Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 4, No. 1, 2023

Research on Sketching Face Headshot Generation based on Improved CycleGAN

Zhen Liao, Guojun Lin *

School of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin, Sichuan 644000, China
*** Corresponding author:** Guojun Lin (Email: 386988463@qq.com)

Abstract: At present, sketch heads generated from realistic heads still has problems such as blurred contours and missing textures. For this reason, this work proposes a sketch head generation method based on CycleGAN. Firstly, the Self-Attention Mechanism (Squeeze-and-Excitation Networks (SENet) module is added to the UNet self-encoder; secondly, the base model is transformed into a supervised learning model so as to add constraints on the generated avatars and the real avatars. The experimental results show that the sketched avatar generated by the method in this paper has a better visual effect on the CUHK student test set with a 0.0274 improvement in SSIM value than the sketched avatar generated by the base model.

Keywords: Feature Extraction; UNet Self-encoder; Sketch Head; Supervised Learning; CycleGAN.

1. Introduction

Sketch avatar generation is to generate a corresponding sketch avatar given a real avatar. With the rapid development of image generation technology, sketch avatars have been widely used in the field of digital entertainment. Sketched avatars are used as personal avatar profiles are favored by more and more Internet users; various social networking software that converts real avatars into sketch style are also welcomed by the majority of Internet users.

At present, there are two main approaches for sketch avatar generation, which are model-driven approach and data-driven approach. The model-driven methods are mainly based on Bayesian learning [1] and multivariate output regression methods [2]. Wang and Tang [3] proposed a data-driven approach. The method is a Markov random field model based on probabilistic graphs and a local linear embedding [4] synthesis method based on subspace learning. The sketched head generated by the model-driven approach and the data-driven approach cannot capture the head detail information well compared with the artist's hand-drawn one, making the generated sketched head not similar enough to the real one, while the generated sketched head is excessively smooth lacking the sketch art style [5].

In 2014, Goodfellow et al [6] proposed Generative Adversarial Network (GAN), which has achieved great success in the field of image generation due to its powerful generative power and also one of the most rapidly developing directions of deep neural networks[7].GAN and its variants have achieved good results in image generation[8] and other fields have achieved good results and compensate the shortcomings of traditional methods .Pix2Pix[9] has achieved good results in the field of image generation, but the images generated by Pix2Pix tend to be blurred. The reason is that Pix2Pix is a single network conversion structure, which cannot guarantee the structural consistency of images before and after conversion. cycleGAN is a novel generative adversarial network model proposed by zhu et al [10], which contains a cyclic reconstruction process of two generative adversarial networks. Compared with other models, CycleGAN can map images from one domain to another and then map the synthesized images back. This dual mapping structure of the network maintains the structure of the generated images well. CycleGAN is an unsupervised learning model that uses cyclic consistency loss to constrain the correlation between the generated and input images.

The literature [11] mentions that using cyclic consistency loss, the generated images have the problem of feature hiding. UNet [12] self-encoder consists of an encoder and a decoder, and the hopping layer connection between the same layer in the encoder and decoder can greatly improve the quality of the generated images.

Therefore, in this paper, we propose a sketch head generation method based on CycleGAN and UNet self-encoder. Firstly, a self-attentive mechanism (Squeeze-and-Excitation Networks (SENet) [13] module is embedded in the UNet self-encoder to improve the model's ability to extract features. Secondly, the model is converted to a supervised learning model so as to add L1 constraints on the generated avatars and the real avatars.

