Face Image Generation with Features of Specific Group from Small Dataset*

Yuichi Kato[12], Takuya Shuto[3], Masahiko Mikawa[1] and Makoto Fujisawa[1]

***Abstract*— The research on face image generation has gained widespread attention in the field of image generation. However, large datasets are required to train a generative model to produce images of high quality and resolution. It is especially difficult to collect multiple face images of a specific group. It may also be difficult to establish an efficient model by training with a small dataset.**

This paper presents a novel method for the generation of face images with the features of a specific group. The face images of a dataset are embedded into a latent space as sets of latent variables by Image2StyleGAN and are expressed as a distribution of the sets of latent variables in the latent space of a pre-trained StyleGAN2. Principal Component Analysis (PCA) is used to extract the features of the distribution. The generation of images for different groups does not require re-training of the model, as a pre-trained model is used instead. Furthermore, the proposed method can generate high-quality images with the features of the group from a dataset with only about 100 face images. However, the quality and the variety of the generated images can vary depending on a Cumulative Contribution Rate (CCR) of the PCA. Therefore, this study also proposes a metric called the Fréchet Inception Distance in Principal Component Space (FID-PCS), which can evaluate the generated images even with a small dataset. The FID-PCS can be used to determine the CCR that generates images with a good balance between the quality and the variety. The face images of three groups were collected as datasets to evaluate the validity of the proposed method, which include male idols, female idols, and male mixed martial artists. It was observed that images with the features of the group are generated by the face distributions extracted by the PCA, and the images with high quality and wide variety are generated by determining the appropriate CCR by the FID-PCS.

I. INTRODUCTION

Generative Adversarial Networks (GANs) [1]–[7] can generate images with high quality and resolution. However, a large dataset and large-scale calculation are required to train these models. Especially, when we want to generate face images of people belonging to a particular group, it is a hard work to collect many face images as a dataset. For the number of input images, Karras et al. [8] successfully generated high-quality images by training the model with only 1000 input images. However, this method involved a large amount of computation and required multiple high-performance GPUs. Noguchi et al. [9] and Zhao et al. [10] proposed a method of generating images from at least 25 input images using the transfer learning process. However, some of the generated images were blurred or unnatural. Other studies [11], [12] proposed an semantic editing method to manipulate attributes in latent space. In particular, some studies [13]–[15] proposed methods that do not require any training by using pre-trained models. Voynov et al. [13] achieved an unsupervised semantic interpretation of the latent space without training model. GANSpace [14] successfully supports the semantic interpretation of latent spaces by performing dimensionality reduction by principal component analysis (PCA) using pre-trained model. However, when a user wants to generate or edit images that do not belong to a general face dataset, such as FFHQ [6] used in GANSpace, but belong to a dataset with the features of a specific group, it is necessary to train the model well with the dataset of the group. In addition, semantic editing is not appropriate for the purpose of generating images with variety.

Therefore, this study proposes a novel image generation method using a pre-trained generative model and a small

