2D Virtual Youtuber Character Generation Using Generative Adversarial Networks

Natdanai Thedwichienchai[a], Thitirat Siriborvornratanakul*[a]
[a] Graduate School of Applied Statistics, National Institute of Development Administration,
148 SeriThai Rd., Bangkapi, Bangkok, 10240, Thailand

ABSTRACT

Virtual YouTubers (VTubers) offer significant growth potential as a new type of content creator. However, the financial aspect poses a hurdle for aspiring VTubers. This article proposes a cost-effective solution by utilizing generative models to create full-body images of 2D VTuber characters. Notably, studies have achieved remarkable results using Generative Adversarial Networks (GANs), including Deep Convolutional GAN (DCGAN) and Style-based GAN 2 (StyleGAN2), for anime face generation. To address the lack of image synthesis systems for full-body anime characters, experiments were conducted with DCGAN and StyleGAN2 on the Danbooru dataset. The results demonstrate that StyleGAN2 models outperform DCGAN, yielding superior Fréchet Inception Distance (FID) scores of 25.06, 28.03, and 24.52, compared to DCGAN's 159.21 FID score. This research contributes to reducing the cost associated with becoming a VTuber and offers insights into generating 2D full-body anime characters for VTubers.

Keywords: Deep learning, generative model, generative adversarial network, virtual youtuber, anime

1. INTRODUCTION

The surge in popularity of online content on video platforms like YouTube, Twitch, and Mixer can be attributed to the rise of amateur content creators. These creators can generate income through partnerships with platforms, advertisements, and viewer subscriptions [12]. With a shift from platform-focused to individual-focused media consumption, content creators have emerged as "micro-celebrities," leveraging social media platforms to build communities and increase their earnings [13]. By 2020, YouTube Gaming accumulated 6.19 billion hours of views, Facebook Gaming had 3.1 billion hours, and Twitch reached 18.41 billion hours [3]. These figures highlight the significant growth potential of the digital platform market, enabling content creators to monetize their creations. Among various content creation formats, streaming stands out as a popular choice, allowing creators to interact with viewers in real-time using visual, audio, and text-based communication tools. Streamers also benefit from virtual gifts purchased by their audience, distinguishing them from other content creators like video bloggers (vloggers) or YouTubers [12].

Virtual YouTubers (VTubers) are content creators who utilize virtual avatars, either in 2D or 3D, which mimic their movements and speech in real-time during streaming sessions. VTubers often adopt avatars to maintain privacy or adhere to corporate policies. Many VTubers have expanded their virtual presence to various channels, including anime, manga, video games, novels, and even traditional radio broadcasting, blurring the boundaries between digital and real-world experiences. VTubers' popularity in East Asia can be attributed to their characteristics, reminiscent of anime, manga, and video games, while also offering unique and unconventional content. For instance, VTubers can engage in avatar swapping during co-streaming sessions, providing a novel viewing experience [23].

However, one significant hurdle to becoming a VTuber is the associated cost, ranging from $500 to $2000 for equipment [7]. An avatar is necessary to conceal one's true identity, with 2D avatar models typically costing $200 to $2000 and 3D avatar models ranging from $3000 to $5000, depending on the artist. Although novice or amateur illustrators may offer lower prices, the quality of their work often aligns with the cost. Custom-sized models can be created by artists at a cost of around $200 to $300, while pre-built avatar models are available on websites like nizima (https://nizima.com/) or r/VirtualTubers on Reddit (https://www.reddit.com/r/VirtualYoutubers/), with varying costs [19].

Recent advancements in machine learning, particularly in Generative Adversarial Networks (GANs) [8], have enabled the generation of new images. Leveraging the power of GANs, we can facilitate growth and opportunities within VTuber communities. For example, GANs can help create virtual avatars, reducing the cost of model design for aspiring VTubers. Additionally, GANs can inspire designers with fresh avatar concepts. Instead of a conventional unconditional

Fifth International Conference on Computer Vision and Computational Intelligence (CVCI 2024),
edited by Chin-Chen Chang, Chutisant Kerdvibulvech , Proc. of SPIE Vol. 13169,
131690A · © 2024 SPIE · 0277-786X · doi: 10.1117/12.3023881

GAN, using a conditional GAN allows more control over image synthesis and can be used for dataset generation [24[. In this research, our aim is to utilize GANs, specifically StyleGAN2, to generate a 2D avatar model for VTubers. The contribution of this study lies in the development of StyleGAN2-based models for synthesizing full-body VTuber character images, which, to the best of our knowledge, has not been previously explored in the literature.

