Computers and Electrical Engineering 95 (2021) 107338

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Sketch to portrait generation with generative adversarial networks and edge constraint

Qingyun Liu [a,b], Huihuang Zhao [a,b,*], Ying Wang [a,b], Feng Zhang [a], Manimaran Ramasamy [a], Zhijun Qiao [c]

[a] *College of Computer Science and Technology, Hengyang Normal University, Hengyang 421002, China*
[b] *Hunan Provincial Key Laboratory of Intelligent Information Processing and Application, Hengyang 421002, China*
[c] *University of Texas Rio Grande Valley, School of Mathematical and Statistical Sciences, TX, 78520, USA*

ARTICLE INFO

Keywords:
Sketch to portrait
Conditional GANs
Edge constraint
Pix2Pix
Laplace operator

ABSTRACT

A novel method for generating color portraits from sketch images using the edge constraint algorithm and generative adversarial networks (GANs) is proposed in this paper. In converting sketches into color portraits, the details of portrait output by GANs are often blurred and unrealistic. A new method with edge constraint is proposed in this work to address this issue. The image generated from generator network and its edge generated from the followed edge network are combined and provided to the discriminator for authenticity identification. Experiments show that the portrait output by the proposed method provides a more clear and realistic edge than a Pix2Pix model and has a better ability to generate color portraits from sketches compared with other common methods. The average structural similarity index measure (SSIM) value of the proposed method is 82.78%, while the values obtained by other methods and Pix2Pix are 42.99% and 78.60%, respectively.

1. Introduction

Automatically converting sketch images into portrait images has significant application value in the fields of digital entertainment, art, law enforcement, and other industries. In the law enforcement industry, there is a big difference between an image drawn by the police or a medical expert according to an oral description from an eyewitness and a portrait image with low recognition. Manually generating a real color portrait requires a skilled labor force with experience in drawing and painting as well as investigation. Automatic generation of realistic facial portraits based on sketch images with generative adversarial networks (GANs) [1,2] can improve the possibility of identification and enhance police efficiency in solving cases. As such, converting sketches into portraits is widely studied in the field of artificial intelligence [3–7].

Employing convolutional neural networks (CNNs) for image style transfer [8–10] is a common image generation technique used across various fields [9–10]. However, in the process of image generation by CNNs, it is necessary to provide the original image to be transferred, as well as the corresponding target image. For example, in the case of style conversion by Gatys et al. [9], the neural network must be told what the target image for image generation is, and an appropriate loss function is then designed correspondingly.

This paper is for regular issues of CAEE. Reviews processed and approved for publication by the co-Editor-in-Chief Huimin Lu.
* Corresponding author at: College of Computer Science and Technology, Hengyang Normal University, Hengyang 421002, China.
E-mail address: happyday.huihuang@gmail.com (H. Zhao).

https://doi.org/10.1016/j.compeleceng.2021.107338
Received 15 March 2021; Received in revised form 23 July 2021; Accepted 24 July 2021
Available online 10 September 2021

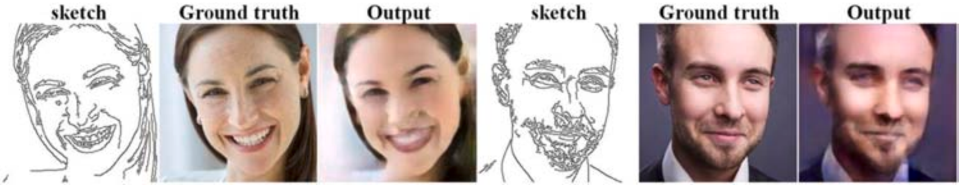

Fig. 1. Unsatisfied results, with fine details and true textures missing.

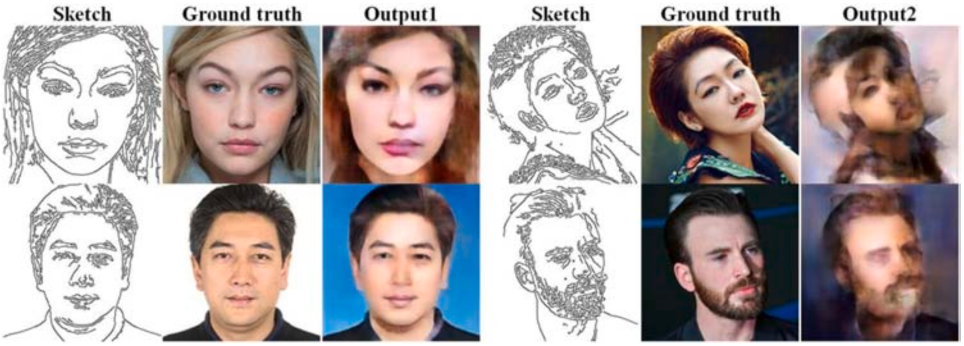

Fig. 2. Unsatisfied results with different degrees of failure.

