Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230391

Research advanced in image style transfer based on deep learning

Yidan Gao[1,†], Luyao Wang[2,†] and Lechun You[3,4,†]

[1]Dalian University of Technology, Dalian, 116024, China
[2]Fuyang Normal University, Fuyang, 236041, China
[3]Beijing Institute of Technology, Beijing, 100081, China

[4]1120190855@bit.edu.cn
[†]These authors contributed equally

Abstract. Image style transfer (IST) aims to combine the style of one image with the content of another to make the resulting image has both the attributes of the style image and the characteristics of the content image. Benefited from rapid developments of deep learning, the methods based on Generative Adversarial Networks (GAN) have greatly improved the accuracy of IST. In this paper, to promote the research of IST based on Convolutional Neural Networks (CNN), some main methods of it are presented. The developments of CNN and GAN are reviewed, and basic principles and techniques of IST based on CNN and GAN are introduced in detail. Moreover, some representative datasets for this task are summarized, then the evaluation metrics and results are presented. Moreover, some problems of current methods are discussed. Finally, application prospects in related fields and future research directions of IST are presented.

Keywords: image style transfer, deep learning, generative adversarial network.

1. Introduction

IST aims to fuse the style (texture, color, etc.) of one picture with the information (outline, sturcture, etc.) of another picture, so that the generated new picture has both the attributes of the picture and the characteristics of the content image. At present, with the quick advancement of multimedia technology, this task has been already applied in the image generation, artistic style reproduction, animation production and other fields.

Early style transfer methods are generally based on image processing and machine learning frameworks, including non-photorealistic graphics, texture conversion, computer graphics, computer vision, digital image processing and other technologies. However, the processing effect and processing speed of these methods can not meet the needs of practical application. Thanks to the strong ability of feature expression of CNN, the effects of style migration methods have been remarkably improved. Industry and academia have paid great attention to the technique proposed by Gatys et al. [1]. Since then, numerous studies that focus on both image-based iteration and model-based iteration have been proposed. In this article, we compare the development and changes of this work, discuss numerous applied IST approaches, and provide an overview of some classic datasets and their potential applications.

DOI: 10.54254/2755-2721/6/20230391

The most representative framework for transferring visual styles is GAN, which has 2 modules: a discriminator and a generator. To be specific, the generator is always designed as a Deep Neural Network (DNN), with a low-dimensional input vector as well as a high-dimensional output vector (picture, text or speech). The discriminator is a DNN as well, with a high-dimensional vector as input and a scalar as output. The more authentic the used image (or text, voice) is, the greater the scalar is. The generator and the discriminator conduct adversarial learning. The generator continuously evolves and strives to generate false images, so that the discriminator can be scammed. The discriminator also iteratively evolves, trying to identify the false images. Through the adversarial learning between the two, the fake generated images are more and more like the real images, and the discriminantor is more and more able to distinguish the false images that are very close to the real images. Both capabilities can be greatly enhanced during iterations.

Focusing on the GAN-based IST methods, in this paper, the representative IST methods are systematically introduced. We analyze the latest research results and its application prospects. Additionally, the representative datasets of this task and common evaluation methods are introduced. Besides, we report the output of different transfer methods and compare their advantages and disadvantages. Then we give an in-depth discussion on the problems that remain to be solved and put forward some suggestions with practical reference value, which lays a foundation for further studies. Finally, the challenges and trends of development in the future are summarized.

2. Methods

2.1. CNN-based methods

Convolutional Neural Networks (CNN), namely the network that can perform the operation of convolution and has a certain hierarchical structure, belongs to the Feedforward Neural Network (FNN). The ability to minimize data dimensions while still extracting and retaining picture information is one of the strongest points of CNN. The application of CNN in IST is mainly based on slow neural style transfer of image iteration. Firstly, extract image features use the DNN, besides the CNN is used to iteratively update the noise image pixels on a random noise image, so that it has the characteristics which match the expected image.

CNN is a multi-layer network structure, and each layer of it is made up of multiple images representing different features. CNN has the characteristics of fewer parameters and simpler structure. When processing images, it can directly use the image as the input signal, which overcomes the difficulties of traditional algorithms in image feature extraction. The traditional CNN algorithm generally includes three key steps, including feature extraction, classification recognition and prediction.

For IST, the generation of neural networks overcomes the limitations of manual modeling. Gatys et al. [1] first modeled the information of a picture as a characteristic response from a pre-trained CNN. So far, results of researches have proved that CNN has the capacity to extract picture style information from the network. Because of this finding, Gatys et al. suggested reorganizing the information extracted from given photos and the styles of best known artworks using CNN activation, thus opening up a new field. The content features of pictures are represented by the result of VGG network, including its general structures and outlines. Then they are presented with Gram matrix.

2.2. GAN-based methods

GAN is mainly composed of a generator G as well as a discriminator D. G generates an image according to the real images. At the same time, D learns to differentiate between the generated image and the one that is used as training samples. Its goal is to determine the optimal values for D and G to make the resulting image as realistic as feasible [2].

2.2.1. CycleGAN. To solve the problem of threshold migration of unpaired images, GAN learns a mapping $G: x -> y$. If the generated $G(x)$ can not be distinguished by the discriminator as true or false, it is effective [2]. As the Figure 1 shown, CycleGAN [3] is composed of two GANs to realize the

DOI: 10.54254/2755-2721/6/20230391

conversion of unpaired images. In order to make GAN more stable, another mapping is introduced which is $F: y -> x$. Cycle-consistency loss is also added to the network in order to prevent mode collapse.

