ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

GENERATING MASCOT IMAGES USING STYLEGAN2-ADA WITH SMALL DATASET

Jeoung Gi Kim[1], Je Kyung Lee[2], Kyung Ae Cha[3] , and Jeong Tak Ryu[4*]
[1,2,3]Department of Artificial Intelligence Daegu University Gyeongsan-si, Korea
[4]Department of Electric Engineering Daegu University Gyeongsan-si, Korea
[1]kjk9982097@gmail.com, [2]jkzzang0203@naver.com, [3]chaka@daegu.ac.kr and [4*] jryu@daegu.ac.kr
Corresponding author: jryu@daegu.ac.kr

ABSTRACT

The mascot image is a small, concise design intended to clearly convey meaning. Designing a concise yet meaningful mascot image often requires significant time and consideration during the design process. Additionally, more rigorous feedback processes and data-driven approaches are essential. In this paper, the goal is to reduce manual work and obtain creative ideas when designing mascot images by conducting experiments using the GAN (Generative Adversarial Network) model to generate mascot images. In particular, the experimental results provide information on image-based applications. It is meaningful in that it can be used to develop a living support system for developmentally disabled people who need to concentrate. To this end, this study utilized the StyleGAN2-ADA network to generate new mascot images from a small set of training images. The results confirmed that usable mascot images could be generated through training with a relatively small amount of image data. This approach demonstrates the potential to leverage GAN models, especially StyleGAN2- ADA, to efficiently generate purpose-built images with limited datasets.

Keywords: image generation, generative model, StyleGAN2-ADA

I. INTRODUCTION

Artificial intelligence technology is advancing considerably, undergoing innovative developments. In the midst of such advancements, the emergence of large-scale AI and the attention towards generative AI have garnered significant interest. This is reshaping the paradigm of artificial intelligence technology and catalyzing innovation across various fields. The generative adversarial network (GAN) [1, 2] is a technology that generates or transforms content, such as videos, images, and text, based on training data. GAN is an algorithm that can produce various outputs based on the constructed training data, making it useful in creative fields that require creativity and originality. Furthermore, the application of GAN technology can serve as an intelligent tool for collaborative work in fields such as content creation and design by providing creative ideas [3].

This paper constructs a mascot generation AI system using the StyleGAN2-ADA [4, 5], a GAN algorithm capable of producing various output results through image augmentation techniques. The system aims to create mascot images that represent companies or products in modern society, conveying strong identity and rapport to consumers. StyleGAN2-ADA possesses the characteristic of regenerating by applying various augmentation techniques to the training data. Therefore, it can be effectively utilized in creating mascots of diverse forms and styles. In creative work, similar ideas often appear, which can make it challenging for individuals without design expertise for novice artists to undertake such work effectively. Therefore, in this paper, new images were created by training a GAN model to generate mascot images.

This approach provides design ideas to novice creators and saves production time. The generative artificial intelligence (AI) model implemented in this study used 3,000 mascot images as training data, and the quality of the newly created mascot images was evaluated by comparing them with the original images. The performance of generative AI was verified through this method.

The paper is structured as follows: Section II discusses related research. Next, Section III outlines the process of constructing training data and model generation, and Section IV presents the experimental results. Finally, Section V concludes the paper.

II. RELATED RESEARCH

A. StyleGAN2-ADA

The GAN model is a neural network consisting of generator and discriminator. When the generator creates an image, the discriminator determines the generated image. This process is repeated to create high-quality images.

Within this context, StyleGAN2-ADA has emerged as an advanced model that incorporates the ADA mechanism into the discriminator. This strategic integration serves to mitigate the limitations observed in the original StyleGAN [6] model, resulting in a significantly improved model. Additionally, the augmentation capabilities of ADA offer the notable advantage of enabling the training of the model with a reduced number of images compared to the conventional requirements in traditional GANs.

In the typical scenario, training a high-quality GAN model necessitates a dataset of approximately 100 000 to 1 million images [6]. However, the this study uses a variety of augmentation techniques, including cutout, pixel blitting (blit), geometric transformation (geom), color, image spatial filtering (filter), and noise, within the StyleGAN2-ADA model. Using these techniques, we successfully trained a model using a dataset containing 3000 mascot images. This approach allows generating a diverse range of mascot images.