The purpose of designing the loss function is to provide a clear direction of convergence during image style transferring and give the model an accurate understanding of the characteristics of the target image. In this paper, an improved Pix2Pix model [11] is selected for converting a sketch to a color portrait. The loss function is also used to constrain the sketch to generate the portrait.

Standard graphics theory can be used to simulate the contours, skin color, and light within a portrait and can be used to render a realistic photo image [12–14]. Although existing graphics algorithms perform well in practice, it is expensive and time-consuming to build and edit the visual environment, and every aspect of the world must be modeled clearly. Pix2Pix can attempt to generate portraits with the help of models learned from a lot of training, but a large number of test results indicate that it still struggles to perform successfully: the contours of the portraits are not clear, and fine details and portrait textures are missing in some results. We improve the Pix2Pix generation model to solve these problems in this work. In the process of image generation, the edge information of the image is extracted, its contour is optimized, and the new model constrains the convergence direction of portrait generation. Experiments show that our modification based on Pix2Pix can effectively solve the edge blur problem of facial image generation and provides some reference value for other similar image generation application scenarios.

We will introduce the related work of this paper in the next Section 2, and describe the proposed method, improved model and loss function in Section 3. Section 4 shows the experimental results of the improved model compared with other models, and finally is the conclusion of this work.

2. Related work

2.1. Pix2Pix mode

As a special GANs model, Pix2Pix explains the problem of image conversion as the mapping relationship between the pixels of the input and the output directly. The generator uses a U-Net structure, while the discriminator uses a convolutional "PatchGANs" classifier. Different from CNNs and other GANs, Pix2Pix provides a general solution to the image conversion problem. The loss function of the image conversion problem is learned automatically through a large amount of conditional training, which is used to constrain the direction of image conversion and convergence. However, in the process of image transformation, if the image structure of the semantic image and target image is significantly different, for example, when converting a sketch into a real face image or transforming the style of real photos, fine details and accurate textures are often missing [13], as shown in Fig. 1. In model training, Ting Chun Wang manually added edge information to each label image to solve the problem of edge blur in the converted images [13].

As shown in Fig. 1, the silhouettes and colors of the output are preserved while the fine details and true textures are missing. The output shows that, although the color and outline of the face are already created, the nose, mouth, teeth, and ears of both faces are blurred. This situation also occurs in [13]. In the following work, we attempt to modify the network model and optimize the facial details of the generated portraits to obtain a more realistic portrait.

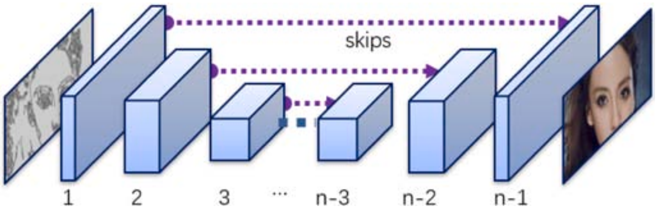

Fig. 3. The U-Net generator with skips between each layer i and layer n-i. The lowest resolution is 1 × 1 pixels, and each blue box corresponds to a multi-channel feature map. The arrows denote the skip connections between mirrored layers.

2.2. Generation of real image

Many researchers have transformed image to image with the help of adversarial learning. Given input-output image pairs as training data, the original images are converted to the target images under the guidance of the style images. For example, a recent Pix2Pix framework [11] used image-conditional GANs [15] for different applications, such as generating satellite views from Google Maps and transforming user sketches into cats. Recently, Chen and Koltun [16] stated that generated results often lacked fine details and realistic textures due to the instability and constraints of training. Wang et al. [13] used multi-scale generators and discriminators to generate high-resolution images and imported edge information images through a boundary map to solve the problem of image edge blur so as to generate images consistent with real photos. In order to obtain an obvious edge of the object after conversion, the corresponding edge image was prepared for each image. A selection of sketch to portrait results is provided in Fig. 2.

As illustrated in Fig. 2, the two output portraits (Output2) on the right are worse than those on the left (Output1). This is because most images used in the training set are front portraits, which are similar to the two images shown on the left side of Fig. 2, while the image on the right shows special cases in the training data set. All the test outputs in Fig. 2 do not exist in the training set. The skin color and background of portraits in the training set are not identical, and, consequently, there are some differences in the skin and background of outputs and the ground truth images. These differences can be observed in Fig. 2 and indicate that using a test image that is similar to the training set will provide better results. Therefore, in order to obtain satisfactory results, a training set with closer contours must be employed, and the test image must be closer to the image of the training set. However, high demand for similarity between the test and training images will limit the flexibility of the model in generating images, and the use of a boundary map will also increase the difficulty of creating a data set. Super resolution is the traditional approach used in the field of computer vision [17, 18]. In this work, we attempt to modify this model and extract edge information from the data set itself.