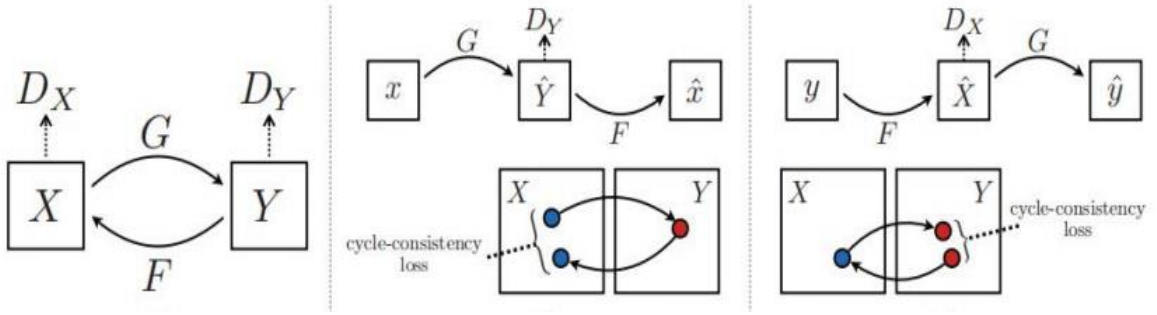

Figure 1. Framework of Cycle-GAN.

The loss functions of CycleGAN include adversarial losses and cycle-consistency loss. To be specific, making the resulting picture as similar to the final expected picture as possible is the goal of the adversarial loss, while cycle-consistency loss can prevent mode collapse and avoid different inputs corresponding to the same output. The discriminator aims to decrease the likelihood that the fake image will be considered to be true while increasing the likelihood that the true image will be. In this process, the generator cannot change the judgement of discriminator of the real picture. For the generator, the larger $D_y(G(x))$ is, the better, and the closer it is to 1, the smaller the loss is [3]. If only confrontation is used to generate network, mode collapse will occur. Therefore, Zhu et al. designed the cycle-consistency loss which enables the transferred picture to maintain features of the original input image, meanwhile, pixels can correspond to each other one by one to prevent mode collapse. For two identical images, it is calculated by making the difference element by element, taking the absolute value and then sum.

The whole model can be regarded as training two automatic encoders. When inputting x, the latent space compresses high-dimensional data into low-dimensional space at the bottom. Then it is converted into high-dimensional space by decoder. The goal of encoder F is to transfer the picture into an oil painting while the goal of G is to convert the oil painting into a photo, and then the converted image is compared to the original image.

2.2.2. StarGAN. There is a problem that CycleGAN can only convert one style. For the style transfer of face, to transfer the style in different parts, such as hair and gender, the model needs to be retrained. StarGAN [4] is more flexible in comparison, which enables multiple datasets to be trained in the same single network, and these datasets usually have different domains. As the Figure 2 shown, one generator and one discriminator are all that are needed for StarGAN to learn the inter-domain mapping. The inputs of the discriminator are two types of data, which are true pictures and fake pictures. The discriminator is used to distinguish whether they are true or false and identify their domain classification. In addition, the target image as well as the input image are inputs of the generator. The generator generates a false image and inputs it as well as the original source domain information to the generator and reconstruct the original image. Moreover, the reconstructed picture and the source image are subject to consistency constraints, and multiple datasets are trained at the same time.

DOI: 10.54254/2755-2721/6/20230391

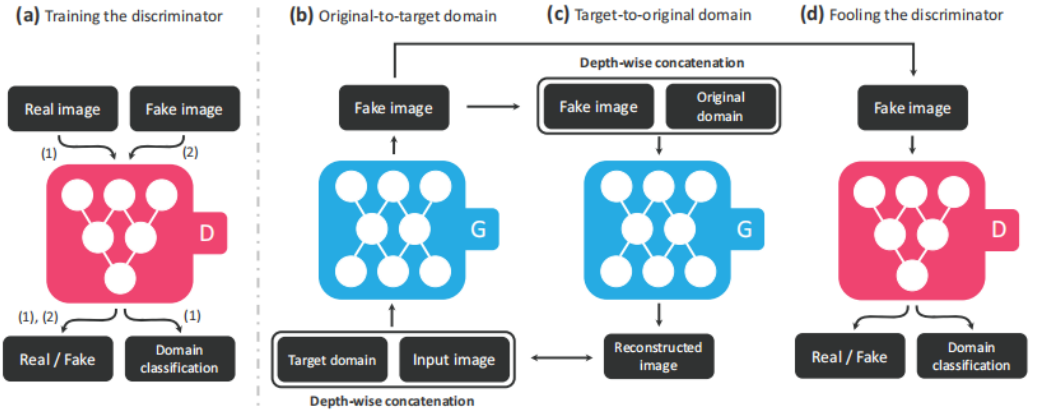

Figure 2. Framework of Star-GAN.

2.2.3. ArtsyGAN and Gated-GAN. ArtsyGAN [5] uses the perceptual loss function, which aims to continuously reduce the distance between the high-level features that have been produced from the pre-trained convolution network. ArtsyGAN replaces the previous reconstructor with perceptual loss, improves the prediction speed of the sampled image and introduces noise and loss functions in order to generate different details. Therefore, the image generated by ArtsyGAN has higher visible light quality, diversity and speed advantages than CycleGAN.

