2023 IEEE 6th Information Technology,Networking,Electronic and Automation Control Conference (ITNEC) | 978-1-6654-6004-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ITNEC56291.2023.10082673

Overview of Cartoon Face Generation

Xianfa Shen[1] ,Sujie Lei[2] ,Jiansong Liu[1]
1. College of Software, Yunnan University, Kunming Yunnan, China
2. College of Economics, Yunnan University, Kunming Yunnan, China
2543382723@qq.com, 1546656627@qq.com,1571233733@qq.com
Corresponding Author: Xianfa Shen Email:2543382723@qq.com

***Abstract*—As a computer art form, animation stylization of human face images is widely used in every aspect of daily life. From children's animation education books to classic animation works, animation style attracts children with a very charming art form, but also promotes the interest of children in exploring. In addition, animation production is also widely used in online games. Scenes and characters in games are often in the style of animation, which can reduce the production cost of games and the memory requirements of computers. In social entertainment, there are more and more people turning self-portrait into animation style as their social network profile pictures, which can not only attract the attention of others, but also protect the privacy of the portrait. However, drawing cartoon portraits by hand is very laborious and requires a lot of artistic skills, even with photo editing software. Therefore, how to perform face cartoonization efficiently and with high quality is an important issue. This article describes the development overview of face cartoonization; gives the application of face cartoonization; lists the commonly used datasets of face cartoonization; and discusses the methods of face cartoonization from three aspects; finally, the research direction and development trend of face cartoonization are prospected from the aspects of dataset, generated image definition, generated image details and model training time.**

Keywords— face cartoonization; unsupervised learning; deep learning; generative adversarial network

I. INTRODUCTION

The so-called face cartoonization is the conversion of a face avatar into a cartoon (animated) image. In the traditional face cartoonization methods, part of them are mainly based on the information of the five facial features to generate simple line drawings [1~4], but do not have a specific artistic style; the other part is based on the machine learning method, which establishes the matching relationship between the face image block and the cartoon image block through simple learning of the sample, finding the cartoon image block that best matches the original image block, and then carries out the synthesis of the cartoon avatar [5~9]. Although this method can achieve style conversion, the similarity with real faces is low, and there are also problems with unnatural expressions and single effect, and the processing steps are complicated and less efficient. And deep learning, as a new branch of machine learning, because of its powerful learning ability, the algorithms based on deep learning have all improved in performance and played a key role in the research related to image processing [10] field, such as image classification [11], image segmentation [12], target detection and tracking [13~14], etc. In addition, convolutional neural networks [15] are capable of converting images to each other through automatic coding and decoding of images, thus enabling end-to-end conversion of image styles, which makes it possible to automatically generate cartoon drawings. However, due to the special nature of face images, there are large differences and uniqueness between individuals, especially in the five facial features, so further research is needed for the style conversion of face images. On the other hand, the current research on the automatic generation method of face caricature is insufficient, and it still remains at the stage of manual deformation processing and cannot realize automatic exaggeration processing. All in all, it is a challenging challenge to realize automatic style conversion of face images in a fast and high quality way.

In summary, research on the automatic generation method of face cartoons and exaggerated cartoons has important scientific significance and application value.

Fig. 1. Example diagram of face cartoonization [16]

II. APPLICATION OF FACE CARTOONIZATION

Comic faces were originally just a kind of graphic art, mostly using exaggerated, satirical and realistic techniques to sketch the characteristics of characters. Caricature faces are commonly used in various entertainment projects, such as comic books, street cartoons, celebrity caricature covers, etc. This method, which is usually obtained by hand, requires a lot of time for composition, tracing, coloring and other steps, and is quite expensive [17~18]. This method, although able to obtain the most detailed and natural cartoon avatar, is not suitable for today's era of rapid technological development, resulting in a large number of personal customization software. For example, "Face Moe" is a cartoon image generation software that allows users to select their own features and compose an exclusive cartoon avatar, which has become a popular topic in people's lives and is loved by users. The following is a detailed description of the application of face cartoons in companies and people's daily lives.

In some corporate companies, due to the influence of cartoon cartoons, companies have begun to use cartoon style characters or animals, etc. as image endorsement, because compared with the way to promote the brand through celebrity endorsement, cartoon images have the characteristics of low cost, flexibility and low risk. A cartoon image as a logo has the following advantages: First, cartoon image is more concise and can quickly convey the brand concept; Second, cartoon image has affinity and certain symbolic characteristics, which have strong design advantages and high design added value for attracting consumers' attention [19~20].

In people's daily life, people will design personalized cartoon avatars specifically for their social accounts, and cartoon avatars are usually cute, which can attract others' attention and protect the portrait's privacy. In video conferencing, cartooning the avatar of the presenter in the meeting increases the vividness of the meeting scene by realizing real-time conversion; in online games, by collecting the real image information of the player, the avatar of the game character is replaced with the cartoon image of the player, which makes it easier for the player to find a sense of belonging in the game, and at the same time increases the realism of the game scene, as if being in life.