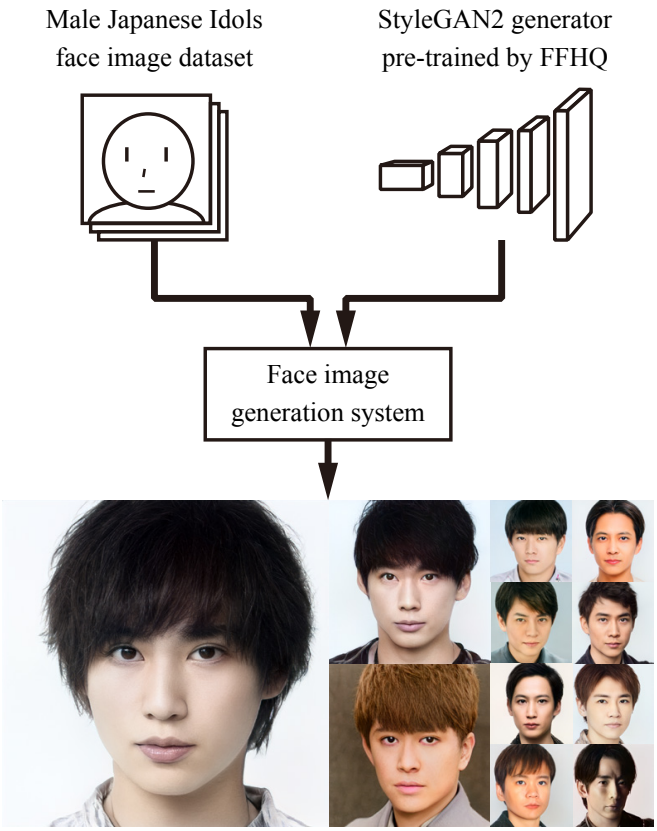


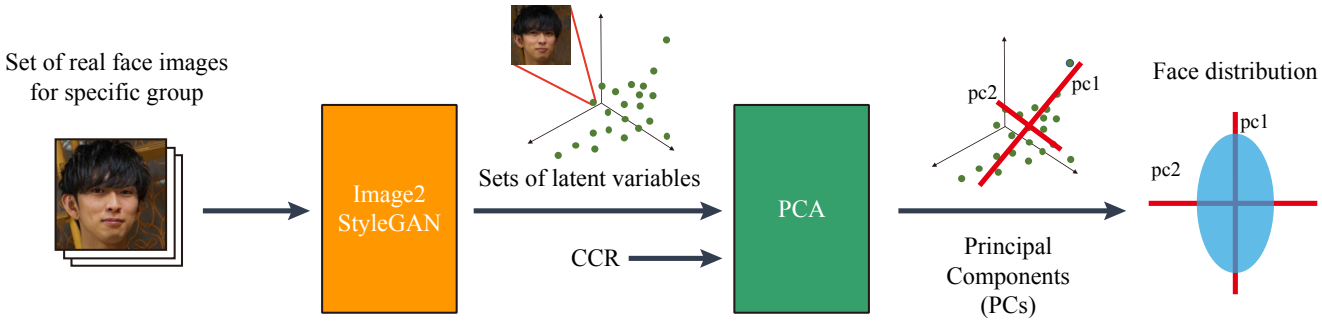
Fig. 1: Face image generation from dataset of male idols as example of specific group

*This work was not supported by any organization

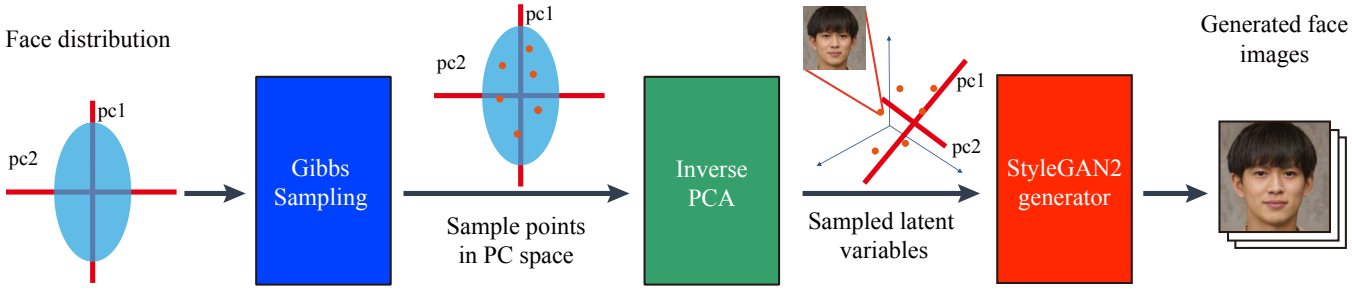
[1]Graduate School of Comprehensive Human Sciences, University of Tsukuba, 1-2 Kasuga, Tsukuba City, Ibaraki Prefecture, 305-0821, Japan

[2] kato.yuichi.sg@alumni.tsukuba.ac.jp

[3]Graduate School of Library, Information and Media Studies, University of Tsukuba, 1-2 Kasuga, Tsukuba City, Ibaraki Prefecture, 305-0821, Japan. This author graduated in March 2021.