Since the traditional GAN generator only supports the transfer to one style, Chen et al. [6] proposed Gated-GAN so that multiple styles can be transferred using a single model. There are three modules in the generative network and the whole framework of it is shown in Figure 3. To be specific, the gated-transformer enables multiple styles by inputting different images through its different branches. For stable training, the auto-encoder is used for reconstruction of input images, which brings together the encoder and the decoder. To identify the style of the resulting image, an auxiliary classifier is introduced, which benefits the generator for its generation in multiple styles. By learning from diverse art styles, the Gated-GAN may further develop a new style.

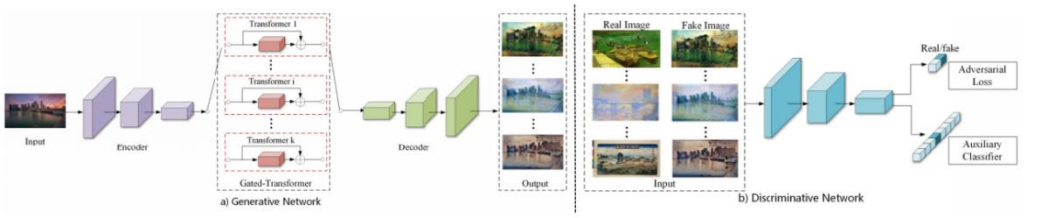

Figure 3. Framework of Gated-GAN.

2.2.4. SDP-GAN and DRB-GAN. To better preserve the structural and detailed content characteristics and achieve a complete transformation of style in all the regions of the picture, Li et al. [7] proposed Saliency Detail Preservation GAN (SDP-GAN), which trains a saliency network along with a generator. The purpose of saliency network is to keep details in salient regions and help the generator with salient features, which can be achieved by two newly introduced loss functions, including saliency-constrained content loss and saliency content loss.

Xu et al. [8] designed DRB-GAN that is composed of three modules (See Figure 4). They modeled 'style code' to fit the style encoding network together with the style transfer network, which are defined as shared parameters. An attention strategy which is on the basis of an auxiliary classifier is introduced in the style encoding network. In addition, there are several Dynamic ResBlocks, which are used to fuse

DOI: 10.54254/2755-2721/6/20230391

'style code' and extract CNN semantic features, and input them to the SW-LIN decoder to generate style transferred image. In the discriminative network, the inputs of the discriminator include the generated image and samples of style image, which then will pass the feature extractor so that feature maps are generated and will be accessed by a small Convolution network.

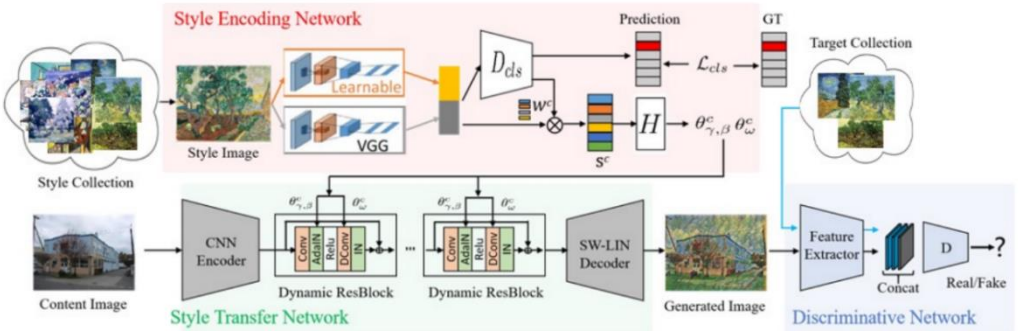

Figure 4. Framework of DRB-GAN.

2.2.5. CartoonGAN and AnimeGAN. Unlike other art styles, cartoon pictures tend to be highly abstract, with clear edges and even coloring. Existing GAN-based methods are not very satisfactory when dealing with this specific task. Therefore, Chen et al. [9] proposed a new technique - CartoonGAN, which specializes in the generations of high-quality cartoon images. As the Figure 5 shown, the whole framework is composed of two CNNs. Significantly, two novel loss functions are introduced. To be specific, the semantic content loss is calculated using the l_1 sparse regularization of the high-level features of VGG. Besides, the adversarial loss is improved to better preserve the edge information. The training images will go through three steps: edge detection using Canny edge detector [10], edge dilation and Gaussian smoothing in edge dilated regions. Therefore, the discriminator must be able to tell apart not only actual pictures and cartoon images, but also cartoon images with smooth edges. The authors also proposed a novel initialization phase which is to pre-train the network using only the semantic content loss.