A novel method for generating color portraits from sketch images with edge constraint algorithm and generative adversarial networks is presented here. The main contributions of this paper are summarized as follows:

- It is determined that Pix2Pix models acquire learning ability by training, but the contours and details of the output image will be blurred when the input images and targets images are not similar enough. The edge constraint is added in the process of sketch to portrait generation, and experiments show that this addition to the model can improve the clarity of the generated portrait.
- The conditional GANs model is modified by adding an edge network (E) to obtain the contour information in the real portrait and generated portrait, which is output from the generator network (G). We then modify the loss constraint in the Pix2Pix model. The new loss function can evaluate the deviation of G output results and add the image contour information, which is output from E. The quality of the contour is improved, the outline is clear, and the details are accurately maintained.
- In a large number of experiments, we observe that the targeted constraints in the loss function can train and strengthen the model according to our expectations. The modified model can output satisfactory results in a small data set. For example, with less than 200 images in the training set, the model is able to finish training in less than 3 hours.

3. Method and design

In many Pix2Pix experiments, we find that the edges of the output image are blurred. The loss function can be learned by the Pix2Pix model automatically and is used to guide the generation from sketch to real portrait. The loss function represents the deep correspondence between the original images and the targets. However, from the perspective of human understanding of images, the difference between the input and the target images is multifaceted. Therefore, the model attempts to balance various differences, such as color, texture, contour, and so on. If the input images have a more similar structure, the difference can be understood more precisely by the loss function. However, the more similar the structure of the input images, the worse the scalability of the model. Therefore, edge constraint is added to the model to enhance its attention to the image edge. When the discriminator network judges whether the image is true or false, it must judge the image itself while also taking the edge as an important basis for judgment.

3.1. U-Net generator

The generator contains multiple convolution and deconvolution layers and is trained to convert the input sketches into corresponding color portraits [16]. From the perspective of representation, the input sketches and the output portraits are quite different,

0	1	0
1	-4	1
0	1	0

(a)

1	1	1
1	1	1
1	1	1

(b)

Fig. 4. The kernels of max pool (a) and average pool (b).

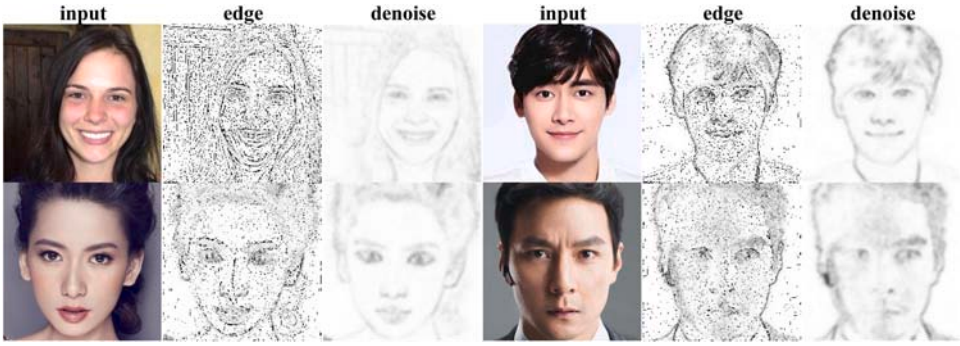

Fig. 5. A selection of intermediate results showing the internal process of E.

but the underlying structures of the images are similar. Therefore, skips are added in the U-Net generator between the symmetric convolutional layer (i) and deconvolutional layer (n-i) to share low-level information about the image, where n is the maximum number of layers in the generator. Transmitting this information in the generator network can provide some deep-seated feature constraints for image conversion [11]. The generator network is shown in Fig. 3. In our work, the input sketches and the output portraits are 256 * 256 pixels, and the expansion and contraction paths both have eight layers.

As shown in Fig. 3, the U-Net generator consists of encoding and decoding networks. In training, the dimension of data is reduced by the encoding network, while the decoding network recovers metadata from low-dimensional data. In both networks, weights are initialized randomly and trained by minimizing the difference between the original data and its reconstructed data. Skip connections concatenate the activation values from layer i (encoder) to layer n-i (decoder). The energy function of the joint distribution (v, h) of a visible layer and hidden layer can be expressed as Eq. (1).

$$E(v,h) = -\sum_{i\in pixels} b_i v_i - \sum_{j\in features} b_j h_j - \sum_{i,j} v_i b_j w_{ij} \tag{1}$$

where v_i and h_j are the binary states of pixel i and feature j, respectively; b_i and b_j are their biases; w_{ij} is the weight between them.