2. RELATED WORKS

GANs [8] are deep learning models composed of two sub-models: the discriminator and the generator. The generator learns to create images from random noise, while the discriminator learns to differentiate between generated (fake) and real images. Radford et al. [1] introduced a variant of GAN called Deep Convolutional GAN (DCGAN) that stabilizes training and enables the generation of higher-resolution images across diverse datasets. Jin et al. [20] utilized DCGAN to generate facial images of anime characters using training images from the Japanese game-selling website Genshu. Although the results were satisfactory, further improvements are needed, especially in terms of dataset distribution and final image resolution.

Another noteworthy GAN-based image generation architecture is Style-based GAN (StyleGAN) [17], which maps latent code to the intermediate latent space, promoting localized styles. Unlike traditional GAN architectures, StyleGAN excels in quality metrics like Fréchet Inception Distances (FID) scores. However, the original StyleGAN exhibited blob-like artifacts due to adaptive instance normalization (AdaIN) in the generator network. This led to the development of StyleGAN2 [18], featuring an alternative normalization design called weight modulation. StyleGAN2 also introduced regularization terms that improve computation cost and memory usage. It maintains progressive growth, with the generator focusing on low-resolution features before refining finer details. StyleGAN2 has been fine-tuned for generating cartoon faces from datasets acquired from Never webtoon and Disney's website [9]. Although cartoon faces have been successfully generated, the entire body of cartoon characters remains unexplored.

While GANs have shown potential in generating both photorealistic and anime face images, anime images possess unique characteristics, such as clear and sharp lines, unlike the textured nature of photorealistic images that GAN excels at [5]. Most existing works on anime generation focus solely on anime faces. Zhuang and Yang [14] employ a few-shot technique for image-to-image translation in generating cartoon faces. Men et al. [21] utilize an unpaired image synthesis method to transform human portraits into anime faces. Tang et al. [6] explore unpaired image-to-image translation using attention GAN. To our knowledge, few studies have addressed the generation of full-body anime characters. Hamada et al. [10] utilize Progressive Structure-conditional Generative Adversarial Networks but require complex anime datasets with precise 2D keypoints extracted from the DeepFashion dataset. Liu's work [22] employs StyleGAN to synthesize full-body anime characters, but the dataset only includes standing characters without background noise and necessitates a large collection of 12,000 high-quality images. Hence, our research aims to contribute an approach that can generate full-body anime character images using smaller anime datasets with background noise. Table 1 presents a comparison between our work and previous studies in full-body anime character generation.

Table 1. Comparison between our work and related works regarding full-body anime character synthesis.

Work	Method	Model	Dataset
Hamada's work [10]	GANs with structural conditions	Progressive Structure-conditional GAN	- Anime dataset with Exact Pose Keypoints - DeepFashion dataset to extract keypoints
Liu's work [22]	Image synthesis	StyleGAN	- 12,000 images of the full-body anime character with no background noise
Our work	Image synthesis	StyleGAN2	- 3,000 images mixing between full-body images and upper-half-body images of the anime character despite any background

3. PROPOSED METHODS

This paper presents a system for generating full-body anime images for VTubers, utilizing two models: DCGAN (Section 3.1) and StyleGAN2 (Section 3.2). The choice of these models stems from their respective merits in generating anime faces, as demonstrated by Jin et al. [20] for DCGAN and Back [9] for StyleGAN2. Moreover, both GAN models offer pre-trained weights specifically tailored for anime faces.