2. Organization of the Text

2.1. Overall Network Framework Diagram

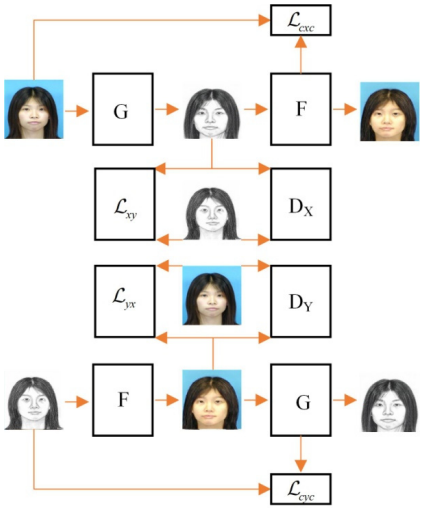

Figure 1. The overall network framework

The overall network framework of the model is shown in Fig. 1. For the purpose of later description, the set of realistic

avatars is set as X domain and the set of sketched avatars is set as Y domain. Let the distribution satisfied by the face avatars in domain X be P(x) and the distribution satisfied by the face avatars in domain Y be Q(y). The whole network consists of two generators and two discriminators. The generator G represents the mapping of X-domain image to Y-domain image, and the generator F represents the mapping of Y-domain image to X-domain image. D_X discriminator discriminates the real image from X-domain and the transformed image F(y); D_Y discriminator discriminates the real image from Y-domain and the transformed image G(x).

The conversion process of the image from X domain image to Y domain image is explained as an example. A cyclic consistency loss is introduced for the image x in the X domain to ensure that it can remain relevant to x after conversion to the Y domain. Firstly, the image is converted from the X domain to the Y domain by the generator G, i.e., the image x is converted to G(x). Then, the image G(x) converted to the Y domain is converted back to the X domain by the generator F, i.e., the image G(x) is converted to F(G(x)) by the generator F. For image x, F(G(x)) ≈ x should be satisfied after two conversions.

2.2. Embedding SENet Module in UNet Self-coding

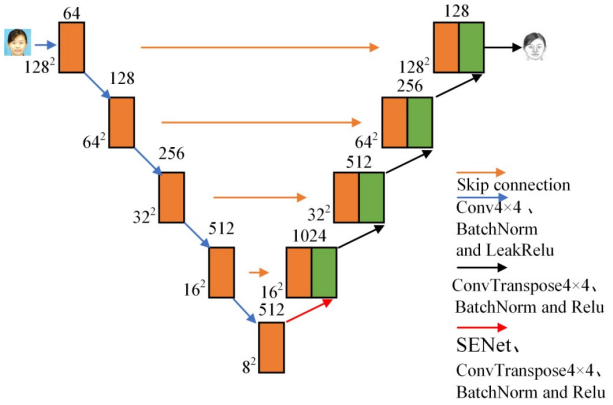

Figure 2. Embedded SENet module in UNet self-coding

The network structure of embedding SENet module in UNet self-coding is shown in Fig. 2. UNet self-coding compressively encodes the input image, and the feature map decreases continuously at this time. When the feature map reaches a minimum, the SENet module is embedded to enhance the feature extraction ability of the model; after up sampling the feature map to achieve cross-domain conversion of the image.

2.3.The Model is Converted into a Model for Permutation Supervised Learning

CycleGAN constrains the correlation between the generated image and the input image by cyclic consistency loss. The literature [14] mentions that there are hidden problems with using cyclic consistency loss for this reason, here the model is converted to a supervised learning model to improve the quality of the generated avatars.

2.4. Loss Function

2.4.1. Image Space Constraint for Generated Avatar and Real Avatar

To alleviate the cyclic consistency loss constraint, which brings about the problem of feature hiding of generated images, image L1 constraint is added to generated avatars and real avatars. l1 constraint is calculated as

$$\mathcal{L}_{xy} = \lambda_{xy}\mathbb{E}_{x_i\sim P(x)}\left\|G(x_i)-y_i\right\|_1 \quad (1)$$

$$\mathcal{L}_{yx} = \lambda_{yx}\mathbb{E}_{y_i\sim Q(y)}\left\|F(y_i)-x_i\right\|_1 \quad (2)$$

where: λ_{xy} and λ_{yx} are the weighting coefficients.