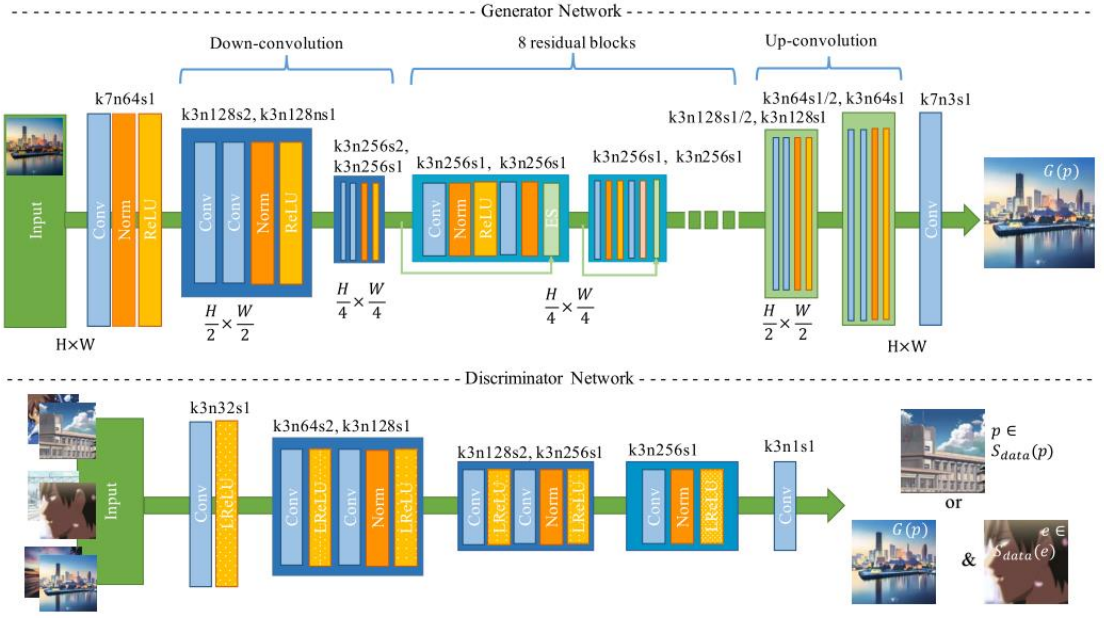

Figure 5. Framework of CartoonGAN.

DOI: 10.54254/2755-2721/6/20230391

Afterwards, Chen et al. [11] proposed AnimeGAN which is on the basis of CartoonGAN while achieved better results. The storage capacity and running time are reduced significantly. The discriminator network is the same as CartoonGAN while the generator network has been improved by adding a series of units including depthwise separable convolutions as well as inverted residual blocks (IRBs). Moreover, three novel loss functions are introduced.

3. Experiments

3.1. Classical datasets

Every neural network training project requires a substantial amount of data assistance. With the development of technology, the corresponding data sets will naturally be enriched. The common datasets for IST mainly include: (1) The computer vision recognition project known as ImageNet is one of the largest image databases in the world. It was founded by Li Feifei, a computer scientist at Stanford University. It stimulates human recognition system, which can identify objects from images. At present, it contains 14,197,122 images, becoming the largest existing image database. The ImageNet competition held every year is haunting the hearts of domestic and foreign famous schools and large IT companies. It can be used in the research direction of this paper or many other fields such as the vehicle identification and the building identification. (2) PASCAL VOC is a benchmark for visual object classification, recognition and detection, which provides standard evaluation system and image annotation datasets for detection algorithm and learning performance. (3) FDDB is mostly used for the research of constrained face detection. 5,171 photos of faces are chosen from 2,845 images acquired in different scenes by the datasets. (4) The Chinese University of Hong Kong created the benchmark dataset Greater Facial, which has a wider variety of face data. It contains 32,203 images and 393,703 face images, showing huge differences in scale, occlusion, pose, expression, dress and care. WIDER FACE has 61 event categories. It selected 40% as the training set, 10% for cross-validation, and finally 50% as the test set for each event type. The metrics used by the PASCAL VOC dataset and the WIDER FACE dataset are the same. As with MALF and Caltech datasets, there is no corresponding background bounding box for test images. This dataset can be used for face occlusion and recognition.

3.2. Evaluation metrics

The following are the most typical IST evaluation metrics:

(1) Subjective evaluation (human eye estimation). User research is a main objective evaluation method, which is to randomly invites subjects to assess the visual quality of different resulting images of the same original image. In the experiment, images for evaluation are randomly extracted from the test set. All the participants are asked to observe these images in the random arrangement and then vote for any image that is of good quality, therefore the results are determined by the users.

(2) IS. The working principle of GAN is that the generator can generate false images that deceive the discriminator so that it cannot distinguish the true from the false, but the classification clarity and diversity of the generated images cannot be guaranteed. Since the mode collapse may occur, the generated images tend to be consistent. The IS is used to reflect the clarity and diversity of the image. The IS is defined as:

$$\mathrm{IS(G)} = \exp(\mathrm{E_{x\sim P_g}} \mathrm{D_{KL}}(\mathrm{P(y|x)||P(y)})), \tag{1}$$

where x is the data element and y represents the label. Every time an image is input to the neural network, the result of the output layer is the category to which it belongs. $P(y|x)$ is the probability of determining which label it is, and $P(y)$ is the distribution of labels. The larger IS value represents that the image classification is clear and the images are more diverse [12-13].