Thus, cartoon cartoons are not only widely used in people's lives and entertainment, but also bring great fun to people. Therefore, the study of face cartoon cartoon generation technology has important research significance and application value.

III. COMMONLY USED FACE CARTOON DATASET

Table 1 lists some datasets commonly used in face cartoon generation experiments, which include real face datasets and cartoon face datasets.

TABLE I. FACE CARTOONIZATION DATASET

Dataset Name	*Dataset Introduction*
FFHQ [21]	FFHQ is a high-quality dataset of yellow faces, containing 70,000 high-definition face images in PNG format at 1024×1024 resolution, with rich variety and obvious differences in age, race and image background, and very many variations in face attributes, with different ages, genders, races, skin tones, expressions, face shapes, hairstyles, face poses, etc.
Celeb A [22]	A large face recognition dataset published by Prof. Xiaogu Tang's lab at the Chinese University of Hong Kong in 2015. It contains 200K face images with more than 40 kinds of face attributes.
CartoonSet10/100k [23]	Released in 2017, there are two subsets, CartoonSet10k and CartoonSet100k, containing 10,000 and 100,000 cartoon face maps, respectively. Each cartoon face map has 16 components, including 12 facial attributes and 4 color attributes.
Self2anime [24]	Released in 2019, this is a comic face dataset that first uses a comic face detection algorithm for images on Anime-Planet1, leaving a total of 3,500 female face images, 3,400 of which are used as training and 100 as testing.
Danbooru2019 [25]	Danbooru2019 is an anime character dataset with more than 300 million images. A researcher selected some images from it to form an anime avatar dataset with a total of 140,000 images and an image size of 512×512.
Cartoon Portraits [26]	This is a cartoon face dataset, including 850 drawings, 730 watercolor drawings and 3000 cartoon drawings.
CUFSF [27]	Published in 2009, this is a portrait sketch dataset with 1195 paired gray frontal portraits and corresponding sketches, originally from FERET.
Anime-Faces [28]	This is a dataset containing 1551 anime faces. There are 21551 in total, all in PNG format, and all images are resized to 64× 64.
IIIT-CFW [29]	Published in 2016, it includes 8,928 cartoon pictures of 100 celebrities and also comes with 1,000 real pictures.

IV. THE METHOD OF FACE CARTOONIZATION

A. Interactive-based approach

Interactive refers to the need for frequent interaction between humans and computer, and its main method is for the input face image to form a cartoon with some humorous effect and high recognizability by manually specifying the deformation area. Image deformation technology has been widely used in animation, which provides stable, realistic and natural distortion effects through deformation control and deformation functions with certain constraints. It is not only widely used in animation, film and television production, but also plays an important role in modeling, medical imaging and other fields, so it has also developed rapidly, and many image distortion techniques have been proposed.

The study of image morphing methods can be traced back to the 1960s. At that time, cross-dissolution [30] provided a smooth transition from one image to another. Cross dissolve was very effective and could produce smooth distortion. However, the visual effect of cross-dissolution was not satisfactory. Strictly speaking, cross-dissolution cannot be regarded as an image distortion method; it can only be regarded as an image transformation. It was not until the late 1980s that Douglas Smythe proposed the grid distortion distortion method [31] in the process of making willow films.

In recent decades, research on image distortion techniques has made rapid progress. Igarashi put the image into a triangular grid, rotated and scaled the transformation separately, and then solved the least square problem to make the transformation of the cartoon image as rigid as possible. Weng extended Igarashi's method to minimize the nonlinear energy equation using the Gauss-Newton iteration [32]. The method is easy to maintain the global and local characteristics of the image and it can be used for the deformation of rigid bodies. Guo [33] further proposed shape deformation based on nonlinear least-square optimization. During the deformation process, the mesh length is constrained to maintain the shape features in the boundary, local and global regions.

Related research work has been carried out mainly on several aspects, such as the selection of control handles [34], energy minimization methods [35~36], multilevel freeform deformation (MFFD) [37], and radial basis function deformation method [38]. Among them, the radial basis function method is one of the most widely used deformation methods. This deformation method is based on the features of the image. Its steps can be summarized as follows: firstly, select the feature points before deformation and establish their correspondence according to these feature points, and then establish the mapping function according to these relations. Finally, each pixel in the image is transformed according to the mapping function. In this transformation, all the deformation functions are radial basis functions, so it looks stiff. Later, Schaefer [39] proposed moving least squares based on the image deformation method. This method deforms the image using control points or control line segments and analyzes the affine transform, similar transform and rigid transform, respectively. More interesting visual effects are improved. However, the method does not consider the topological relationship of image deformation and has some limitations.