(a) Embedding real faces into sets of latent variables and obtaining a face distribution



(b) Sampling from the face distribution and generating face images

Fig. 2: Process of proposed face image generation method

dataset of the face facial images of a specific group. Each image in the dataset is embedded as a latent variable into a latent space of the pre-trained StyleGAN2 [7] model by using Image2StyleGAN [16]. Although the sets of latent variables form a distribution with the features of the specific group, it is difficult to generate a face image with the features of the group directly from the latent space because the distribution is implicitly embedded in the latent space. Therefore, the features included in the distribution are extracted as Principal Components (PCs) by performing the PCA. This enables the generation of high-quality images with the features of the group. Furthermore, the images can be generated from a small dataset without re-training the model owing to the application of the pre-trained model. Conversely, the quality and the variety of the generated images depend on the Cumulative Contribution Rate (CCR) of the PCA, which represents the degree of dimensionality reduction. There exists a trade-off between the quality and the variety, which is similar to the threshold of the truncation trick in BigGAN [5]. Therefore, this study also proposes a novel evaluation metric called the Fréchet Inception Distance in Principal Component Space (FID-PCS) to determine the CCR that can generate a wide variety of high-quality images. The FID-PCS can perform the calculations with a small number of images contrary to the conventional FID that requires a large number of images for evaluation.

Three types of experiments are conducted to evaluate the proposed face image generation system. First, high-quality face images with the characteristics of each of the three groups are generated. Second, the effectiveness of the FID-PCS is evaluated by comparing it with the conventional FID. Finally, the CCR that can generate images with a good balance between the quality and the variety is determined by the FID-PCS.

II. METHOD

This section explains the procedure of the generation of face images from a specific group. Fig. 2 shows the workflow of the proposed method. It consists of two main processes. The first process involves embedding the face images of the specific group into a latent space of StyleGAN2 and obtaining a distribution of the group using the PCA. Face images with the diversity reflecting the features of the group can be generated by adding PCA. This distribution is termed as the face distribution in this paper. The second process involves sampling based on the face distribution and the generation of face images with the features of the specific group.

A. Embedding Face Images into Latent Space and Obtaining Face Distribution

Fig. 2a shows the processes of obtaining the face distribution of the specific group. A set of real face images of the specific group is used as the input. The input face images are embedded into a latent space as sets of latent variables by Image2StyleGAN. Image2StyleGAN minimizes the distance between the image generated from random latent variables and the input image in the pre-learned StyleGAN, which makes it possible to represent the input image as a latent variable. The latent space, which is pre-trained by the StyleGAN2 model with FFHQ [6], is used in this study. The PCA is performed on the sets of latent variables owing to the high dimensionality of the latent space. This