(3) FID. FID is utilized to assess how well the recreated image differs from the actual sample one [14]. Firstly, the goal of the inception network is feature extraction, while the purpose of Gaussian model is feature space modeling and calculation of distance between two features. For each image, the lower

DOI: 10.54254/2755-2721/6/20230391

FID value indicates its better the quality and diversity. It is usually a supplement to IS. The FID is defined as the following formula:

$$\mathrm{FID}(x, g) = ||\mu_x - \mu_g|| + \mathrm{T_r}(\sum_x + \sum_g - 2(\sum_x \sum_g)^{1/2}), \quad (2)$$

where μ_r represents the feature mean value of the real picture, μ_g is the feature mean value of the recreated image, $\sum_x$ is the covariance matrix of the actual picture, $\sum_g$ is the covariance matrix of the generated picture.

SSIM. In order to make the evaluation results closer to the visual effect of the human eye, it is necessary to represent the structural information of the image from its brightness, contrast and structural similarity. Therefore the structural similarity (SSIM) model [15] was established and applied to the quality evaluation, which could better match the subjective sense of the human eyes. The brightness and contrast related to the structure of the object are taken as the definition of the structural information of the image. Therefore, the formula of SSIM is:

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (3)$$

where μ_x and μ_y represent the mean value of x and y, σ_x and σ_y represent the standard deviation of x and y, σ_{xy} represents the covariance of x and y. Moreover, x and y refer to input images and output images respectively. A higher SSIM value signifies that the produced image and the original image are more comparable. (when SSIM=1, the two images are consistent).

3.3. Methods comparison

3.3.1. Quantitative comparison. To evaluate the benefits and drawbacks of various methods, we quantitatively compared the results of representative IST methods on different data set indicators, as shown in Table 1 and Table 2.

In Table 1, the IS and FID scores of different GAN-based methods are compared. The SDP-GAN almost achieves the best IS scores among these methods when it is used to transfer these four kinds of styles. Meanwhile, compared to the previous methods, FID scores of SDP-GAN are reduced, proving its effective improvements. The IS scores of CartoonGAN are relatively high, too.

Table 2 lists the running time and GPU memory of NST, CycleGAN and DRB-GAN. Since NST is the first IST algorithm, its running speed is relatively low, but its pioneering contribution to this field can not be denied. The speeds of transfer using CycleGAN and DRB-GAN are significantly improved compared to it and the needed memories are reduced at the same time.

Table 1. Comparisons of GAN-based techniques using IS and FID scores [7].

Style	IS and FID	*StarGAN*	CartoonGAN	*CycleGAN +Lidentity*	SDP-GAN
Miyazaki Hayao	IS	5.48±0.52	6.09±0.83	4.37±0.58	6.38±0.98
	FID	169.42	159.69	136.82	133.76
Van Gogh	IS	4.28±0.53	4.77±0.48	4.58±0.74	4.86±0.59
	FID	162.89	135.93	106.94	101.59
Ukiyo-e	IS	5.33±0.52	6.10±0.72	5.75±0.64	6.07±0.77
	FID	152.46	123.42	107.25	101.92
MEAN	IS	5.03±0.52	5.65±0.68	4.9±0.65	5.77±0.78
	FID	161.59	139.68	117	112.42

DOI: 10.54254/2755-2721/6/20230391

Table 2. Quantitative comparison of NST, CycleGAN and DRB-GAN (measured on a Titan XP GPU) [8].

	Time (sec)	Memory (MiB)	Model
NST [1]	200	3887	PSPM
CycleGAN [3]	0.07	1391	PDPM
DRB-GAN [8]	0.08	1324	MDPM

3.3.2. Visual analysis. We also visualize the transfer results of different methods, which can be seen in Figure 5-7.

As shown in Figure 5, CycleGAN+$L_{identity}$ and SDP-GAN achieve better effects than the others when they are used to transfer the Van Gogh style. As for the Ukiyo-e style, the results of CartoonGAN and SDP-GAN are better, which not only preserve the content features but also achieve relatively complete style transfer. However, the output of CycleGAN+$L_{identity}$ loses some details on the person's face. Meanwhile, for the two styles, the results of StarGAN seem to maintain more details of outlines and textures of the input content images.

Figure 6 shows the results of three methods which are used to transfer cartoon styles. There are big differences between the hue and luminance of these outputs but the effects are all good. Figure 7 also compares the transfer to cartoon style using different methods. Visibly, the CartoonGAN is especially good at this kind of task and its resulting images are of good quality. The first NST methods proposed by Gays et al. [1] also achieved impressive results even it is not designed to this specific cartoon style transfer task. On the other side, the CycleGAN does not fit the animated transfer, as shown in this picture.