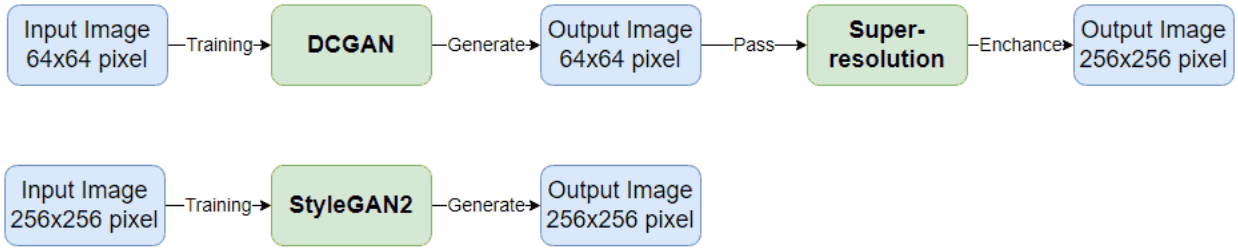

Figure 1. The training process of our DCGAN-based (top) and StyleGAN2-based (bottom) models.

3.1 DCGAN

The initial approach involves the utilization of DCGAN as introduced by Jin et al. [20] for generating anime faces. In our implementation, we employ the pre-trained weight known as the "Anime face weight," which can be accessed at https://github.com/yashyenugu/Anime-Face-GAN. However, since the pre-trained weights and architecture were optimized for output sizes of 64x64 pixels, which are relatively small, we employ the super-resolution model developed by Intaniyom et al. [16] to upscale the generated images from 64x64 to 256x256 pixels. It is important to note that the super-resolution model has already been trained for enhancing anime images, and the corresponding weights have been made available. Therefore, no additional training is required for the super-resolution component. An overview of the training process for the initial approach is illustrated in Figure 1 (top).

3.2 StyleGAN2

StyleGAN2 is a style-based generator comprised of a non-linear mapping network and the synthesis network, denoted as g. The non-linear mapping network takes latent vectors as input and maps them to the intermediate latent space, employing weight demodulation to regulate the generator for each convolutional layer in the synthesis network g. Each resolution is associated with 8 layers of the mapping network f and 18 layers of the synthesis network g. The output from the final layer is then converted into RGB using a separate 1x1 convolution. StyleGAN2 provides three pre-trained weights: ffgh256, Disney, and NeverWebtoon. The ffhg256 weight is trained for human face generation, Disney weight for generating Disney-style and 3D animation faces, and NeverWebtoon weight for Webtoon-style anime faces. However, these weights were originally developed solely for face generation, and their applicability to full-body anime character generation remains unexplored. Consequently, our investigation focuses on evaluating the performance of these three pre-trained weights when trained with a dataset of full-body anime images.

The training process for the second approach, based on StyleGAN2, is depicted in Figure 1 (bottom). Since StyleGAN2 is capable of generating images with a resolution of 256x256 pixels, which provides sufficient detail, there is no longer a need for the super-resolution model. After completing the training, only StyleGAN2 is utilized to generate the desired output images.

3.3 Hyperparameter setting

The experiments conducted in this study utilized Google Colab Pro with a Tesla P100 GPU. The models, incorporating pre-trained weights, were fine-tuned with specific hyperparameters: 1,000 epochs, a batch size of 16, an initial learning rate of 0.002, and output images of resolution 256x256 pixels.

3.4 Dataset

In this study, the Danbooru 2021 dataset (depicted in Figure 2) was employed due to its significant variation between images and the presence of noise. Unlike other anime image datasets that predominantly focus on anime faces, the Danbooru dataset consists of two directories. The first directory contains anime face images, primarily utilized for anime face generation. The second directory encompasses a diverse range of images, including anime landscapes, manga, full-

body anime characters, and anime scenes intended for various purposes. Considering the objective of this research, which involves creating 2D full-body characters for VTubers, the second directory was chosen. From this directory, images featuring a single character with minimal background interference were manually selected. Each selected image encompasses the entirety of the character's body or, at the very least, the head-to-thigh region. For this paper, only images depicting characters in a standing posture were included. Examples of the selected images are illustrated in Figure 3. Ultimately, a total of 4,000 images depicting full-body anime characters were chosen for this study. Among these images, 3,000 were allocated for training, while 1,000 were reserved for testing purposes.