3.2. Discriminator

We employ the Markovian discriminator of pixel to pixel mapping, which contains a discriminator sliding window with the size of n × n (smaller than the picture size). The discriminator is used to determine whether each patch block in the portrait is real or fake. Finally, the average value of these discrimination results is taken as the final output of the discriminator. Testing of different patch block sizes illustrates that the discriminator shows excellent discrimination ability when n is set to 1. We follow the standard approach to optimize the networks, in which the discriminator and generator are trained and optimized alternately [19]. As widely used constraint algorithms, mini-batch stochastic gradient descent (SGD) and the Adam solver are used and applied later [20].

Intuitively, the discriminator is completely composed of convolutional layers, the final output is an N × N matrix, and the mean of the output matrix is taken as the true/false output. Markovian discriminator can maintain high resolution and detail in the style transfer process.

3.3. Edge extraction and denoising

Both the real portrait and generated image output from generator must go through two steps in E: extraction of the edge and denoising. The internal structure of E can be expressed as conv2d > maxpool > conv2d > avgpool. The first iteration of convolution and max-pooling is used to obtain the image edge, while the second process of convolution and average pooling works to denoise the edge. The kernels of the max pool and the average pool are shown in Fig. 4.

As shown in Fig. 4, a convolution operation is carried out using the kernel of max pool (a), and then max pooling is performed. In the subsequent average pooling, we use the mean filtering method to denoise the portrait's output from the max pool. The Laplace

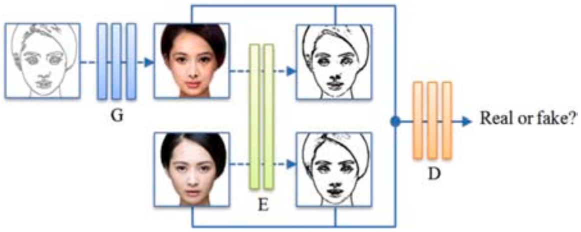

Fig. 6. Improved model architecture. E is used to generate the edge image for the real and fake images from G. Portrait image and edge image are input to D together to judge if the image is real or fake.

Fig. 7. Comparison of generated portraits before and after adding edge constraints.

algorithm is used to obtain the edge of the image for the max pool. The Laplace algorithm can be expressed as shown in Eq. (2).

$$\nabla^2 f = [f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1)] - 4f(x,y) \tag{2}$$

where x and y represent the coordinates of the pixel in the center of the convolution kernel. Mean filtering can be expressed as denoted in Eq. (3).

$$g(x,y) = \frac{1}{M}\sum_{f\in s} f(x,y) \tag{3}$$

where x and y represent the coordinates of a pixel. After denoising by mean filtering, the value of the center pixel is the average value of the nine surrounding pixels.

The function of E is tested separately, and some of these results are shown in Fig. 5.

As can be observed in Fig. 5, the edge output from the first convolution and max-pooling contains a lot of noise initially, and then the second convolution and average pooling denoises it. Finally, these edge images are input into the discriminator (D) for real or fake judgment after adjustment.

3.4. Model and loss

Blurred facial features and contours are unacceptable when generating portraits from sketches. Inspired by the methods and successes in [13], we improve the model [11] and add an edge network for edge detection and constraint.

As shown in Fig. 6, both the real portrait image and the generated image output from G are input into E. The edges of these images are obtained and output after the convolution and pooling process of E. Both the generated image and the real portrait are then combined with the corresponding edge and subsequently input to the D network for real or fake judgment.

Table 1
SSIM values before and after edge constraints are added.

	Pic1	Pic2	Pic3	Pic4	Pic5	Pic6
without Edge Constraint [11]	76.54%	72.34%	77.58%	81.27%	67.77%	61.66%
with Edge Constraint	**81.85%**	**81.50%**	**81.21%**	**83.04%**	**69.54%**	**69.36%**