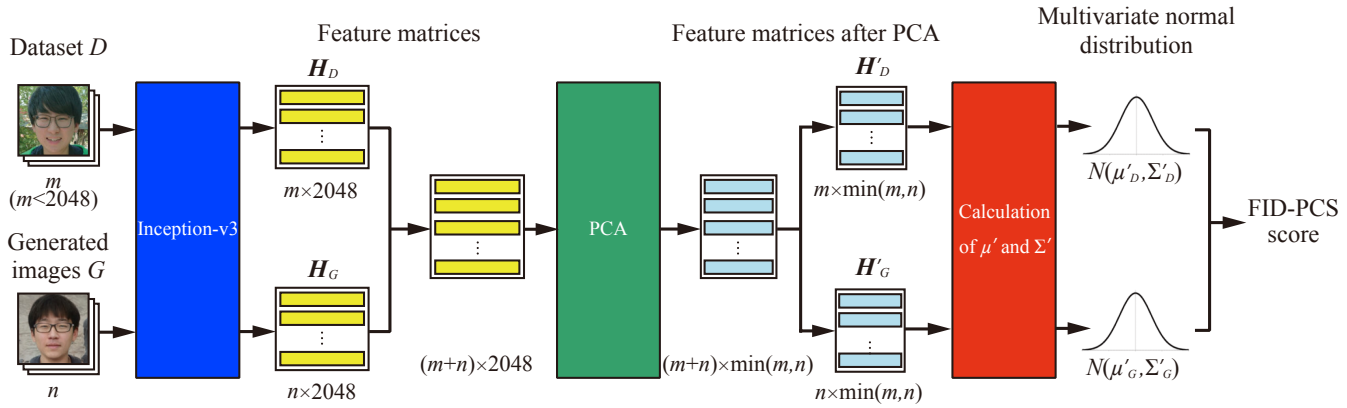


Fig. 3: Calculation of FID-PCS score

method extracts the features of the data distribution and reduces the dimensionality. A CCR is also fed as an input to determine the degree of the dimensionality reduction. The sets of latent variables can be embedded into a PC space of lower dimensions. A multivariate normal distribution is assumed, which is called the face distribution in this study and represents the features of the specific group of the face images in the lower-dimensional space.

B. Face Image Generation with Sampling from Distribution

Fig. 2b shows the process of the face image generation using the face distribution with the features of a specific group. The points in the face distribution obtained from the process in Fig. 2a are randomly sampled using the Gibbs sampling [17] method, which allows for fast sampling in the high-dimensional distributions. Since the sampling points are in the PC space, the inverse PCA transformation is used to project the points into the latent space. Lastly, the face images are generated by feeding the sets of latent variables as inputs into the StyleGAN2 generator.

C. Evaluation Method for Generated Images

This subsection presents the method of determining the CCR to generate the face images with a good balance between the quality and the variety. The Fréchet Inception Distance (FID) [18] is a widely used evaluation metric for deep generative models, and Inception-v3 [19] is used in FID for feature extraction. However, the application of FID is limited because it requires at least 2048 input images. Let the number of input images be k. Since the dimensionality of the feature vector, $\boldsymbol{h}$, extracted by Inception-v3 from one image is 2048, the feature matrix, $\boldsymbol{H}$, consisting of the multiple $\boldsymbol{h}$ is represented by $k \times 2048$. When $k < 2048$, the covariance matrix, Σ, calculated with $\boldsymbol{H}$ may contain some imaginary numbers and NaNs because $\boldsymbol{H}$ is not full-rank; this increases the difficulty of computation. Particularly, the FID is unsuitable for a small dataset such as the one used in this study.

Therefore, this study proposes the FID-PCS as a novel evaluation metric for smaller input images. Fig. 3 shows the process of the FID-PCS. Let the number of images of a dataset, D, be m, and the number of generated images, G, be n, where m is assumed to be less than 2048. The dataset and the generated images are inputted, and $\boldsymbol{H}_D$ and $\boldsymbol{H}_G$ are obtained from Inception-v3. The sizes of $\boldsymbol{H}_D$ and $\boldsymbol{H}_G$ are $m \times 2048$ and $n \times 2048$, respectively. When either the number of the input images in the dataset D or the number of generated images G, is less than 2048, the dimensionality reduction using the PCA on $\boldsymbol{H}_D, \boldsymbol{H}_G$ is performed. $\boldsymbol{H}_D$ and $\boldsymbol{H}_G$ are concatenated into one matrix, and the dimensionality reduction by the PCA is performed on this matrix. Since the size of the feature matrix obtained after the PCA is $(m+n) \times \min(m,n)$, it is divided into two matrices, $m \times \min(m,n)$ and $n \times \min(m,n)$, corresponding to the number of feature vectors. The dimensionally reduced feature matrices, $\boldsymbol{H}'_D$ and $\boldsymbol{H}'_G$, are used to calculate the mean vectors μ'_D, μ'_G and the covariance matrices Σ'_D, Σ'_G. They are then used to calculate the FID-PCS score, d^2, using the following equation:

$$d^2 = ||\mu'_D - \mu'_G||_2^2 + tr(\Sigma'_D + \Sigma'_G - 2(\Sigma'_D\Sigma'_G)^{\frac{1}{2}}). \quad (1)$$

The generated images can be evaluated stable without any imaginary numbers or NaNs by the FID-PCS, even when the number of images in the dataset is small.