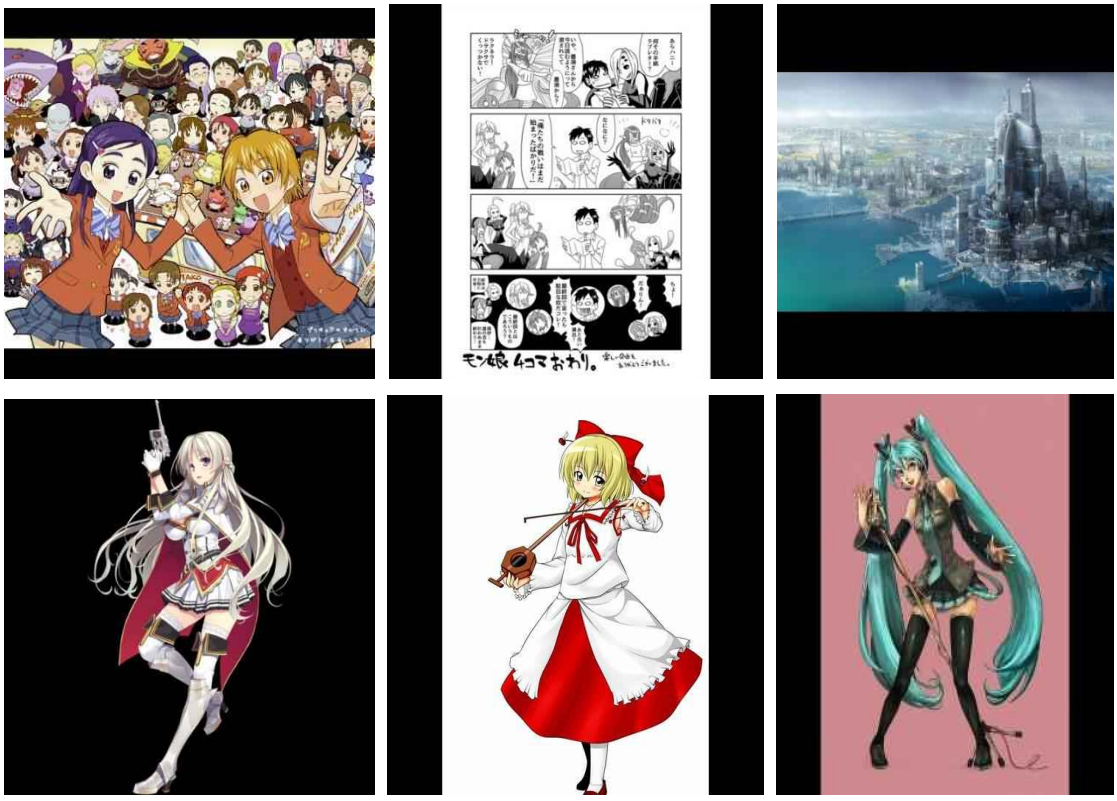

Figure 2. Example images from the original Danbooru dataset from https://www.gwern.net/Danbooru2020.

Figure 3. Example images that are selected for training our models. These images are from the Danbooru dataset in https://www.gwern.net/Danbooru2020.

4. RESULTS

To assess the quality of images generated by our trained models, we employ the Fréchet Inception Distance (FID) [11], an improved variant of the inception score. Our generated images are inputted to an inception model trained on ImageNet. Images containing meaningful objects exhibit lower entropy. The distinction between the standard inception score and the FID score lies in the utilization of a feature extractor layer in FID scores. This layer extracts vision-

relevant features, generating a maximum distribution through mean and covariance known as Gaussian. Subsequently, the model compares the difference between the Gaussian distributions of fake and real images using either Fréchet distance or Wasserstein-2 distance. A lower FID score indicates superior performance.