In the current interactive cartoonization methods, most of the generated cartoons are sketches without color rendering, which lack aesthetic appeal. And most of the methods simply generate the corresponding cartoon drawings for the input image, without considering some subsequent processing after cartoon synthesis.

B. Sample rule-based approach

The sample rule-based approach refers to summarizing the rules for caricature creation based on a certain amount of data set, allowing the algorithm to automatically learn the correspondence between photos and caricatures, and having the computer simulation automatically generate the caricatures. A method of dynamic exaggeration of faces to generate face caricatures was first proposed by Brennan [40], which uses the difference between face images and reference images to generate caricature images. Koshimizu [41] proposed the very famous PICASSO face caricature generation system, which can process 2D and 3D face images and generate face sketches. Liang [42] uses a standard rule-based approach to generate face caricatures in which the standard sample consisted of the image set of both the caricature drawn by the artist and the original face image, and learned the relationship between the face image and the caricature image graph by calculating the partial least squares (PLS) between the two to learn the exaggeration rule. Mo [43] has improved Brennan's face exaggeration method by adopting different strategies of exaggeration for face features of different scale sizes because it is sample-driven. In the case of Liu's work [44], an inter-class Gibbs model (IGM) is learned from a dataset of face images and corresponding caricature images, and the model is used to automatically generate face caricature images from the input face images.

The above methods are all sample rule-based face caricature generation methods, but the focus of these works is on how to develop exaggeration rules for face features. Although exaggeration rules can be developed that are very close to caricaturists' creative techniques, they ignore the rendering effect and artistic style of face caricature images, and do not make obvious optimization of the face caricature generation effect, resulting in a single caricature style. Moreover, some research work requires paired data of faces and caricatures as training samples, and the cost of constructing such datasets is high. These rule-based and sample-based methods focus on finding the geometric shape transformation relationship from real faces to caricature faces, but this geometric transformation relationship is heavily dependent on the data sample, and therefore is also limited by the exaggerated style and degree of caricature faces in the data sample, and the geometric shape transformation is relatively single and not arbitrary.

C. Style migration-based approach

1) Generating adversarial networks (GCN)

GAN has gained great success in image processing in recent years, due to the continuous development of neural networks. They are considered to be the most effective methods for image generation tasks and play an important role in various fields such as image translation and image stylization.

This model has been popular since 2014 when Goodfellow [45] proposed generative adversarial networks (GAN). GAN consists of a generator G and a discriminator D. The general structure of a generative adversarial network is shown in Fig. 2.

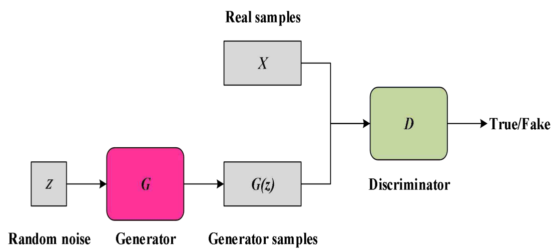

Fig. 2. Structure diagram of generating adversarial network

GAN has a generator Generator and a discriminator

Descriminator. generator G is used to generate realistic samples from random noise and tries to fool discriminator D. discriminator D is used to determine whether this sample is real or generated by generator G. The generator and the discriminator continue competing until the discriminator is unable to distinguish between the image generated by the generator and the real image. The whole process can be seen as a two-person minimax game. Where the main purpose of GAN training is to reach Nash equilibrium [46].

Definition 1. The GAN loss function can be described as

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[log\, D(x)] + E_{z\sim p_z(z)}\left[log\left(1 - D\big(G(z)\big)\right)\right] \quad (1)$$

Where $p_{data}(x)$ denotes the true data distribution and $p_z(z)$ denotes the noise distribution. The ultimate goal of this loss function is to make the sample distribution generated by the generator fit the real sample distribution as much as possible, and the discriminator discriminates as much as possible whether the input sample comes from the real sample or the generated sample.

2) Image to image translation

Many problems in image processing, computer graphics and computer vision can be thought of as "translating" the input image into the corresponding output image. The term "translation" is often used for translating between languages, for example, between Chinese and English. Image translation, however, means conversion from image to image in a different form. The idea of image-to-image translation can be traced back to the image analogy of Hertzmann [47]. This is a non-parametric model that uses a pair of images to implement an image transformation [48]. Many problems involving computer vision and computer graphics applications can be seen as examples of image-to-image translation problems. The task of image translation is to learn a map from a given image *X* to a specific target image *Y*, for example, mapping a grayscale image to an *RGB* image. Learning a mapping from one visual representation to another visual representation requires understanding the underlying properties shared between these representations, which are either domain independent or domain specific. However, learning a map between two or more domains is a challenging task for two reasons. First, collecting a bunch of images may be difficult, or the relevant images may sometimes not exist. The second difficulty is when performing multiple model transformations, i.e., mapping one input image to multiple outputs.