III. Experimental Results

Three types of experiments are conducted to generate a wide variety of high-quality face images with the features of a group. First, the face images are generated from the datasets of three groups to confirm the effects of the CCR on the quality and the variety of the generated images. Second, the FID-PCS scores are calculated based on the datasets and the generated images and compared with the FID scores to verify whether the evaluation performance of the FID-PCS is equivalent to that of the FID. Third, the appropriate CCR is determined based on the evaluations of the FID-PCS to generate face images with an appropriate balance between the quality and the variety.

A. System Configuration and Datasets

The face image generation system is built using Google Colaboratory, and the system configuration is shown in Table I.

TABLE I: Development environment

Development environment	Google Colaboratory
CPU	Intel Xeon 2.20 GHz
Memory	10GB
GPU	Tesla P-100 16 GB
Language	Python 3.9.3
Library	Tensorflow-gpu 1.14.0

The facial images of the following three groups were collected from the Internet: 112 images of male mixed martial artists, 103 images of male idols, and 100 images of female idols. All of them were Japanese. The backgrounds of the images of the male mixed martial artists and female idols were uniform, but the backgrounds of the male idols were varied. Additionally, the lighting conditions in the images of the male mixed martial artists and the female idols were almost identical, but they varied in the images of the male idols. The size of all the images was 1024×1024 [pixel]. The collected face images are not presented in this paper due to portrait rights.

B. Face Image Generation with Features of Group

This subsection shows that face images with the features of the groups can be generated and that the CCR affects the quality and the variety of the generated images. The face images were generated from the three groups described in Section III-A and presented along with the variation of the CCR. First, the CCR was set from 0.05 to 1.00 in increments of 0.05 in this experiment. Second, the face distributions for each CCR were generated from each dataset. Finally, sampling was performed for this distribution, and the face images for each CCR were generated.

Fig. 4 shows the face images generated from each group and the value of the CCR. It is observed that the features of each group are represented well in the generated face images. Furthermore, the CCR affects both the quality and the variety of the generated images. The quality of the resultant images is observed to be high when the CCR is small. However, the variety of the generated face images is reduced because of mode collapse, and many similar images are generated. Therefore, the greater the CCR, the wider the variety of the generated images. However, when the CCR is set above a certain value, some of the face images generated were of poor quality. The details are discussed in Subsection III-D, but it is necessary to determine an appropriate value of the CCR to generate images with a good balance between the quality and the variety.

Table II shows the average processing time required to generate images using the proposed method. Approximately 6 h are required to generate 1000 face images from the dataset of 100 images. Embedding an image requires most of the processing time.

C. Validation of FID-PCS

The evaluation metric is essential to accurately evaluate the quality and the variety of the generated images. This subsection shows that the proposed evaluation metric, FID-PCS, performs equivalently to the conventional FID. For simplicity, DCGAN [2] is used to generate the images, and MNIST [20] is used as the dataset to train the DCGAN model.

Fig. 4: Generated face images from datasets of three groups for each CCR

TABLE II: Details of processing time to generate 1000 face images from 100 images of dataset

Process	Processing time
Image2StyleGAN	17457 [s]
PCA	0.20 [s]
Gibbs sampling	2.50 [s]
inverse PCA	0.08 [s]
StyleGAN2 generator	51.30 [s]

The variation of the conventional FID score with an epoch is shown by the light blue line in Fig. 5. For the calculation, 2048 images in the dataset and 2048 images in generated images were used. It is observed that as the epoch progresses, the training accuracy is improved, the image quality and the variety increase, and the FID score consequently decreases. The other lines show the variations of the FID-PCS scores with the epoch. First, 100 randomly selected images from the