In this study, we compare the results obtained from 1,000 generated images and 1,000 images from our test dataset. We extract vision-relevant feature vectors from the generated images and calculate the Gaussian distribution with the test dataset images, as well as vice versa. Subsequently, we compare the Gaussian distributions using Fréchet distance. The Python-based PyTorch-fid package (https://github.com/mseitzer/pytorch-fid) is utilized for calculating FID scores from both real and generated samples, along with other metrics regarding GPU requirements for training and time per epoch.

The results presented in Table 2 demonstrate that DCGAN, enhanced with the super-resolution model, yields the highest FID score (the least desirable) of 159.21, whereas StyleGAN2 with the NeverWebtoon pre-trained weights achieves the lowest FID score (the most favorable) of 24.52. Concerning training speed, all models were trained using the same Tesla P100 GPU. The outcomes reveal that DCGAN exhibits the fastest training speed, requiring 43 seconds per epoch, as DCGAN possesses a less complex model compared to StyleGAN2. Among the StyleGAN2-based models, the one utilizing Disney pre-trained weights demonstrates the fastest training speed of 228 seconds per epoch, whereas the slowest model is StyleGAN2 with NeverWebtoon pre-trained weights, taking 266 seconds per epoch. As for interface speed, we measured the time each model required to generate 1,000 images using the Tesla T4 GPU provided by Google Colab Pro. DCGAN (without super-resolution enhancement) emerges as the fastest model, taking 20 seconds on GPU and 64 seconds on CPU. As for StyleGAN2-based models running on GPU, the model with ffhq256 pre-trained weights is the fastest using 58 seconds, followed by NeverWebtoon pre-trained weights with 59 seconds, and Disney pre-trained weights with 67 seconds.

Table 2. Our experimental results of models whose pre-trained weights are specified inside parentheses.

Model	FID Score ↓	GPU required	Training speed	Interface speed on CPU	Interface speed on GPU
DCGAN (Anime face weights) + Super-resolution	159.21	No	43 seconds per epoch	64 seconds	20 seconds
StyleGAN2 (ffhq256 weights)	25.06	Yes	247 seconds per epoch	N/A	58 seconds
StyleGAN2 (Disney weights)	28.03	Yes	228 seconds per epoch	N/A	67 seconds
StyleGAN2 (NeverWebtoon weights)	24.52	Yes	266 seconds per epoch	N/A	59 seconds

Based on the findings presented in Table 2, the StyleGAN2 model with NeverWebtoon pre-trained weights exhibits the most favorable overall performance. This conclusion is drawn from its achievement of the lowest FID score. While the StyleGAN2 model with NeverWebtoon pre-trained weights shows slower training speed, its interface speed on GPU is comparable to the fastest StyleGAN2 model with ffhq256 pre-trained weights. On the other hand, the StyleGAN2 model with Disney pre-trained weights may offer faster speed compared to the StyleGAN2 model with NeverWebtoon pre-trained weights, but its FID scores and GPU inference speed are not among the best. Similar observations apply to the DCGAN model, which boasts the fastest training speed but is not selected due to its unfavorable FID score.

Regarding the quality of the generated images, our DCGAN model fails to produce satisfactory results, as depicted in Figure 4. This outcome can be attributed to the low resolution (64x64 pixels) of the original output images generated by the model with pre-trained weights, leading to a noticeable decrease in image quality. The implementation of the super-resolution model allows for the enlargement of the output images from 64x64 pixels to 256x256 pixels, as illustrated in Figure 5. Although the character's heads are noticeable, the overall image generated remains distorted, and even the utilization of the super-resolution module fails to rectify the issue.

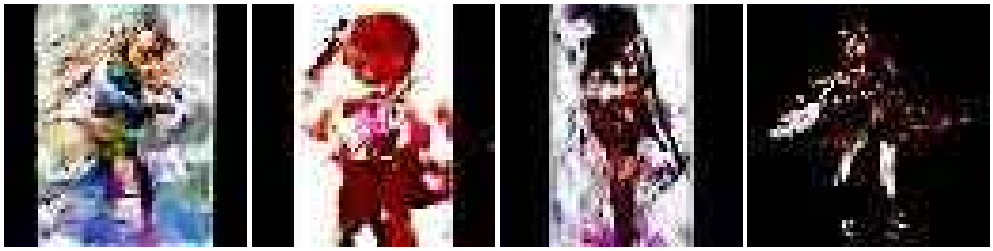

Figure 4. Some examples of generated images (64×64 pixels) from our DCGAN.