In recent years, image-to-image translation has made great progress, and image-to-image translation using generative adversarial networks has attracted significant attention in both supervised and unsupervised learning research. Noise-to-image GAN generates real images from random noise samples, while image-to-image GAN generates different images from images, and many variants of GAN have been proposed that have achieved good results in image-to-image translation tasks [49~51].

3) Generative adversarial network-based approach

Image style migration [52] is a technique that uses algorithms to learn the style of a famous painting and then apply that style to another image. It is an application of image-to-image translation. Early techniques for style migration include mainly non-realistic rendering [53~55] and texture migration [56~58]. Non-realistic rendering (NPR), also known as stylized rendering, aims to simulate an artistic style of drawing. It can digitally simulate various drawing styles such as watercolor painting, ink painting, and classical Chinese style painting; while texture migration techniques fill the input image with texture based on the reference image, making the generated image have a texture style similar to the reference image, which is suitable for images with simple texture. However, these two methods can only render specific styles and are slow. With the rise of neural networks, GAN-based style migration methods have become mainstream. These methods are able to understand the semantic information of an image by deforming the GAN model, reduce the time of model training, and generate high-quality cartoons. In the following, we will introduce some methods to implement face cartoonization based on GAN.

a) GAN-based approach

Chen [59] proposed a photo cartoonization solution called CartoonGAN, which converts photos of real scenes into cartoon style images. The network architecture of CartoonGAN is shown in Fig. 3.

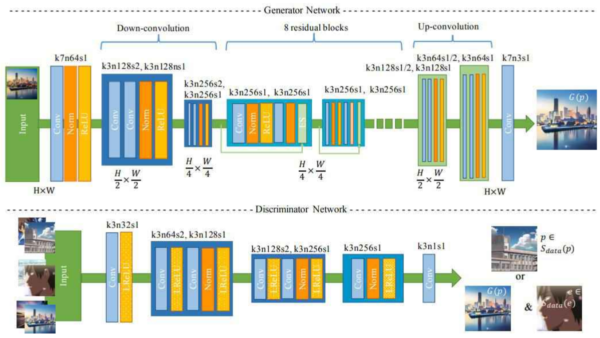

Fig. 3. CartoonGAN network architecture diagram

Definition 2. The loss function of CartoonGAN can be described as

$$L(G,D) = L_{adv}(G,D) + wL_{con}(G,D) \quad (2)$$

CartoonGAN is a cartoon generation method based on generative adversarial networks. The content loss function is proposed based on the original GAN $L_{con}(G,D) = E_{p_i\sim S_{data}(p)}\left[||VGG_l\big(G(p_i)\big) - VGG_l(p_i)||_1\right]$,where l refers to the feature maps of a specific VGG layer. The method uses a pre-trained high-level feature map in the VGG network [60] and defines the semantic loss by sparse regularization [61] of the feature map between the input and output photos, which can ensure that the generated cartoon image retains the semantic content of the input photo and proposes an edge-promoting adversarial loss to maintain clear edges of the generated image to generate high-quality images. CycleGAN requires two-way mapping to train two GAN models, which severely slows down the training process, while for this method, instead of mapping from cartoon back to photo, only one GAN model is needed for one-way mapping and the training time spent is much less using VGG feature maps to constrain the content.

Li [62] proposed a cartoon generation method called self-attentive generative adversarial network, which can generate better cartoon faces by introducing an Attention mechanism [63]. The network architecture of SCGAN is shown in Fig. 4.

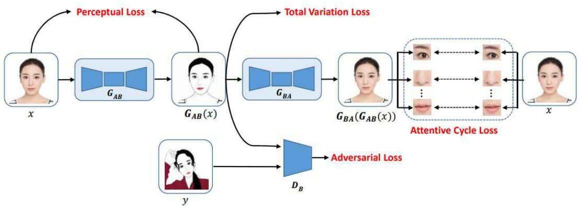

Fig. 4. SCGAN network architecture diagram

Definition 3. The loss function of SCGAN can be described as

$$L = L_{GAN}^{AB} + L_{GAN}^{BA} + a\left(L_{att-cyc}^{AB} + L_{cyc}^{BA}\right) + bL_{tv} + cL_{per} \quad (3)$$

Where $a, b, c > 0$,are the weights of different loss functions. SCGAN proposes a cartoon character generation task implemented using the attention mechanism. It can be divided into two important aspects. On the pixel aspect of the image, Total Variation Loss is introduced as a result of the cartoon image having a highly simplified and uniform color distribution, $L_{tv} = E_{x\sim P_{dt}(A)}[|\nabla G_{AB}(x)|]$,This forces the network to remove unwanted detail regions in the training and retains the important edge parts. At the region level, the authors propose Attentive Cycle Loss, based on Cycle Consistency Loss in CycleGAN, $L_{att-cyc}^{AB} = E_{x\sim P_{dt}(A)}\left[\sum_{j=1}^{k} a_j \,||G_{BA}\left(G_{AB}(x_i^j)\right) - x_i^j||_1\right]$,This loss function guides the generator and discriminator to focus more on detailed facial features such as eyelashes and pupils to produce more detailed cartoonish faces. This method is able to generate different cartoon-style images with high robustness.

b) CycleGAN-based approach

Zhu [64] proposed an unpaired image-to-image transformation method called CycleGAN. The network architecture of CycleGAN is shown in Fig. 5.