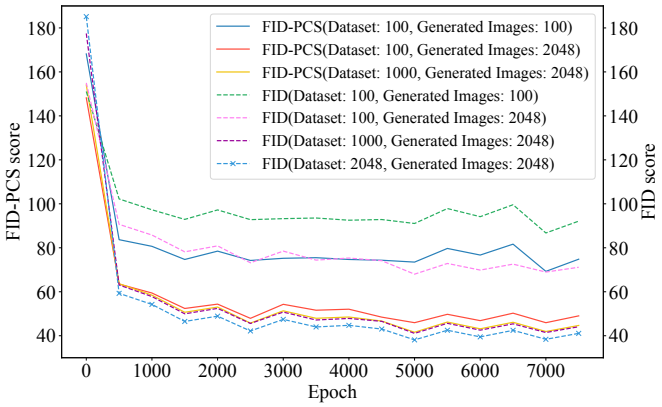
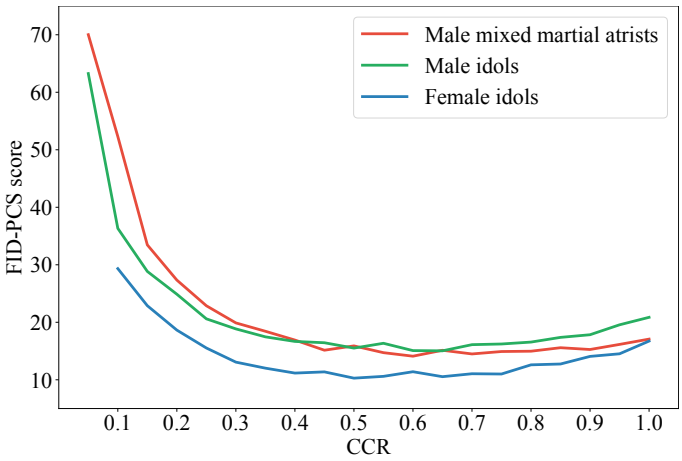


Fig. 5: FID-PCS score variation by varying number of datasets and generated images



(a) Relationship between FID-PCS scores and CCRs

(b) Generated face images in each group when CCR was set to 0.25, 0.45, and 0.65

Fig. 6: Process of proposed face image generation method

dataset and 100 generated images were used to calculate the FID-PCS and the FID scores. The variation of the FID-PCS and the FID scores with the epoch is shown by the blue and the green lines in Fig. 5. The scores and the shapes of these lines are different from those of the light blue line of the FID. Second, 100 randomly selected images from the dataset and 2048 generated images were used to calculate the FID-PCS and the FID scores, which are shown by the red and the pink lines. The shape of the pink line is similar to those of the light blue line of the FID, however the scores are far. In comparison, the scores and the shape of the red line of FID-PCS are similar to those of the light blue line, which indicates that the performance of the FID-PCS is equivalent to that of FID even with a small number of input images. Furthermore, 1000 randomly selected images from the dataset and 2048 generated images were used to calculate the FID-PCS and the FID scores shown by the yellow and the purple lines. It can be stated that the scores and the shape of the yellow line are almost same as those of the purple line, and are more similar to those of the light blue line of the FID. Therefore, the FID-PCS can evaluate the generated images as well as the FID even for a small number of datasets.

D. Determination of best CCR

The CCR affects the quality and the variety of the generated images, as described in Subsection III-B. Therefore, this subsection presents the method of the determination of the appropriate CCR that can generate images with a good balance between the quality and the variety. Since the process of determining the CCR is aimed at preserving both the quality and the variety, the images generated from the CCR may be inferior in quality. However, we tolerate this in consideration of the contribution to the variety. The datasets of the three groups described in Subsection III-B were used.

Fig. 6 (a) shows the relationship between the FID-PCS scores and the CCRs. The variation of the FID-PCS scores with the CCR has some common features among the three groups. The FID-PCS score decreases when the CCR changes from 0.05 to 0.45. No significant change is observed when the CCR is between 0.45 and 0.80, and the score

increases slightly when the CCR is 0.80 or higher. Fig. 6 (b) presents the face images of the three groups generated when the CCR is set to 0.25, 0.45, and 0.65. When the CCR is 0.25, the generated face images do not show significant variety for any of the groups. Conversely, when the CCR is 0.65, although a wide variety of face images is generated, the quality of some of the images is poor. When the CCR is 0.45, the generated images show a good balance between the quality and the variety, although this is the subjective opinion of the authors. Furthermore, this value is common to all three groups. Therefore, it can be stated that the CCR with the best balance between the quality and the variety of the generated images is the one where the FID-PCS score decreases, and the curve begins to flatten. Moreover, since the best CCR is common to all three groups, the value can be determined systematically. Therefore, it is assumed that the images with a good balance between the quality and the variety are generated by setting the CCR to 0.45.