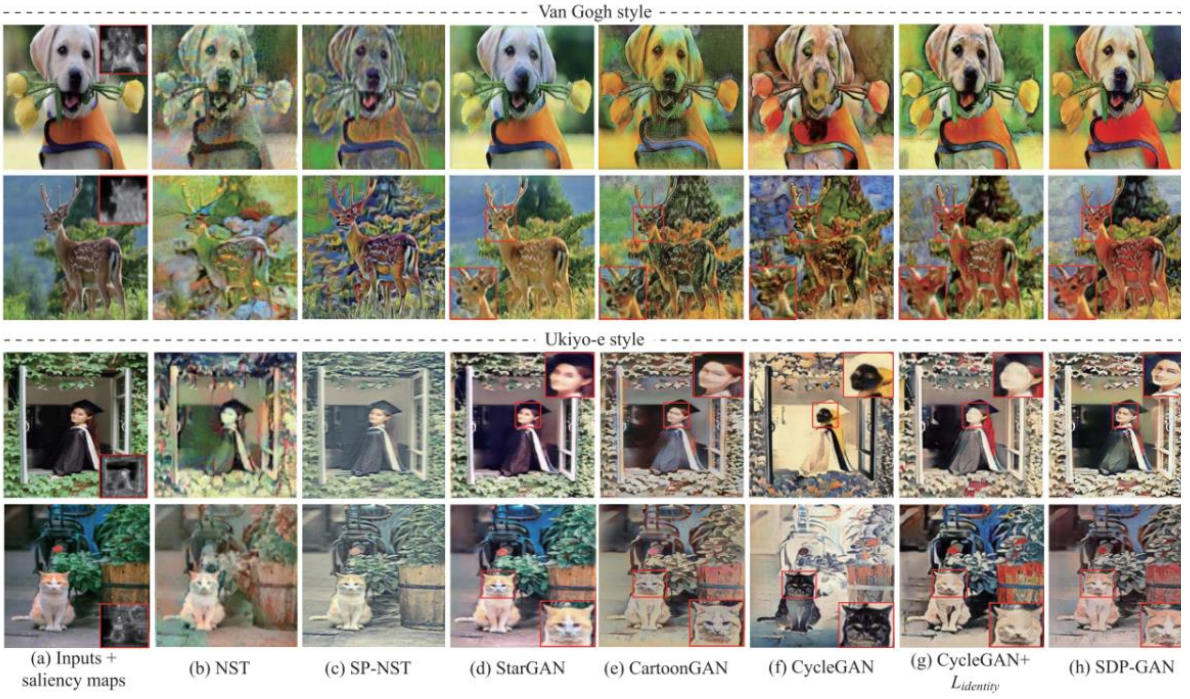

Figure 6. Qualitative comparisons of NST [1], SP-NST [16], StarGAN [4], CartoonGAN [9] and CycleGAN [3] and SDP-GAN [7] for two different styles [7].

DOI: 10.54254/2755-2721/6/20230391

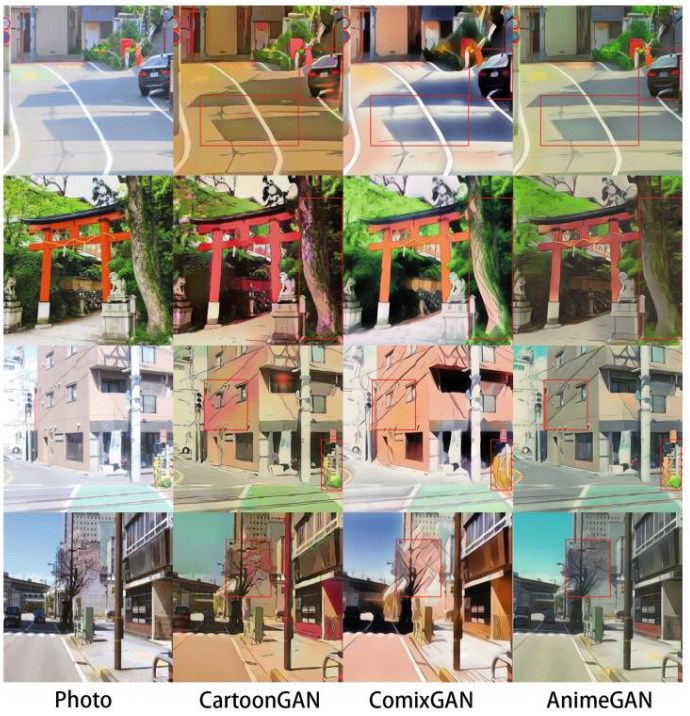

Figure 7. Qualitative comparisons of CartoonGAN [9], ComixGAN [17] and AnimeGAN. [11].