2.4.2. Generating the Adversarial Loss

Using the least squares loss as the adversarial loss of the GAN leads to a more stable training process of the model. Therefore, the GAN loss function of the model is:

$$\begin{aligned}\mathcal{L}_{c\_gan}(G,F,D) &= \mathcal{L}_{gan}(G,D_Y,X,Y)+\mathcal{L}_{gan}(F,D_X,Y,X)\\ &= \mathbb{E}_{x_i\sim P(x)}\left[\log\left(1-D_X\left(G\left(x_i\right)\right)\right)\right]+\\ &\quad \mathbb{E}_{y_i\sim P(x)}\left[\log\left(1-D_X\left(F\left(y_i\right)\right)\right)\right]\end{aligned} \quad (3)$$

2.4.3. Cyclic Consistency Loss

For each face avatar xi in the X domain mapped by the generator G and then mapped by the generator F should be as consistent as possible with the original image xi. Similarly, the same is true for each face avatar yi in the Y domain. The cyclic consistency loss is calculated as

$$\mathcal{L}_{cyc} = \lambda_{cyc}\mathbb{E}_{y_i\sim Q(x)}\left\|G(F(y_i))-y_i\right\|_1 \quad (4)$$

$$\mathcal{L}_{cxc} = \lambda_{cxc}\mathbb{E}_{x_i\sim P(x)}\left\|F(G(x_i))-x_i\right\|_1 \quad (5)$$

where: λ_{cyc} and λ_{cxc} are the weight coefficients.

2.4.4. Ontology Mapping Loss (Identity Loss) Loss

For each face avatar xi in the X domain, it should be as consistent as possible with xi after being transformed by the generator F. Similarly, for each face avatar yi in domain Y, the same is true. The ontology mapping loss is calculated as

$$\mathcal{L}_{yc} = \lambda_{yc}\mathbb{E}_{y_i\sim Q(x)}\left\|G(y_i)-y_i\right\|_1 \quad (6)$$

$$\mathcal{L}_{xc} = \lambda_{xc}\mathbb{E}_{x_i\sim P(x)}\left\|F(x_i)-x_i\right\|_1 \quad (7)$$

where: λ_{yc} and λ_{xc} are the weight coefficients.

2.4.5. The Total Loss Function of the Improved Model

In summary: the total loss function of the improved model is

$$\mathcal{L}_{total} = \mathcal{L}_{c\_gan}+\mathcal{L}_{cyc}+\mathcal{L}_{cyc}+\mathcal{L}_{yx}+\mathcal{L}_{xy}+\mathcal{L}_{yc}+\mathcal{L}_{xc}+\mathcal{L}_{sx}+\mathcal{L}_{sy} \quad (8)$$

3. Experimental Setup, Results and Analysis

To verify the effectiveness of the method in this paper, comparative experiments with different models are conducted in the CUHK student face dataset.

3.1. Data Set Introduction

There are 188 sketched face avatars in the CUHK student dataset. 88 were selected as the training set and 100 as the test set. The sketch images are drawn from real face heads of artists.

3.2. Evaluation Metrics

Structural Similarity (SSIM) [15], is a metric to measure the similarity of two images. The larger the SSIM value is, the more realistic the generated image is to the real image.

3.3. Experimental Setup

3.3.1. Experimental Environment Setup

The experiments are conducted under the windows 10 system environment, the GPU is NVIDIA GeForce RTX 3060, the display size is 12GB, the CPU is Intel(R) Core (TM) i7-

4770, and the deep learning framework of pytorch is used.

3.3.2. Experimental Parameter Settings

In the CUHK student face dataset training, the Adam optimizer with momentum 0.5 is selected for training, the initial learning rate of both the generator and the discriminator is set to 0.0002, the learning rate "MultStepLR" is dynamically adjusted, the batch size is set to 1, and the number of iterations is 200.

In the CUHK dataset training, λ_{cyc} and λ_{cxc} are both taken as 10, λ_{yc} and λ_{xc} are both taken as 0.5, and λ_{xy} and λ_{yx} are taken as 2.