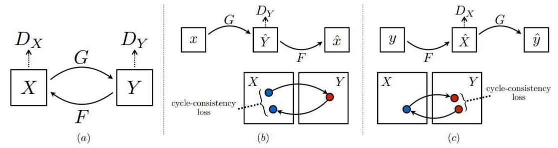

Fig. 5. CycleGAN network architecture diagram

Definition 4. The overall loss function of CycleGAN is defined as

$$G^*, F^* = arg\min_{F,G}\max_{D_X,D_Y} L\,(G, F, D_X, D_Y) \quad (4)$$

CycleGAN is an innovation in the field of unsupervised image translation research and is also used in style migration, which converts one class of images into another class of images. That is, now there are two sample spaces, X and Y, and we want to convert the samples in X space into samples in Y space. In fact, it is an $A \rightarrow B$ one-way GAN plus a $B \rightarrow A$ one-way GAN. two GAN share two generators, and then each carries a discriminator, in addition to introducing Cycle Consistency Loss, $L_{cyc}(G,F) = E_{x\sim p_{data}(x)}[||F(G(x)) - x||_1] + E_{y\sim p_{data}(y)}[||G(F(y)) - y||_1]$, It is used to ensure that the generated image retains as much information as possible from the source image as a way to break the dependence on paired data in the transfer of the two domains. Training can be achieved in the absence of sample pair information. The method does not require the input of one-to-one correspondence between the real image of the face and the cartoon image, which greatly reduces the cost of sample production. However, it lacks the understanding of face semantics, and the generated cartoon images are not effective.

Wu [65] proposed a method to generate cartoon faces using feature points of faces to assist CycleGAN, which uses the disparity between real faces and cartoon faces to train the data, and their network architecture is shown in Fig. 6.

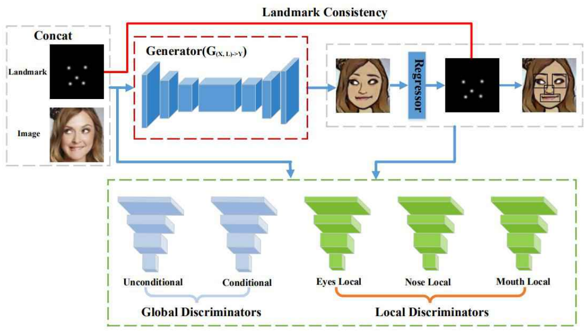

Fig. 6. Landmark-assisted CycleGAN network architecture diagram

Definition 5. The Landmark consistency loss proposed by this method can be defined as

$$L_c\big(G_{(X,L)\to Y}\big) = |R_Y\big(G_{(X,L)\to Y}(x,l)\big) - l|_2 \tag{5}$$

where $L(l \in L)$ denotes the input landmark heat map and is the pre-trained U-Net-like Landmark regressor with 5-channel output in each domain(R_Y is the domain Y). Under the constraint of this formula, images from different regions can be made to present similar facial features. The method solves the huge structural difference between two domains of real faces and cartoon faces by introducing Landmark-assisted CycleGAN, using face key points to constrain the face structure between two domains, and guiding the local discriminator training, which results in high quality of the generated cartoon face images.

c) U-GAT-IT based approach

Kim [66] proposed a way to generate realistic cartoon avatars by using adaptive layer instance normalization. The U-GAT-IT network architecture is shown in Fig. 7.