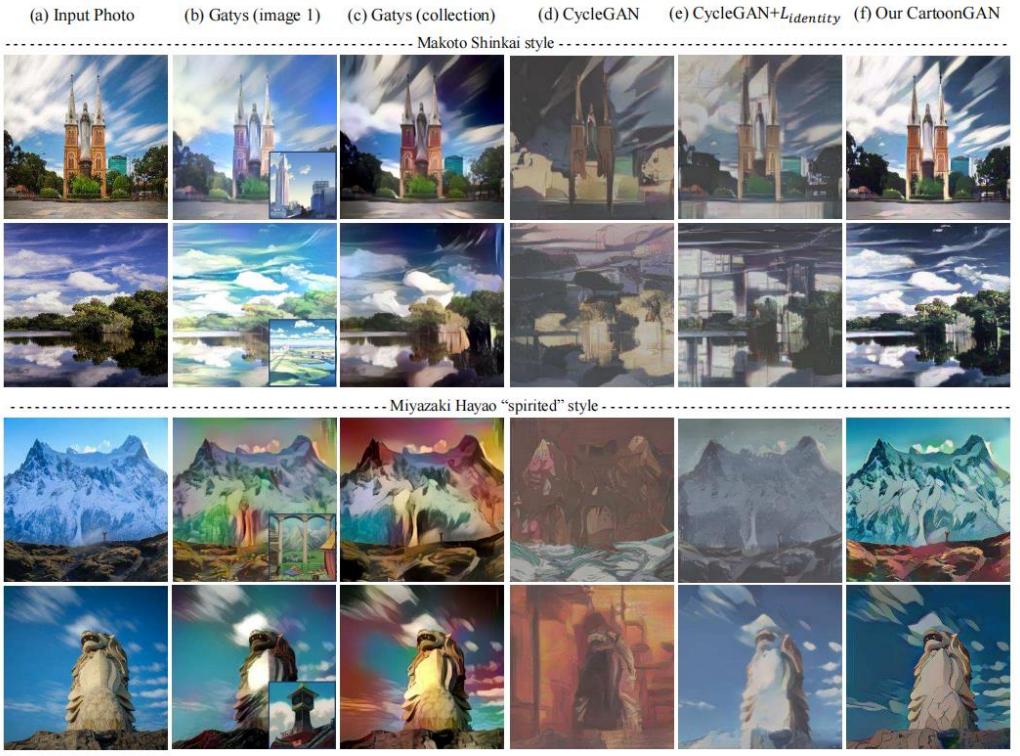

Figure 8. Comparison of NST [1], CycleGAN [3] and CartoonGAN [9] with for two different styles. [9].

DOI: 10.54254/2755-2721/6/20230391

4. Discussions

4.1. Current problems

At present, one of the problems of image style is the versatility of the model. Many models perform well for specific styles while are not good at some other styles. The CNN-based method is better in the case of irregular styles, but for regular patterns, such as polygons and pixelized styles in cartoon-style images, it usually produces distorted results, which is due to the characteristics of CNN in reconstructing images. [18]

The problems of IST task mainly lie in the style extraction stage. The combination of different style images and content images needs to retain different features, such as texture, color, shape, and edge information. Different requirements are put forward for the model. In the style extraction stage, specific algorithms can be added to retain the required features, and different weights can be assigned to different types of features. The combination is carried out according to the image characteristics, so as to achieve the desired effect.

Another problem is the speed of style transfer. For models that have been trained for a specific style, the migration speed is faster. However, if the two content pictures and style pictures are randomly chosen and the style and content characteristics should be learned, the speed will be slower.

Moreover, as the Figure 8 shown, for the adhesion objects, the objects close to each other after the style transfer are fused together. There were only horses and zebra in the training set, people and horses could not be distinguished correctly, so they were all transferred into zebras. Another problem is that the color and texture migration of the images is more successful, while the shape features cannot be completely transferred.

Figure 9. Failure cases of CycleGAN.

4.2. Future works

The technologies of IST have developed rapidly in the field of artificial intelligence and has made remarkable achievements. It has been widely used in all aspects, but there are still some problems to be solved. The following are some problems currently faced in the field of IST, as well as its future research directions and challenges.

4.2.1. Video style transfer. In the area of video style transfer, CNN has a lot of potential. A stylized network and a loss network are employed as the key components of the real-time video style migration. The stylized network is responsible for converting a single video frame into a stylized video frame, and vgg-19 is used as the loss network to extract reliable and meaningful contents of the original frame, stylized frame. The space and time loss of stylized image are calculated to train the stylized network

DOI: 10.54254/2755-2721/6/20230391

[19]. Among them, temporal consistency is an important measure of style transfer method used in the video field. The stylized network is trained using the loss network, which also determines the temporal and spatial loss. The styled network can produce time-consistent video frames after receiving enough training. Video style migration makes the original technology more widely used.

4.2.2. Model compression. At present, most IST methods rely on deep CNN models that are pre-trained on a huge amount of image datasets to obtain rich feature representation. Therefore the application of IST methods in mobile devices is limited by the huge scales of parameters. It also indirectly hinders the development of IST. Pei et al. improved the network structure through more efficient network computing and improved the timeliness (as a mobile terminal application), based on the real-time style migration model proposed by Johnson [20]. The improved method overcomes one disadvantage of IST, which is that the model is difficult to be applied in the mobile terminal to a certain extent without reducing the image quality.

4.2.3. Algorithm evaluation. There are two kinds of image quality evaluation: qualitative evaluation and quantitative evaluation. The method of qualitative evaluation is carried out by a large number of participants expressing their subjective feelings, so as to judge the image quality. The results of this evaluation method may vary from person to person, so it is not an ideal evaluation method. Quantitative evaluation is to reflect the difference between image quality through data comparison. Currently, the main indicators used are algorithm training and reasoning time. Yeh et al. [21] proposed the effectiveness and coherence of style migration algorithm and constructed a comprehensive quantitative evaluation program, which promoted the development of algorithm evaluation in the field of IST.

4.2.4. Theoretical perfection. Li et al. [22] regard IST as a domain adaptive work. They have theoretically proved that the Gram matrix of matching maps of characteristics is tantamount to minimizing the maximum average difference (MMD). Domain adaptation belongs to the field of transfer learning, so IST task can also be seen as an application of transfer learning. Therefore, improvements of transfer learning theory can provide a more comprehensive mathematical explanation and theoretical support for the style transfer algorithm, which has a significant impact on further developments of the field of IST.

4.2.5. Preprocessing and postprocessing. In order to achieve a better migration effect, some preprocessing and postprocessing techniques can be adopted, including image semantic segmentation [23], image fusion [24] and image smoothing processing, etc. Some of these methods have already been adopted. In the future, different preprocessing and postprocessing methods can also be combined to adapt to the characteristics of different transfer tasks.