3.4. Analysis of Experimental Results

In this paper, we conducted comparison experiments with the mainstream image generation methods (LLE [16], Pix2Pix, CycleGAN and Combogan [17]). The sketched face avatars generated by each model on the CUHK student face test set are shown in Figure 6, SSIM values Table 1.

From Table 1, it can be seen that: the SSIMs of the sketched avatars generated by the method in this paper are all higher than those of other models.

Table 1. SSIM values of sketched heads generated by each model on CUHK student

Models	SSIM
LLE	0.4624
Pix2Pix	0.4683
CycleGAN	0.5931
Combogan	0.4075
Methods in this paper	0.6205

It can be seen from Fig. 3 that the sketch face head generated by this paper is clearer, more distinctive and richer in facial details than the sketch face head generated by other models, and has a better subjective feeling.

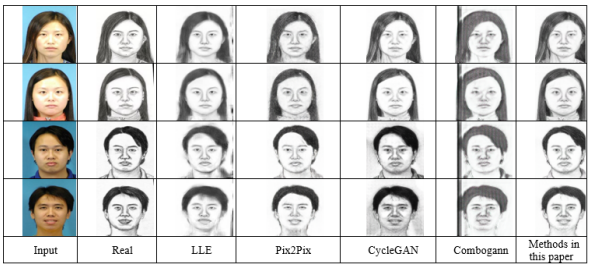

Figure 3. Sketch heads generated by each model on the CUHK student test set

4. Conclusion

In this paper, a sketch head generation method is proposed. Firstly, in. Second, in order to further improve the quality of generated sketch face avatars, this paper converts the model into a supervised learning model by adding the generated avatar with real avatar L1 constraint. The experimental results show that the sketch avatar generated by the method in this paper is better than the base model in both subjective perception and objective evaluation index.

References

[1] Zhang S, Gao X, Wang N, et al. Face sketch synthesis via sparse representation-based greedy search [J]. IEEE transactions on image processing, 2015, 24(8): 2466-2477.

[2] Chang L, Zhou M, Deng X, et al. Face sketch synthesis via multivariate output regression [C] // Human-Computer Interaction. Design and Development Approaches: 14th International Conference, HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part I 14. Springer Berlin Heidelberg, 2011: 555-561.

[3] Wang X, Tang X. Face photo-sketch synthesis and recognition [J]. IEEE transactions on pattern analysis and machine intelligence, 2008, 31(11): 1955-1967.

[4] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. science, 2000, 290(5500): 2323-2326.

[5] Zhou H Q, Cao L, Du C N. Multi-discriminator recurrent generative adversarial network for sketch face synthesis[J]. Computer Engineering and Applications,2021,57(03):231-238.

[6] Creswell A, White T, Dumoulin V, et al. Generative adversarial networks: An overview[J]. IEEE signal processing magazine, 2018, 35(1): 53-65.

[7] Yan B., Zhang J. Lin. A study on the status of image translation based on generative adversarial networks[J]. Foreign Electronic Measurement Technology,2019,38(06): 130-134. DOI: 10.19652/j.cnki.femt.1801330.

[8] Xia Guangyou. Research on colorization method of grayscale images of farmers' paintings based on generative adversarial network [D]. Qinghai Normal University, 2022. DOI: 10. 27778 / d.cnki.gqhzy.2022.000550.

[9] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.

[10] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]// Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.

[11] Chu C, Zhmoginov A, Sandler M. Cyclegan, a master of steganography [J]. arXiv preprint arXiv:1712.02950, 2017.

[12] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.

[13] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[14] Mao X, Li Q, Xie H, et al. Least squares generative adversarial networks [C]// Proceedings of the IEEE international conference on computer vision. 2017: 2794-2802.

[15] Wang L T, Hoover N E, Porter E H, et al. SSIM: A software levelized compiled-code simulator[C]//Proceedings of the 24th ACM/IEEE Design Automation Conference. 1987: 2-8.

[16] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. science, 2000, 290(5500): 2323-2326.

[17] Anoosheh A, Agustsson E, Timofte R, et al. Combogan: Unrestrained scalability for image domain translation [C]// Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018: 783-79.