application Scenario	Strength or Weakness
CariGAN	✓	WebCaricature	Photo animation	CariGAN generate plausible caricatures with large diversity but easily neglect tiny facial features
AnimeGAN	✓	ImageNet1000		AnimeGAN generated images are affected by the brightness
CycleGAN	✓	ImageNet1000	Oil painting	Image style transfer classic algorithm, strong versatility
DRB-GAN	✓	Place365		exhibit its superior performance in terms of visual quality
ProGAN	✓	CELEBA	Face synthesis	ProGAN progressively learn high-resolution image;
StarGAN	✓	CelebA/RaFD		StarGAN trains the data across multiple domains;
StyleGAN	✓	FFHQ		StyleGAN'generator is a really high-quality generator;
C-GAN	-	COX-S2V		it provides a higher level of accuracy for data augmentation;

Table 1. Comparison of the methods of GAN-based style transfer

5. Future and Discussion

In this paper, we first introduce the basic model of GAN. Then, we explore some popular GAN-based style migration models and related real-life applications. Finally, we summarize the development trends of GAN-based methods for image style migration and list the real-world applications.

Although GAN models have been developed a lot and have a wide range of applications in image-to-image transformation. However, GAN also has many problems, such as slow training speed, sensitivity to noise, and difficult evaluation of transfer results. It is worth noting that there are no scientific and uniform evaluation criteria for GAN models. Various researchers are proposing judgment methods applicable to current research results, and some of them are even judged by human eyes. We need a standardized evaluation system to evaluate future GANs.

In the future, GANs can be further applied to medical, educational, and transportation fields. Similar frameworks can be further extended and applied to more domains, but not limited to style transfer. Last but not least, we made a hypothesis about how we could migrate a similar approach to style transfer to audio.

By applying audio content to someone's speaking style, the output is audio that is similar to that person's speaking style. Based on content-based audio and music signals, a neural network framework similar to GANs is trained, which can automatically generate the person's voice as the output according to the provided voice content. The framework of StarGAN or MUNIT can be extended to audio as well so that multiple audios of different characters are produced simultaneously. By addressing the limitations mentioned and achieving audio extensions, we believe style transfer based on GANs can have an ideal development.

References

[1] Jie Chen, Gang Liu, and Xin Chen. Animegan: a novel lightweight gan for photo animation. In *International symposium on intelligence computation and applications*, pages 242–256. Springer, 2020. 6

[2] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8789–8797, 2018. 2, 4

[3] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014. 5

[4] Leon Gatys, Alexander S Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks. *Advances in neural information processing systems*, 28, 2015. 2

[5] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016. 5, 7

[6] Bruce Gooch and Amy Gooch. *Non-photorealistic rendering*. AK Peters/CRC Press, 2001. 1

[7] Aaron Hertzmann, Charles E Jacobs, Nuria Oliver, Brian Curless, and David H Salesin. Image analogies. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 327–340, 2001. 1

[8] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 172–189, 2018. 2, 5

[9] Goodfellow Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, and David Warde-Farley. Generative adversarial nets." in advances in neural information processing systems. 2014. 2, 3, 4, 5, 7

[10] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 2, 3, 4

[11] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. Neural style transfer: A review. *IEEE transactions on visualization and computer graphics*, 26(11):3365–3385, 2019. 1, 2

[12] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016. 2

[13] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017. 7

[14] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 7

[15] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. In *International conference on machine learning*, pages 1857–1865. PMLR, 2017. 5

[16] Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Transactions on graphics (TOG)*, 33(4):1–11, 2014. 5

[17] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018. 7

[18] Wenbin Li, Wei Xiong, Haofu Liao, Jing Huo, Yang Gao, and Jiebo Luo. Carigan: Caricature generation through weakly paired adversarial learning. *Neural Networks*, 132:66–74, 2020. 6

[19] Ming-Yu Liu and Oncel Tuzel. Coupled generative adversarial networks. *Advances in neural information processing systems*, 29, 2016. 2

[20] Jean-Francois Mertens and Shmuel Zamir. The value of two-person zero-sum repeated games with lack of information on both sides. *International Journal of Game Theory*, 1(1):39–64, 1971. 2

[21] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 3, 4

[22] Fania Mokhayeri, Kaveh Kamali, and Eric Granger. Cross-domain face synthesis using a controllable gan. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 252–260, 2020. 7

[23] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016. 4, 5

[24] Javier Portilla and Eero P Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*, 40(1):49–70, 2000. 1

[25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 4

[26] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016. 7

[27] Bingzhe Wu, Haodong Duan, Zhichao Liu, and Guangyu Sun. Srpgan: perceptual generative adversarial network for single image super resolution. *arXiv preprint arXiv:1712.05927*, 2017. 2

[28] Wenju Xu, Chengjiang Long, Ruisheng Wang, and Guanghui Wang. Drb-gan: A dynamic resblock generative adversarial network for artistic style transfer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6383–6392, 2021. 7

[29] Shuai Yang, Zhangyang Wang, Zhaowen Wang, Ning Xu, Jiaying Liu, and Zongming Guo. Controllable artistic text style transfer via shape-matching gan. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4442–4451, 2019. 7

[30] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE international conference on computer vision*, pages 2849–2857, 2017. 5

[31] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer, 2016. 5

[32] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 2, 5