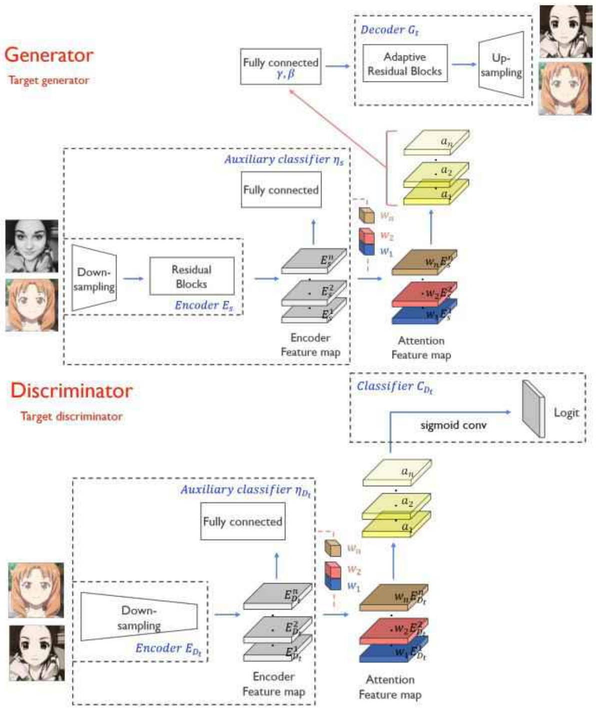

Fig. 7. U-GAT-IT network architecture diagram

Definition 6. The overall loss function of U-GAT-IT is defined as

$$\min_{G_{s\to t},G_{t\to s},I_s,I_t}\ \max_{D_s,D_t,I_{D_s},I_{D_t}}\ a_1 L_{l\,lsgan} + a_2 L_{cycle} + a_3 L_{identity} + a_4 L_{cam} \tag{6}$$

Where a_1, a_2, a_3, a_4 are the weight coefficients of different loss functions. U-GAT-IT is a new unsupervised method for image conversion across domains, and its main innovations are as follows: firstly, by adding an auxiliary classifier to the original GAN and thus introducing the Attention mechanism into the generator and discriminator, the geometric variation between domains can be processed, and images with overall or large shape variation can be converted; for various data sets with different amounts of shape and texture variation, the normalization The choice of the normalization function has a significant impact on the quality of the conversion results for various data sets with different amounts of shape and texture variations. Inspired by batch instance normalization (BIN) [67], the paper proposes the adaptive layer instance normalization (AdaLIN) algorithm, which can automatically adjust the weight of Instance Norm and Layer Norm by selectively maintaining or changing the content information, which helps to solve a wide range of image-to-image migration problems, enhances model robustness, and can achieve fine human This helps to solve a wide range of image-to-image migration problems, enhances model robustness, and enables beautiful portrait-to-manga style transformation.

V. THE TREND OF FACE CARTOONIZATION

With the rise of artificial intelligence, deep learning, augmented learning, machine learning and other artificial intelligence fields are increasingly in demand for face cartoonization, and the importance of face cartoonization is constantly highlighted. However, it also faces some challenges and problems in the process of its development, which are shown below:

(1) In terms of data collection, there are no publicly available paired datasets on human selfies and cartoon portraits, and real-face images are easily available on the web, while cartoon faces of specific styles are difficult to collect or design. In addition, the number of various cartoon faces is unbalanced, and cartoon faces of young women are more common on the web than other groups; on the other hand, most of the datasets are European face datasets, which are less applicable to Asian face cartoonization.

(2) In terms of generated image clarity, as most of the face cartoon style images developed by websites nowadays have lower resolution, which leads to lower resolution of the network output images, improvements can be made in the direction of collecting higher resolution face cartoon style images or super-resolution reconstruction of the generated images, so that the network can output higher quality target domain images; most of the current face cartoonization methods are only for face cartoonization, how to generate high quality and efficient full body cartoon images is also a problem. In addition, how to process video in real time to support interesting applications is another promising research direction.

(3) In terms of generating image details, most of the real face images in the dataset are frontal images of human faces without excessive expressions and without problems such as masking and complex backgrounds. In real life, the captured face images are variable in pose and expression, and when the expressions of the test images are more exaggerated, the test results obtained by many

algorithms have certain deviations. In order to make the conversion of cartoon cartoons more general, the training data should be diverse in content; in addition, since the hair color of the animation is almost fixed during the training process, the generated results may be unstable when the hair color of the real face image is different from that of other training samples. In this way, the current cartoon face generation algorithm does not guarantee that every generated image is of high quality, and a few generated images will be over-styled, resulting in unclear texture of the whole face and somewhat blurred contours of the five features, which is also a difficult point in the field of cartoon face generation at present and needs to be studied in depth.

(4) In terms of model training, the model network of the deep learning-based face cartoonization method is too deep and the training time is too long, and there is room for further optimization in the practicality of the algorithm.

VI. CONCLUSIONS

After years of development, image synthesis technology has made considerable progress, and face cartoonization has become one of the key research topics in the field of computer vision, which has been more and more widely used in recent years in fields such as online games and network avatars. Artificial intelligence technology brings great convenience to daily life, and the human demand for social entertainment has driven the research work on online games and intelligent beauty, while the application of converting self-portraits or character portraits into anime-style photos adds fun to life and improves spiritual life. We expect that future face cartoonization technology will have better development in the quality of anime-style image generation, so that the generated images have more delicate detail features; hardware and software will be continuously improved to make the face cartoon dataset with better resolution and shorter network training time.

REFERENCES

[1] W. Deng, Y.M. Wei, Y.G. Zhang, Generating personalized face cartoons using image morphing[J]. Computer Engineering and Applications,2011,47(24):132-135.

[2] F. Yan, G.Z. Fei, T.T. Liu, W.H. Ma, M.Y. Shi, Algorithm for face portrait generation in cartoon style[J]. Journal of Computer-Aided Design and Graphics,2007(04):442-447.

