Face Image Generation for Illustration by WGAN-GP Using Landmark Information

Miho Takahashi
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
miho.takahashi@akane.waseda.jp

Hiroshi Watanabe
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
hiroshi.watanabe@waseda.jp

***Abstract*— With the spread of social networking services, face images for illustration are being used in a variety of situations. Attempts have been made to create illustration face images using adversarial generation networks, but the quality of the images has not been sufficient. It would be much easier to generate face images for illustrations if they could be generated by simply specifying the shape and expression of the face. Also, if images can be generated using landmark information, which is the location of the eyes, nose, and mouth of a face, it will be possible to capture and learn the features of the face. Therefore, in this paper, we propose a method to generate face images for illustration using landmark information. Our method can learn the location of landmarks and produce high quality images on creation of illustration face images.**

I. Introduction

Face illustrations are used as icons for social networking services, avatars for games, posters, and so on. Generative Adversarial Networks (GANs) [1] can be used to generate illustration face images, although GANs are used as a natural image generation method. However, some of the images produced by these methods cannot be recognized as faces because they do not have eyes or the mouth is in the wrong position. Another problem is that it is not possible to specify the shape and expression of a face when generating a regular GAN. Therefore, the ability to easily create face illustrations by simply specifying the shape and expression of the face will expand the possibilities of face images for illustration. Therefore, there is no example of face image generation using facial landmarks contained in a single image.

II. Related Work

Another study that incorporates facial landmarks into image generation is Landmark Assisted CycleGAN for Cartoon Face Generation [2] proposed by Ruizheng Wu et al. in 2019. In their work, they successfully generated high-quality images by using facial landmark information in CycleGAN [3]. However, CycleGAN requires data from two different images, one before and one after the transformation; this method cannot be used when generating images with GAN because there is no paired correct image.

III. Wasserstein GAN-GP

Wasserstein GAN (WGAN) [6] was proposed by Martin Arjovsky et al. in 2017 as a stabilization method for GANs. WGAN improves the instability of GANs such as gradient loss and mode collapse by measuring and learning the Wasserstein distance between the generator output and the dataset.

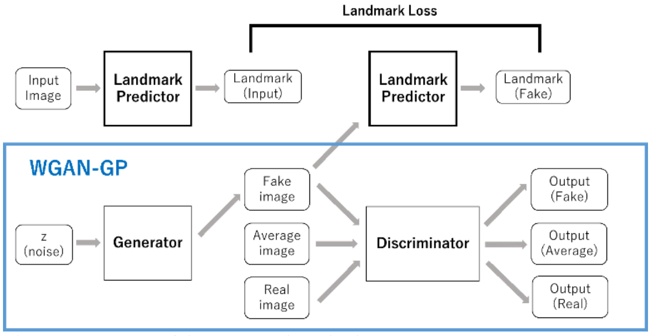

Fig. 1. Structure of the proposed method

The WGAN improves the instability of GAN such as gradient loss and mode collapse. However, depending on the parameters, the Weight Clipping used in WGAN may not learn well, or may suffer from instability such as gradient loss. In 2017, they proposed Wasserstein GAN-GP (WGAN-GP) [7], which is an improved version of WGAN. WGAN-GP successfully stabilizes WGAN by adding Gradient Penalty.

In this paper, we focus on the improvement of WGAN-GP in face illustration generation.

IV. Proposed Method

The structure of the method proposed in this paper is shown in Fig. 1. The Landmark Predictor is added to the basic structure of WGAN-GP and Landmark Loss is added to the loss function so that the generated image can reproduce the landmarks in the input image. Landmark Loss is a loss function that represents the distance between the landmarks in the input image and the landmarks in the image created by the generator. The loss function, L_{Land}, will be of the form:

$$L_{Land} = \lambda \| L(\tilde{x}) - l \|_2 \quad (1)$$

where $\tilde{x}$ is the generated image, $L(\tilde{x})$ is the landmarks in the generated image predicted by Landmark Predictor, and l is landmarks in the input image.

The Landmark Predictor consists of Convolutional Neural Networks (CNN) [6] which are layered with convolutional and pooling layers. It is pre-trained to output landmarks when an image is input.

V. Experiment

A. Dataset

We use 4,084 images from the Bitmoji face image dataset Bitmoji Faces [7], which is available on kaggle. In this paper,

978-1-6654-3676-2/21/$31.00 ©2021 IEEE

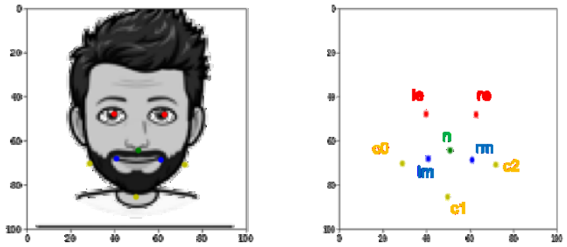

Fig. 2. How to designate landmarks.