4.2.6. Cross-modal style transfer. Cross-modal transfer is an additional feasible option to transferring styles in the picture mode in the future, such as describing the specific styles in text form and realizing automatic transfer, without the need to specify a style image.

5. Conclusion

Regarding computer vision, IST is a hot topic for research. Its goal is to fit the style of one picture and the substance of another together, then make the recreated image maintain the attributes of them. At present, the methods based on GAN remarkably improve the quality of results of this task and GAN has become the mainstream framework of IST. Focusing on the framework of IST, we first introduce the basic principles of CNN-based and GAN-based methods. Secondly, we summarize the main methods of IST and present some representative datasets and evaluation metrics. Then we compare the different methods and analyze the shortcomings of existing techniques. Last but not least, we talk about the task's potential for use in adjacent domains as well as its future research objectives.

DOI: 10.54254/2755-2721/6/20230391

References

[1] L. A. Gatys, A. S. Ecker and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2414-2423, doi: 10.1109/CVPR.2016.265.

[2] L. Zhao, Y. Jiao, J. Chen and R. Zhao, "Image Style Transfer Based on Generative Adversarial Network," 2021 International Conference on Computer Network, Electronic and Automation (ICCNEA), 2021, pp. 191-195, doi: 10.1109/ICCNEA53019.2021.00050.

[3] J. -Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2242-2251, doi: 10.1109/ICCV.2017.244.

[4] 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," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8789-8797, doi: 10.1109/CVPR.2018.00916.

[5] H. Liu, P. N. Michelini and D. Zhu, "Artsy-GAN: A style transfer system with improved quality, diversity and performance," 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 79-84, doi: 10.1109/ICPR.2018.8546172.

[6] X. Chen, C. Xu, X. Yang, L. Song and D. Tao, "Gated-GAN: Adversarial Gated Networks for Multi-Collection Style Transfer," in IEEE Transactions on Image Processing, vol. 28, no. 2, pp. 546-560, Feb. 2019, doi: 10.1109/TIP.2018.2869695.

[7] R. Li et al., "SDP-GAN: Saliency Detail Preservation Generative Adversarial Networks for High Perceptual Quality Style Transfer," in IEEE Transactions on Image Processing, vol. 30, pp. 374-385, 2021, doi: 10.1109/TIP.2020.3036754.

[8] W. Xu, C. Long, R. Wang and G. Wang, "DRB-GAN: A Dynamic ResBlock Generative Adversarial Network for Artistic Style Transfer," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 6363-6372, doi: 10.1109/ICCV48922.2021.00632.

[9] Y. Chen, Y. -K. Lai and Y. -J. Liu, "CartoonGAN: Generative Adversarial Networks for Photo Cartoonization," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 9465-9474, doi: 10.1109/CVPR.2018.00986.

[10] J. Canny, "A Computational Approach to Edge Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 6, pp. 679-698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.

[11] L. Zhang, Y. Ji, X. Lin and C. Liu, "Style Transfer for Anime Sketches with Enhanced Residual U-net and Auxiliary Classifier GAN," 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), 2017, pp. 506-511, doi: 10.1109/ACPR.2017.61.

[12] Salimans T, Zhang H, Radford A, et al. Improving GANs using optimal transport[J]. arXiv preprint arXiv:1803.05573, 2018.

[13] Barratt S, Sharma R. A note on the inception score[J]. arXiv preprint arXiv:1801.01973, 2018.

[14] Heusel M, Ramsauer H, Unterthiner T, et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium[J]. Advances in neural information processing systems, 2017, 30.

[15] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment[C]//The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. Ieee, 2003, 2: 1398-1402.

[16] M.-M. Cheng, X.-C. Liu, J. Wang, S.-P. Lu, Y.-K. Lai, and P. L. Rosin, “Structure-preserving neural style transfer,” IEEE Trans. Image Process., vol. 29, pp. 909–920, 2020.

[17] Pesko, M., Svystun, A., Andruszkiewicz, P., Rokita, P., Trzcinski, T.: Comixify: transform video into a comics. CoRR abs/1812.03473 (2018). http://arxiv.org/abs/1812.03473

[18] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu and M. Song, "Neural Style Transfer: A Review," in IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 11, pp. 3365-3385, 1 Nov. 2020, doi: 10.1109/TVCG.2019.2921336.

DOI: 10.54254/2755-2721/6/20230391

[19] Haozhi Huang, Hao Wang, Wenhan Luo, Lin Ma, Wenhao Jiang, Xiaolong Zhu, Zhifeng Li, Wei Liu; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 783-791

[20] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual Losses for Real-Time Style Transfer and Super-Resolution. Lecture Notes in Computer Science, 694–711.

[21] Yeh M C, Tang S, Bhattad A, et al. Improving style transfer with calibrated metrics[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020: 3160-3168.

[22] Yanghao Li, Naiyan Wang, Jiaying Liu, and Xiaodi Hou. 2017. Demystifying neural style transfer. In Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17). AAAI Press, 2230–2236.

[23] Garcia-Garcia, Alberto & Orts, Sergio & Oprea, Sergiu & Villena Martinez, Víctor & Rodríguez, José. (2017). A Review on Deep Learning Techniques Applied to Semantic Segmentation.

[24] Levin A，Lischinski D，Weiss Y． A closed-form solution to natural image matting［J］．2008，30，IEEE Trans on Pattern Analysis and Machine Intelligence，(2)：228-242．