[3] R.Q. Zhou, J.Y. Zhou, Y.Q. Chen, J.F. Liu, Feature discovery-based cartoon face portrait generation[J]. Journal of Computer-Aided Design and Graphics,2006(09):1362-1366.

[4] H. Chen, N.N. Zheng, L. Liang, Y. Li, Y.Q. Xu, X.Y. Shen, Image-based personalized cartoon system[J]. Journal of Software,2002(09):1813-1822.

[5] C.T. Zhang, Research on face cartoonization method based on machine learning [D]. University of Electronic Science and Technology, 2011.

[6] M. Chen, Research on video face cartoonization method[D]. University of Electronic Science and Technology,2015.

[7] Q.Y. Li, D.S. Chen, Y.Y. Wu, Face cartoonization technology based on image distance matching[J]. Microcomputers and Applications,2014,33(10):44-46.

[8] S. Liu, Research on learning-based face expression animation generation method[D]. University of Electronic Science and Technology, 2013.

[9] L. Hu, Performance analysis and optimization design of face personalized cartoon model system based on sample learning[D]. Southeast University, 2005.

[10] K.C. Santosh, S.K. Antani, Recent trends in image processing and pattern recognition[J]. Multimedia Tools and Applications, 2020, 79(47-48):1-3.

[11] Y.Y. Sun, A review of deep learning and its research in image classification and recognition[J]. Information Technology and Informatization,2018(01):138-140.

[12] Y.Q. Li, X.H. Feng, Z. Wang, Advances in the application of computer vision technology[J]. Artificial Intelligence,2019(02):18-27.

[13] A. AKaff, D. Martin, F. Garcia, et al, Survey of computer vision algorithms and applications for unmanned aerial vehicles[J]. Expert Systems with Applications, 2018, 92: 447-463.

[14] M. Chen, L. Zhu, Z. Zeng, et al, Spatial target detection and tracking algorithm based on fully convolutional neural networks[J]. Transducer and Microsystem Technologies, 2019.

[15] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[16] L.L. Fan, Y. Li, X.Q. Zhang, Algorithm for cartoon stylized generation of key face contour regions[J]. Journal of Graphology,2021,42(01):44-51.

[17] J. Niu, Hand-drawn effects in traditional two-dimensional animation[J]. Art Education Research,2017(11):90.

[18] X.L. Chen, Research on Computer Aided Design Method of Hand-Painted Art Animation[J]. Journal of Residuals Science & Technology,2016,13(6).

[19] D.H. Song, A.N. Wang, Research on the application of cartoon image under brand concept[J]. Marketing Weekly,2021,34(01):36-38.

[20] P. Huang, The advantages and applications of cartoon images in commodity packaging design[J]. Educational Teaching Forum,2015(41):93-94.

[21] Https://github.com/NVlabs/ffhq-dataset.

[22] Z. Liu, P. Luo, X. Wang, et al, Deep learning face attributes in the wild[C]//Proceedings of the IEEE international conference on computer vision. 2015: 3730-3738.

[23] Https://google.github.io/cartoonset/download.html.

[24] Https://github.com/taki0112/UGATIT

[25] Https://www.gwern.net/Danbooru2019#danbooru2018

[26] X. Li, W. Zhang, T. Shen and T. Mei, "Everyone is a Cartoonist: Selfie Cartoonization with Attentive Adversarial Networks," 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 2019, pp. 652-657.

[27] X. Wang and X. Tang, Face Photo-Sketch Synthesis and Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 31, no. 11, pages 1955-1967, 2009.

[28] Https://pan.baidu.com/s/1NUzn_7C3cH21bRqHV3TZ0g

[29] A. Mishra, S.N. Rai, A. Mishra, et al, IIIT-CFW: A benchmark database of cartoon faces in the wild[C]//European Conference on Computer Vision. Springer, Cham, 2016: 35-47.

[30] T. Igarashi, T. Moscovich, J. Hughes, As-rigid-as-possible shape manipulation[J]. ACM Transactions on Graphics, 2005, 24(3):p. 1134-1141.

[31] D. Smythe, " A two-pass mesh warping algorithm for object transformation and image interpolation," ILM Technical Memo, 1990.

[32] Y. Weng, W. Xu, Y. Wu, K. Zhou, and B. Guo, " 2D shape deformation using nonlinear least squares optimization, " The Visual Computer, vol.22, no. 9, pp. 653-660, 2006.

[33] H. Guo, et al., ″As-rigid-as-possible shape deformation and interpolation,″ Journal of Visual Communication & Image Representation, vol.19, no. 4, pp. 245-255, 2008.

[34] S. Schaefer, T. Mcphail, J. Warren, Image deformation using moving least squares[J]. ACM Transactions on Graphics, 2006, 25(3): p. 533-540.