Figure 5. Some examples of generated images (256×256 pixels) from our DCGAN after being enhanced with the super-resolution module.

Figures 6-8 showcase the generated images from our StyleGAN2 models, utilizing the ffhq256, Disney, and NeverWebtoon pre-trained weights. The top rows of the figures display satisfactory results, while the bottom rows exhibit less desirable outcomes. When compared to the DCGAN model, the overall quality of the images produced by the StyleGAN2-based models is superior. The images generated with the ffhq256 pre-trained weights (Figure 6, top row) present well-defined characters in both full-body and half-body forms. However, the model fails to generate the character's arms. Conversely, the images with less favorable outcomes (Figure 6, bottom row) exhibit unclear body parts, despite the legs and head being discernible. In Figure 7, the images produced with the Disney pre-trained weights (top row) successfully depict characters in both full-body and half-body forms, but like the ffhq256 model, the arms are absent. Comparatively, the unsatisfactory results from the Disney model (bottom row) are even worse than those from the ffhq256 model, as no body part can be reliably identified. Finally, in Figure 8, the images generated using the NeverWebtoon pre-trained weights (top row) portray characters in both full-body and half-body forms, with improved detail in terms of cosmetics compared to the ffhq256 and Disney pre-trained weights. However, similar to the previous models, the arms of the characters are not generated. The unsatisfactory results (bottom row) include images where the legs and head can still be identified, albeit with excessive distortion that makes it difficult to specify other body parts.

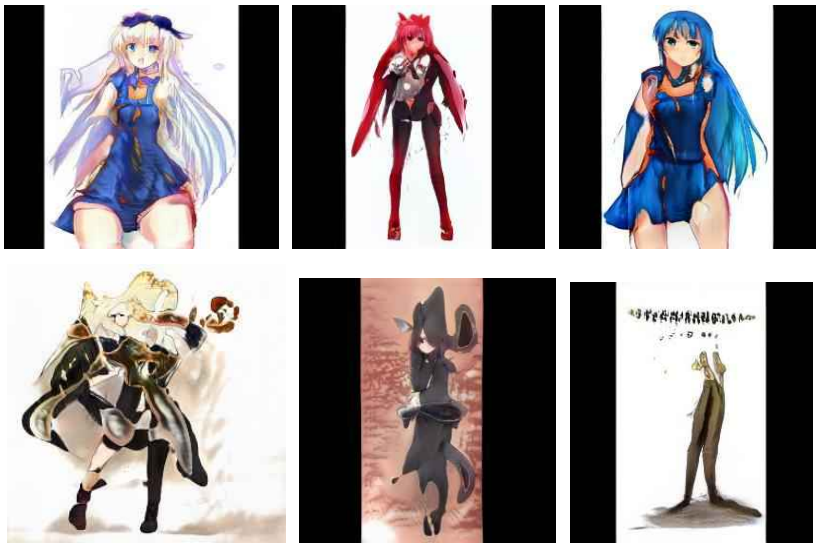

Figure 6. Some examples of 256×256 images generated from our StyleGAN2 model using the ffhq256 pre-trained weights.

Figure 7. Some examples of 256×256 images generated from our StyleGAN2 model using the Disney pre-trained weights.

Figure 8. Some examples of 256×256 images generated from our StyleGAN2 model using the NeverWebtoon pre-trained weights.