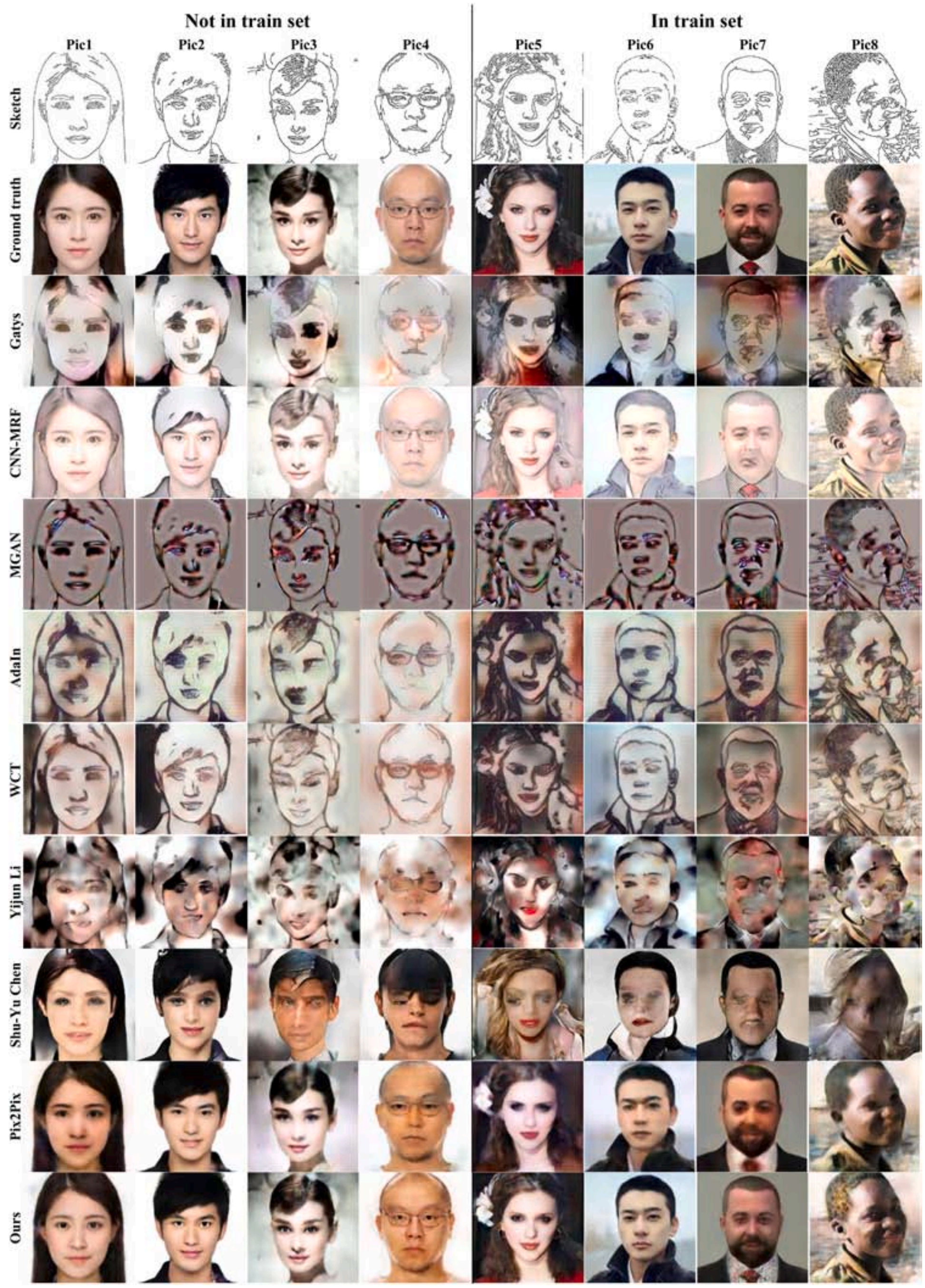

Fig. 8. Comparison of some results from different methods for the same sketch-portrait pairs.

The common conditional GANs can be expressed as Eq. (4).

$$\mathscr{L}_{cGAN}(G,D) = E_{x,y\sim P_{data}(x,y)}[logD(x,y)] + E_{x\sim P_{data}(x),z\sim P_z(z)}[\log(1-D(x,G(x,z)))] \tag{4}$$

where G attempts to minimize the objective, while D attempts to maximize it. Finally, after continuous training and constraint, G and D reach a balance as the final output. Over a large number of experiments from sketch to portrait, we find that the balanced results of G and D provide a comprehensive output, and the model attempts to balance the color, structure, and illumination of the generated portrait. However, the blurring of facial features in the generation of portraits is obvious, which is contrary to our aim to generate realistic portraits. In response, we improve the objective and add judgment of the image edge. The design of the improved model is denoted in Eq. (5).

$$G^* = argmin_G max_D \mathscr{L}_{cGAN}(G,D) + \lambda_1 \mathscr{L}_{L1}(G) + \lambda_2 Edge_{L1}(G) \tag{5}$$

where λ1 and λ2 control the weight of the two items. The importance of the edge can be adjusted by controlling λ1 and λ2. The constraint of image edge information is added to the objective, and it is used to guide the generator to synthesize the portrait with clearer and more vivid contours.