[35] S.Y. Lee, K.Y. Chwa, J. Hahn, et al, Image Morphing Using Deformation Techniques[J]. The Journal of Visualization and Computer Animation, 1996.

[36] D. Terzopoulos, J. Platt, A. Barr, and K. Fleischer, ″Elastically de-formable models,″ ACM SIGGRAPH Computer Graphics, vol. 21, no.4, pp. 205C214, 1987.

[37] S. Y. Lee, K. Y. Chwa, and S. Y. Shin, ″Image metamorphosis using snakes and free-form deformations,″ Cell, vol. 29, pp. 439-448, 1995.

[38] J. H. Kwon, B. G. Lee, J. Yoon and J. J. Lee, ″Image deformation using radial basis function interpolation,″ Vaclav Skala - UNION Agency,2015.

[39] T. Liu, Research on moving least squares image deformation method[D]. Dalian University of Technology, 2008.

[40] S.E. Brennan, Caricature generator: The dynamic exaggeration of faces by computer[J]. Leonardo, 1985, 18(3): 170-178.

[41] H. Koshimizu, M. Tominaga, T. Fujiwara, et al, On KANSEI facial image processing for computerized facial caricaturing system PICASSO[C]// IEEE International Conference on Systems. IEEE, 1999.

[42] L. Liang, H. Chen, Y.Q. Xu, et al, Example-based caricature generation with exaggeration[A]. 10th Pacific Conference on Computer Graphics and Applications, 2002. Proceedings[C]. IEEE, 2002: 386-393.

[43] Z. Mo, J.P. Lewis, U. Neumann, Improved automatic caricature by feature normalization and exaggeration[A]. Siggraph Sketches[C]. 2004: 57.

[44] Z. Liu, H. Chen, H.Y. Shum, An efficient approach to learning inhomogeneous Gibbs model[A]. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings[C]. IEEE, 2003, 1: I-I.

[45] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, et al, Generative Adversarial Networks[J]. Advances in Neural Information Processing Systems, 2014, 3:2672-2680.

[46] L.J. Ratliff, S.A. Burden, S.S. Sastry, Characterization and computation of local Nash equilibria in continuous games[C]// Communication, Control, & Computing. IEEE, 2006.

[47] A. Hertzman, C.E. Jacobs, N. Oliver, B. Curless, D.H. Salesin, Image analogies. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles,CA, USA, 12–17 August 2001; pp. 327–340.

[48] J.Y. Zhu, T. Park, P. Isola, et al, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks[J]. IEEE, 2017.

[49] X. Huang, M.Y. Liu, S. Belongie, et al, Multimodal Unsupervised Image-to-Image Translation[C]// European Conference on Computer Vision. Springer, Cham, 2018.

[50] 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.

[51] P.W. Wu, Y.J. Lin, C.H. Chang, et al, RelGAN: Multi-Domain Image-to-Image Translation via Relative Attributes[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019.

[52] S.H. Chen, Y.K. Wei, L. Xu, et al, A review of image style migration research based on deep learning[J]. Computer Application Research,2019(8).

[53] S. Shekhar, H. Xiong, Non-Photorealistic Rendering[J]. A K Peters.

[54] T. Strothotte, S. Schlechtweg, Non-Photorealistic Computer Graphics: Modeling, Rendering, and Animation. Morgan Kaufmann Publishers Inc. 2002.

[55] P. Hall, A.S. Lehmann, Image and Video-Based Artistic Stylisation (Computational Imaging and Vision[J]. Computational Imaging & Vision, 2012.

[56] A. Efros, W. Freeman, Image quilting for texture synthesis and transfer[C]//Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, 2001：341-346.

[57] I. Drori, D. Cohen-Or, H. Yeshurun, Example-based style synthesis[C]//2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. IEEE, 2003, 2: II-143.

[58] O. FRIGO, N. SABATER, J. DELON, et al, Split and match：Example-based adaptive patch sampling for unsupervised style transfer[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016：553-561.

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

[60] K. Simonyan, A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[61] J.W. Liu, L.P. Cui, Z.Y. Liu, X.L. Luo, Regularized sparse model[J]. Journal of Computer Science,2015,38(07):1307-1325.

[62] X. Li, W. Zhang, T. Shen, et al, Everyone is a Cartoonist: Selfie Cartoonization with Attentive Adversarial Networks[J]. 2019.

[63] H. Zhang, I. Goodfellow, D. Metaxas, et al, Self-Attention Generative Adversarial Networks[J]. 2018.

[64] J.Y. Zhu, T. Park, P. Isola, et al, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks[J]. IEEE, 2017.

[65] R. Wu, X. Gu, X. Tao, et al, Landmark Assisted CycleGAN for Cartoon Face Generation[J]. 2019.

[66] J. Kim, M. Kim, H. Kang, et al, U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation[J]. 2019.

[67] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International conference on machine learning. PMLR, 2015: 448-456.