Table 3 presents a comparative analysis of the FID scores between our models and the FID scores obtained from three previous studies: Jin et al.'s research [20] on anime face generation using DCGAN, BroadGAN for anime face generation [15], and StyleGAN for full-body anime character generation [22]. The results indicate that our StyleGAN2 models outperform the anime face models generated in previous works [15,20] in terms of FID scores, but fall short when compared to the full-body anime character generation [22]. However, it is important to consider the nature of anime face generation, where the generated images typically consist of full-screen faces with intricate facial details that contribute to the FID computation. In contrast, our full-body character generation includes the anime character in a standing posture, occupying less image space and incorporating non-relevant background areas into the FID computation. This discrepancy may account for the lower FID scores observed in our full-body anime generation compared to the face-only anime generation. The same rationale applies to the results of our full-body character generation, as our current model can generate both full-body and upper-half body images with varying non-related backgrounds. If our StyleGAN were specifically trained to exclusively generate full-body images, it would likely yield results with lower (better) FID scores.

Table 3. The comparative results of our models with different pre-trained weights compared to previous works on anime face generation [15,20] and the full-body anime character generation [22].

Model	FID Score ↓
DCGAN (2017) (Jin et al., 2017) [20]	122.96
BroadGAN (2022) (Jin et al., 2022) [15]	255.06
StyleGAN (2020) (Liu, 2020) [22]	5.02
DCGAN (Anime face weights) + Super-resolution (ours)	159.21

Based on the conducted experiments, it is evident that generating the full-body 2D VTuber character poses a challenge due to its intricate details. To overcome this, a larger dataset comprising full-body characters is essential to facilitate more effective training. Presently, our models are unable to generate arms, primarily because the character images in the Danbooru dataset exhibit various posting styles. Generating arms proves to be more complex compared to legs, as arms can be positioned in diverse locations, whereas legs consistently appear in the lower part of the body. Additionally, it is important to note that our current results are confined to female characters, as the Danbooru dataset exclusively comprises such characters. To incorporate male characters, additional datasets featuring male characters need to be acquired. Moreover, to enhance the performance of DCGAN, exploring alternative pre-trained weights with higher output resolution holds promise in improving the quality of the generated images.

5. CONCLUSION

This study investigates the efficiency and effectiveness of adapting GAN models, originally trained for generating face images, to the task of generating full-body anime characters. The evaluation reveals that StyleGAN2, utilizing the NeverWebtoon weight, achieves the lowest FID score, whereas DCGAN exhibits the highest FID score. In terms of time per epoch, DCGAN demonstrates a shorter duration, while StyleGAN2 with the NeverWebtoon weight requires more time per epoch. However, despite DCGAN's favorable training efficiency, it falls short in generating suitable VTuber characters. Consequently, employing pre-trained weights with higher resolutions becomes crucial to enhance the quality of the generated images.

Given the effectiveness of StyleGAN2 as a prominent representative of GANs, future research can explore alternative generative models, such as Diffusion Probabilistic Models, for VTuber character generation. Additionally, the investigation of character creation from sketches or the adoption of a text-to-image approach holds the potential for exerting greater control over the generated characters. These avenues offer promising prospects for advancing the field, enabling more diverse and customizable approaches to VTuber character generation.

REFERENCES

[1] A. Radford, L. Metz, and S. Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” International Conference on Learning Representations 2016, San Juan, Puerto Rico, May. 2-4, 2016.

[2] D. P. Jaiswal, S. Kumar, and Y. Badr, “Towards an Artificial Intelligence Aided Design Approach: Application to Anime Faces with Generative Adversarial Networks,” Procedia Computer Science, vol 168, Amsterdam,Netherlands: Elsevier, 2020, pp. 57-64

[3] E. May, “Streamlabs and Stream Hatchet Q4 Live Streaming Industry Report” streamlabs.com. https://blog.streamlabs.com/streamlabs-and-stream-hatchet-q4-live-streaming-industry-report-a898c98e73f1 (accessed Jan. 16, 2022).

[4] G. Branwen, “Danbooru2021: A Large-Scale Crowdsourced and Tagged Anime Illustration Dataset” gwern.net. https://www.gwern.net/Danbooru2021 (accessed Feb. 19, 2022).

[5] G. Branwen, “Making Anime Faces With StyleGAN” gwern.net. https://www.gwern.net/Faces (accessed Jan. 16, 2022).