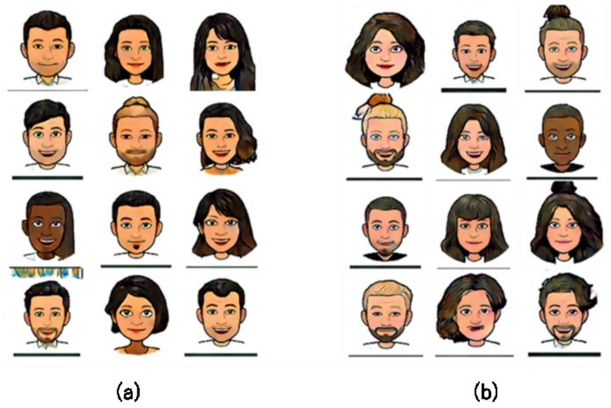

Fig. 3. Generated images of a regular quality group by (a) the proposed method and (b) WGAN-GP.

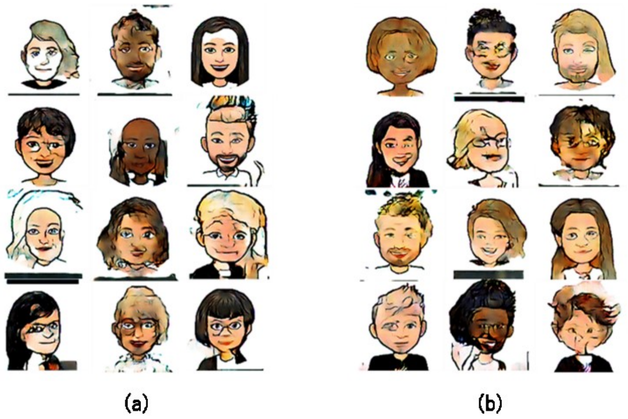

Fig. 4. Generated images of a poor quality group by (a)the proposed method and (b) WGAN-GP.

TABLE I. EVALUATION RESULT

	Land mark Loss	*FID*
WGAN-GP	0.2203	136.1
Proposed Method	0.2180	132.8

we normalized the size of the images to 128 × 128 [pixels] before using them.

The landmarks to be detected are the eyes, nose, mouth, and chin, and the specification method is shown in Fig. 2.

B. Result

The generated images of a regular and poor quality groups by the proposed method and WGAN-GP are shown in Fig. 3 and Fig. 4. The classification of regular and poor quality groups was determined by the human eye. The evaluation result using the distance between the landmarks and FID [8] of the input and generated images are shown in Table I.

VI. DISCUSSION

As shown in Table I, the landmark distance of the proposed method was smaller than that of WGAN-GP, which means slightly better. However, it was not possible to generate an image with a perfect match between the input image and the landmarks just by getting a little closer.

Since the landmark information was only incorporated into the loss function and not directly given to the generator, the generator could not learn detailed information about the landmarks. Therefore, as shown in Fig. 3, there is no significant difference between the generated images of the proposed method and WGAN-GP.

However, as shown in Fig. 4, there was a difference in the poor quality group. In images generated by WGAN-GP, the eyes, nose and mouth are blurred and cannot be recognized as a face. On the other hand, in images generated by the proposed method, the eyes, nose, and mouth are not collapsed compared to the WGAN-GP, although the whole image is distorted in some cases. In addition, the shape of the eyes, nose and mouth became clearer and the value of FID became smaller.

Therefore, the proposed method can be considered to have learned to generate images that can recognize landmarks by adding Landmark Loss.

VII. CONCLUSION

In this paper, we proposed a method for generating face images for illustration using WGAN-GP with landmark information, and compared it with WGAN-GP. In order to incorporate landmark information in the generated adversarial network, we added Landmark Loss to the loss function.

In our experiments, we could not generate images that exactly match to the landmarks in the input image. However, with the addition of Landmark Loss, the proposed method could learn the location of landmarks and produce high quality images. Therefore, we confirmed the effectiveness of the proposed method on creation of illustration face images. For future work, it is necessary to consider how to provide landmark information directly to the generator in order to generate images with a higher degree of freedom.

REFERENCES

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Networks, ” Neural Information Processing Systems (NIPS), Dec 2014.

[2] R. Wu, X. Gu, X. Tao, X. Shen, Y. W. Tai, and J. Jia, “Landmark Assisted CycleGAN for Cartoon Face Generation,” The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul 2019.

[3] J. Y. Zhu, T. Park, P . Isola, and A. A. Efros, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, ” IEEE International Conference on Computer Vision (ICCV), 2017.

[4] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. Courville, “Improved Training of Wasserstein GANs,” arXiv preprint arXiv: 1704.00028, 2017.

[5] M. Arjovsky, S. Chintala and L. Bottou, “Wasserstein GAN, ” arXiv preprint arXiv:1701.07875, 2017.

[6] Koki Saito, Deep Learning from scratch: theory and implementation of deep learning in Python, O'Reilly Japan, 2016, p.205. (Japanese)

[7] M. Mozafari, “Bitmoji Faces,”kaggle, 3 Jan. 2021.

[8] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium,” arXiv preprint arXiv:1706.08500v6, 2018.