4. Experiments and discussion

In the experiments of training and testing, both sketches and color portraits were 1–3 channel images, with a resolution of 256 × 256. Our training set consisted of 200 pairs of sketches and portraits, and the model could be trained quickly on a dataset of this size. Training took approximately 3 hours with 200 epochs on a single NVIDIA GPU (GeForce GTX 1080) and took less than 1 minute to complete the synthesis of 100 real portraits in the test.

4.1. Analysis of the edge constraint

We initially selected the same dataset as in Fig. 2 for the training and testing. The results using the improved model are shown in Fig. 7 and illustrate the effect of the model improvement on the output. Both the original model and the improved model adopted the same training set and parameters.

In Fig. 7, "a/b" represents the sketch-portrait pair images for the test. In the image, "without Edge Constraint" are the output results before adding the edge constraint [11], and "with Edge Constraint" are the output results after adding the edge constraint. It can be observed that the output portraits without and with edge constraints have similar contours and colors. However, it can be seen that with edge constraint, the output portrait obtains clearer contour and detail features. For example, the facial features, hair, and clothes in the portrait have more realistic details. The edge outputs of Pic3 and Pic5 from Fig. 7 are shown in Fig. 5. The edge constraint in Fig. 5 can provide the generated image with clearer contour features. We also tested the structural similarity index measure (SSIM) value between the results before and after edge constraint and the real image. The results are provided in Table 1.

Table 1 shows that the SSIM value between the original image and the generated image is higher than that without edge constraint. The change of these data shows that adding edge optimization to the model is conducive to generating more realistic portraits.

Through numerous tests, we compared the output results before and after edge constraint, and found that this improvement was common. Therefore, we can draw a conclusion that the portrait output from the Pix2Pix model with edge constraint has clearer and more realistic details. Thus, the edge constraint in the objective can significantly improve synthesis quality.

4.2. Comparison with other methods

In order to evaluate the effectiveness and performance of our improved model, we used the same dataset for training and testing. Some of the images for testing were from the training set, while others were not. As an extensive research field, portrait transformation has been widely studied and applied. We compared our results with a selection of these other methods from sketch to portrait. A selection of comparison results are provided in Fig. 8.

It can be seen from the comparison results in Fig. 8 that the methods of Gatys [9] and MGAN [21] do not possess the ability to synthesize a sketch into a portrait. The CNN-MRF [22] method attempts to achieve a balance between the sketch and the target portrait. When generating a new portrait, the target portrait corresponding to the sketch must be prepared in both Gatys and CNN-MRF methods. MGAN attempts to learn the texture features of the portrait, but the final results are far from our expectations. The methods of AdaIn [23] and WCT [23] try to keep the content information of sketches in style conversion but as a result, the model only attempts to thicken the contour lines and the results are far from the real portraits. These two methods attempt to obtain the overall texture of the portrait, however, according to the results, they do not accurately understand the conversion relationship between the sketch and the portrait. The method of Li et al. [24] obtains the general outline of the portrait in the transformation, but the details remain very vague. The method of Chen et al. [25] can convert the sketch into a portrait when the lines of the sketch are simple, such as in Pic 1 and 2. However, the facial features of the output portrait are very fuzzy when the line sketch is complex, such as in Pic 3-8. Comparatively, the results from Pix2Pix [11] and our model have real facial features. When these portraits are generated, the corresponding portraits for input sketches are not required in the two models ([11] and ours). Moreover, we can also observe that, compared with the output of the Pix2Pix model, our results are clearer with more realistic details, such as facial features. In the experiments using our method, the portrait image corresponding to the sketch image is not prepared, and the automatic conversion ability of the model is completely

Table 2
SSIM values between the results and the ground truth in Fig. 8.

	Pic1	Pic2	Pic3	Pic4	Pic5	Pic6	Pic7	Pic8
Gatys[9]	45.59%	45.79%	50.02%	41.35%	64.81%	51.77%	24.68%	28.56%
CNN-MRF[22]	62.01%	46.63%	68.39%	64.83%	85.28%	66.30%	41.26%	58.71%
MGAN[21]	38.45%	48.35%	45.76%	43.10%	53.54%	47.19%	21.40%	26.66%
AdaIn[23]	19.99%	23.86%	28.69%	48.57%	12.30%	27.33%	27.99%	17.15%
WCT[23]	33.01%	28.76%	39.40%	62.53%	20.08%	44.40%	43.77%	23.07%
Yijun Li[24]	38.95%	34.74%	38.57%	63.42%	30.71%	44.75%	45.24%	29.22%
Shu-Yu Chen[25]	59.91%	66.29%	50.02%	58.13%	27.42%	57.89%	57.62%	33.04%
Pix2Pix[11]	77.11%	79.72%	74.05%	84.75%	81.53%	78.91%	69.58%	76.67%
Ours	**79.91%**	**85.77%**	**78.11%**	**85.42%**	**85.16%**	**85.91%**	**74.51%**	**79.60%**

learned from the training and objective. Some experimental results are shown in Fig. 8 in the column "Not in train set". The column "In train set" means that these test images also exist in the training set. Both "Not in" and "In" results obtain fine detail and accurate texture compared with the results of Pix2Pix. According to our results, the portraits in the training set are close to the ground truth in color, while the portraits not in the training set are obviously different from the ground truth in color. However, compared with pix2pix, these two types of output obtain realistic portrait details. Moreover, we test the SSIM values between the results and the ground truth, and the results are provided in Table 2.