[6] H. Tang, H. Liu, D. Xu, P. H. S. Torr, and N. Sebe, "AttentionGAN: Unpaired Image-to-Image Translation Using Attention-Guided Generative Adversarial Networks," IEEE Transactions on Neural Networks and Learning Systems, New Jersey, USA: IEEE, 2021
[7] H. Tariq, "How to Become a Successful Faceless Virtual Star. Entrepreneur" entrepreneur.com. https://www.entrepreneur.com/article/375553 (accessed Jan. 16, 2022).
[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," Advances in neural information processing systems 27. New York, NY, USA: Curran Associates, Inc, 2014. [Online]. Available: https://papers.nips.cc/
[9] J. Back, "Fine-Tuning StyleGAN2 For Cartoon Face Generation," 2021, arXiv:2106.12445.
[10] K. Hamada, K. Tachibana, T. Li, H. Honda, and Y. Uchida, "Full-Body High-Resolution Anime Generation with Progressive Structure-Conditional Generative Adversarial Networks," Computer Vision – ECCV 2018 Workshops, Munich, Germany, Sep., 8-14, 2018
[11] M. Heusel, H. Ramsauer, T. Unterthiner, and B. Nessler, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," Advances in Neural Information Processing Systems 30. New York, NY, USA: Curran Associates, Inc, 2017. [Online]. Available: https://papers.nips.cc/
[12] M. Törhönen, J. Giertz, W.H. Weiger, and J. Hamari, "Streamers: the new wave of digital entrepreneurship? Extant corpus and research agenda," Electronic Commerce Research and Applications, vol 46, Amsterdam,Netherlands: Elsevier, 2021
[13] M. Sjöblom, M. Törhönen, J. Hamari, and J. Macey, "The ingredients of Twitch streaming: Affordances of game streams," Computers in Human Behavior, vol 92, Amsterdam,Netherlands: Elsevier, 2019, pp. 20-28
[14] N. Zhuang, and C. Yang, "Few-Shot Knowledge Transfer for Fine-Grained Cartoon Face Generation," IEEE International Conference on Multimedia and Expo (ICME). [Online]. Available: https://ieeexplore.ieee.org/document/9428473
[15] Q Jin, R. Lin, and F. Yang, "BroadGAN: Generative adversarial networks of discriminating separate features based on broad learning," Engineering Applications of Artificial Intelligence, vol 109. United Kingdom: Elsevier, 2022.
[16] T. Intaniyom, W. Thananporn, and K. Woraratpanya, "Enhancement of Anime Imaging Enlargement Using Modified Super-Resolution CNN," International Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, Thailand, Oct. 14-15, 2021.
[17] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence (CVPR), CA, USA, 2019, pp. 4396-4405
[18] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtien, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), WA, USA, 2020, pp. 8107-8116
[19] Tony, "How Much Does a VTuber Model Cost?" streamscheme.com. https://www.streamscheme.com/how-much-does-a-vtuber-model-cost/ (accessed Jan. 16, 2022).
[20] Y. Jin, J. Zhang, M. Li, Y. Tian, H. Zhu, and Z. Fang, "Towards the Automatic Anime Characters Creation with Generative Adversarial Networks," 2017, arXiv:1511.06434
[21] Y. Men, Y. Yao, M. Cui, Z. Lian, X. Xie, and X. Hus, "Unpaired Cartoon Image Synthesis via Gated Cycle Mapping," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, Louisiana, Jun., 18-24, 2022
[22] Z. Liu, "Generating Full-Body Standing Figures of Anime Characters and Its Style Transfer by GAN," M.S. thesis, Dept Computer Science and Communications Engineering, the Graduate School of Fundamental Science and Engineering of Waseda University, Tokyo, Japan, 2020
[23] Z. Lu, C. Shen, J. Li, H. Shen, and D. Wigdor, "More Kawaii than a Real-Person Live Streamer: Understanding How the Otaku Community Engages with and Perceives Virtual YouTubers", CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 2021, pp. 1-14
[24] Y. Ikeda, K. Doman, Y. Mekada, and S. Nawano, "Lesion image generation using conditional GAN for metastatic liver cancer detection," Journal of Image and Graphics, vol 9, no 1, 2021, pp. 27-30