On the other hand, the FID-PCS also faces some limitations. First, when the CCR is high, although the percentage of images with poor quality is high, the FID-PCS score does not increase significantly. This implies that the FID-PCS cannot evaluate the quality of images accurately at the high CCR. Second, the FID-PCS scores and shapes of the three groups are similar in this experiment. However, since the number of groups used for validation is small, it is necessary to verify whether the same results can be obtained in more groups as those used in this experiment.

IV. Conclusion

This study proposes a method to generate face images from a small dataset based on the pre-trained StyleGAN2 model. The input images are embedded into the latent space and transformed into sets of latent variables by Image2StyleGAN. Since these sets of latent variables have high dimensionality, a PCA is performed to reduce their dimensionality. Additionally, a multivariate normal distribution is assumed from the reduced sets of latent variables, and Gibbs sampling to the distribution is performed. The face images were generated by feeding the sample points as inputs into the StyleGAN2 generator.

In the experimental analysis, the face image datasets of male mixed martial artists, male idols, and female idols were created. Face images with the features of each group were generated from each dataset. It was observed that the CCR of the PCA affects the quality and the variety of the generated images. Therefore, this study also proposed the FID-PCS to evaluate the images generated from the proposed method. The evaluation performance of the FID-PCS was equivalent to that of the FID even when fewer images are inputted in the experiment. Furthermore, the appropriate CCR to generate images with a good balance between the quality and the variety was determined by the FID-PCS.

The proposed method faces certain limitations. The FID-PCS cannot accurately evaluate the quality of the generated images when the CCR is high. Additionally, it is necessary to generate images from more groups and examine the correlation between the FID-PCS and the CCR. In future work, we will extend the application of the proposed generation method, beyond human faces, to groups in different domains such as cars, cats, and horses.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, pages 2672–2680. 2014.

[2] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[3] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *ICLR*, 2018.

[4] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. In *ICML*, pages 7354–7363, 2019.

[5] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *ICLR*, 2019.

[6] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, pages 4401–4410, 2019.

[7] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *CVPR*, pages 8110–8119, 2020.

[8] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. In *NeurIPS*, volume 33, pages 12104–12114, 2020.

[9] Atsuhiro Noguchi and Tatsuya Harada. Image generation from small datasets via batch statistics adaptation. In *ICCV*, pages 2750–2758, 2019.

[10] Miaoyun Zhao, Yulai Cong, and Lawrence Carin. On leveraging pretrained GANs for generation with limited data. In *ICML*, volume 119, pages 11340–11351, 2020.

[11] Paul Upchurch, Jacob Gardner, Geoff Pleiss, Robert Pless, Noah Snavely, Kavita Bala, and Kilian Weinberger. Deep Feature Interpolation for image content changes. In *CVPR*, 2017.

[12] Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the latent space of gans for semantic face editing. In *CVPR*, pages 9243–9252, 2020.

[13] Andrey Voynov and Artem Babenko. Unsupervised discovery of interpretable directions in the gan latent space. In *ICML*, pages 9786–9796, 2020.

[14] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. GANSpace: Discovering interpretable gan controls. In *NeurIPS*, volume 33, pages 12104–12114, 2020.

[15] Hui-Po Wang, Ning Yu, and Mario Fritz. Hijack-gan: Unintended-use of pretrained, black-box gans. In *CVPR*, pages 7872–7881, 2021.

[16] Rameen Abdal, Yipeng Qin, and Peter Wonka. Image2stylegan: How to embed images into the stylegan latent space? In *ICCV*, pages 4432–4441, 2019.

[17] Stuart Geman and Donald Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *PAMI*, 6(6):721––741, 1984.

[18] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NeurIPS*, pages 6629–6640, 2017.

[19] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *CVPR*, pages 2818–2826, 2016.

[20] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *IEEE*, 86(11):2278–2324, 1998.