As can be seen from Table 2, the SSIM values of our results perform better than other methods, and this result is consistent with the generation ability of the models shown in Fig. 8.

5. Conclusion

Our experimental results showed that the boundaries and contours of the portrait were clearer and more natural after the edge constraint was added to conditional GANs. Compared with eight other general models, the portrait output of the proposed model obtained better portrait features. The findings also illustrated that different constraints and weights were required for loss in different applications, and this constraint enabled the model to pay more attention to specific details in the process of generating new images. However, although we added edge information constraints to the model, some differences between the final test results and the real photos remained. On the one hand, the sketch provided to the model may not have fully described all the information required by the model to generate a vivid portrait. On the other hand, the model may need further constraints to generate a realistic portrait. We believe that our method and the improved model can provide some assistance in the fields of forensic mapping and police investigation.

Declaration of Competing Interest

We declare that we have no conflicts of interest to report regarding the present study.

Acknowledgment

This work is supported by National Natural Science Foundation of China (No. 61772179), the Hunan Provincial Natural Science Foundation of China (No. 2019JJ40005, 2020JJ4152), Scientific Research Fund of Hunan Provincial Education Department (No. 18A333). the Double First-Class University Project of Hunan Province (No. Xiangjiaotong [2018]469), Postgraduate Research and Innovation Projects of Hunan Province (No. Xiangjiaotong [2019]248-998), the science and technology innovation Program of Hunan Province (No.2016TP1020), Hengyang guided science and technology projects and Application-oriented Special Disciplines (No. Hengkefa [2018]60-31), Postgraduate Scientific Research Innovation Project of Hunan Province(CX20190998), Degree & Post-graduate Education Reform Project of Hunan Province (2019JGYB266,2020JGZD072), Industry University Research Innovation Foundation of Ministry of Education Science and Technology Development Center (2020QT09), Hengyang technology innovation guidance projects(Hengcaijiaozhi [2020]-67), Postgraduate Teaching Platform Project of Hunan Province(Xiangjiaotong [2019]370-321). We also thank Gatys et al. [9], Goodfellow et al. [19], Wang et al. [13] and Zhu et al. [14] for sharing their codes, and NVIDIA for the GPU donation.

Author statement

I have made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND I have drafted the work or revised it critically for important intellectual content; AND I have approved the final version to be published; AND I agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Qingyun Liu, Huihuang Zhao.

References

[1] Kawai Y, Seo M, Chen YW. Automatic generation of facial expression using generative adversarial nets. In: Proceedings of the IEEE 7th global conference on consumer electronics; 2018. p. 278–80.
[2] Lin JP, Liu D, Li HQ, Wu F. Generative adversarial network-based frame extrapolation for video coding. In: Proceedings of the IEEE visual communications and image processing; 2018. p. 1–4.
[3] W. H. Xia, Y. J. Yang and J. H. Xue, "Cali-sketch: stroke calibration and completion for high-quality face image generation from poorly-drawn sketches," arXiv preprint *arXiv:1911.00426,*2019.
[4] Wang LD, Sindagi V, Patel VM. High-quality facial photo-sketch synthesis using multi-adversarial networks. In: Proceedings of the IEEE international conference on automatic face & gesture recognition; 2018. p. 83–90.
[5] Lu YY, Wu SZ, Tai YM, Tang CK. Image generation from sketch constraint using contextual GAN. In: Proceedings of the European conference on computer vision; 2018. p. 213–28.
[6] Chen WL, Hays J. SketchyGAN: towards diverse and realistic sketch to image synthesis. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition; 2018. p. 9416–25.
[7] Han X, Hou K, Du D, Qiu Y, Cui S, Zhou K, Yu Y. Caricatureshop: personalized and photorealistic caricature sketching. IEEE Trans Visual Comput Graph 2019;26 (7):2349–61.
[8] Zhao HH, Liu H. Multiple classifiers fusion and CNN feature extraction for handwritten digits recognition. Granular Comput 2020;5(3):411–8.
[9] Gatys LA, Ecker AS, Bethge M. Image style transfer using convolutional neural networks. In: Proceedings of the conference on computer vision & pattern recognition; 2016. p. 2414–23.
[10] Lou GX, Shi HZ. Face image recognition based on convolutional neural network. China Communications 2020;17(2):117–24.
[11] Isola P, Zhu JY, Zhou TH, Efros AA. Image to image translation with conditional adversarial networks. In: Proceedings of the conference on computer vision & pattern recognition; 2017. p. 5967–76.
[12] Kazemi H, Taherkhani F, Nasrabadi NM. Unsupervised facial geometry learning for sketch to photo synthesis. In: Proceedings of the international conference of the biometrics special interest group; 2018. p. 1–5.
[13] Wang TC, Liu MY, Zhu JY, Tao A, Kautz J. High-resolution image synthesis and semantic manipulation with conditional GANs. In: Proceedings of the conference on computer vision & pattern recognition; 2018. p. 8798–807.
[14] Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision; 2017. arXiv:1703.10593.
[15] Mirza M, Osindero S. Conditional generative adversarial nets. In: Proceedings of the international conference of the biometrics special interest group; 2014. p. 2672–80.
[16] Chen QF, Koltun V. Photographic image synthesis with cascaded refinement networks. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 621–32.
[17] Guo Y, Chen J, Wang J, Chen Q, Cao J, Deng Z, Xu Y, Tan M. Closed-loop matters: dual regression networks for single image super-resolution. In: Proceedings of the Conference on Computer Vision & Pattern Recognition; 2020. arXiv:2003.07018.
[18] Maeda S. Unpaired image super-resolution using pseudo-supervision. In: Proceedings of the conference on computer vision & pattern recognition; 2020. arXiv: 2002.11397.
[19] Goodfellow IJ, Abadie JP, Mirza M, Xu B, Farley DW, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: Proceedings of the international conference of the biometrics special interest group; 2014. p. 2672–80.
[20] Zhang ZJ. Improved adam optimizer for deep neural networks. In: Proceedings of the IEEE/ACM 26th international symposium on quality of service; 2018. p. 1–2.
[21] Li C, Wand M. Precomputed real-time texture synthesis with markovian generative adversarial networks. In: Proceedings of the IEEE international conference on computer vision; 2016. p. 702–16.
[22] Li C, Wand M. Combining markov random fields and convolutional neural networks for image synthesis. In: Proceedings of the conference on computer vision & pattern recognition; 2016. p. 2479–86.
[23] An J, Huang S, Song Y, Dou D, Liu W, Luo J. ArtFlow: unbiased image style transfer via reversible neural flows. In: Proceedings of the conference on computer vision & pattern recognition; 2021. arXiv:2103.16877.
[24] Li Y, Fang C, Yang J, Wang Z, Lu X, Yang MH. Universal style transfer via feature transforms. In: Proceedings of the31st international conference on neural information processing systems; 2017. p. 386–96.
[25] Chen SY, Su W, Gao L, Xia S, Fu H. DeepFaceDrawing: deep generation of face images from sketches. ACM Trans Graph 2020;39(4). https://doi.org/10.1145/3386569.3392386.

Qingyun Liu received the M.S. degree in computer application technology from Jiangsu University, China, in 2014. He is currently an Assistant Professor with the College of Computer Science and Technology, Hengyang Normal University, and also an Assistant Professor with the Hunan Provincial Key Laboratory of Intelligent Information Processing and Application. His research interests include image style transfer and deep learning.

Huihuang Zhao received the Ph.D. degree from Xidian University, in 2010. He was a Sponsored Researcher with the School of Computer Science and Informatics, Cardiff University. He is currently an Associate Professor with the College of Computer Science and Technology, Hengyang Normal University. His current research interests include solder joint inspection, compressive sensing, machine learning, and image processing.

Ying Wang received the M.S. degree in computer application technology from Hunan University, China, in 2012. She is currently an Associate Professor with the College of Computer Science and Technology, Hengyang Normal University. She current research interests include machine learning and image processing.

Feng Zhang graduated from Hengyang Normal University with a bachelor's degree in 2019. Currently, studying at Hengyang Normal University, School of Computer Science and Technology, majoring in Electronics and Communication Engineering, studying for a master's degree. Her current research interests include image style transfer and video image style transfer.

Manimaran Ramasamy received the MCA degree from Bharathidasan University, India, in 2004. He is currently a Lecturer with the College of Computer Science and Technology, Hengyang Normal University, Hengyang, China, and also a member with the Hunan Provincial Key Laboratory of Intelligent Information Processing and Application. His current research interests include Human Computer Interface, Cognitive Psychology and Innovative Teaching Methodologies.