B. Emoticon Image Production Study

R Chandra and et al.[7] addressed the reseach addresses the process involves leveraging CNNs for facial expression detection and feature extraction, followed by generative models for emoji generation, enabling the creation of emojis that reflect the detected facial expressions accurately. However, this research does not address the creation of new images such as mascots. On the other hand, in our research, we aim to utilize GANs to train emoji-shaped images and produce creative content.

III. MASCOT IMAGE GENERATION AI MODEL

A. Construction of Training Data

The training dataset primarily comprises editable materials with copyright notices. The material is predominantly mascots and characters related to local governments, sports, universities, and other organizations. Additionally, we used freely available images from platforms like Pixabay [8] and Freepik [9] to construct the training dataset. A total of 15 000 mascot images were collected, from which images with frontal facial expressions were selected. For training, each image was converted to 256x256 pixels in red, green, and blue (RGB) (24-bit) format. After data refinement, the final training dataset consisted of 3000 images.

Collecting a sufficient number of training images to suit the intent of a project can be very difficult. However, by taking advantage of StyleGAN2-ADA, we effectively trained the model with only 3000 images, which is approximately 1/50 of the number of images typically required for traditional GAN training. The specifications of the mascot image are illustrated in Fig. 1, depicting a static, front-facing image measuring 256x256 pixels in RGB (24-bit) color format.

Fig. 1: Example of Training images

B. Architecture of the Mascot Image Generation AI Model

The structure of the mascot image generation AI implemented in this paper is illustrated in Fig 2. constructed learning data and augmentation type obtained by collecting and refining images suitable for the purpose of mascot production are input into the StyleGAN2-ADA model for training. After that, a learning result and a learning network are created and a mascot image is created. The resulting output image is a creative representation that has been transformed and synthesized by the AI through a learning model.

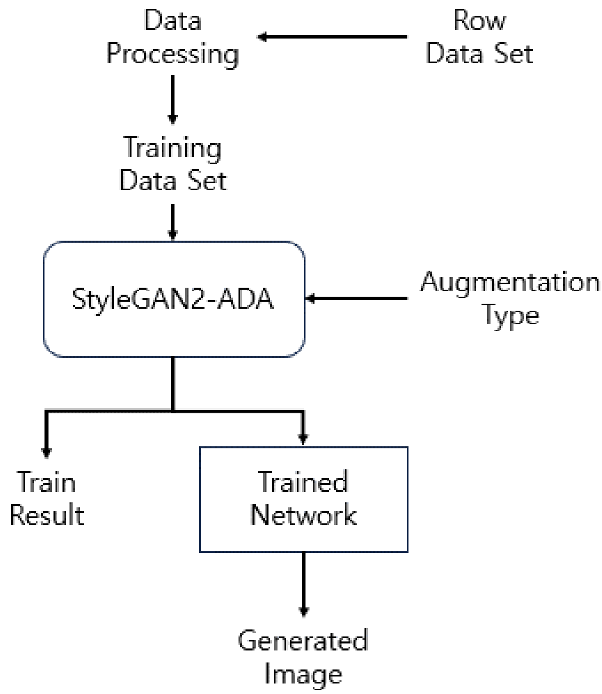


Fig. 2: Image Generation Model Structure

C. Style GAN2-ADA Training

The structure of the mascot image generation AI implemented in this paper is displayed in ADA model involves collecting and refining a training dataset containing images matching the mascot creation intent. The final output image features creative transformations, as AI transforms the image through a model learned from the initial training dataset. Table I presents the training model environment using Python and CUDA Engine/PyTorch on a graphics processing unit (GPU).

Table I Experiment Environment of Model Training

CPU	11th Intel(R) i9-11900k
RAM	64.0GB
GPU	NVIDIA RTX 3080Ti
Anaconda Python v.	3.9.0
CUDA Engine v.	CUDA 11.3
PyTorch v.	11.3

The StyleGAN2-ADA model is a generative AI, which trains on the entire training dataset and transforms and combines styles to create new images. Training is equivalent to the concept of epochs and is performed in Kimg units, representing the number of training repetitions.

The value of kimg determines how many images the discriminator evaluates in one step, and training continues until this value is reached. In addition, 1Kimg means 1000 repetitions, and the generated images are generated at intervals of 100kimg (100 000 repetitions) [10].

To obtain the final learning model, we inspected the generated images, and images that did not meet the mascot specifications were refined to reconstruct the training dataset. This process was repeated three times. Table II represents the training process.

Table II: Stylegan2-Ada Training Processes

Data set	Training amount	Count of augmentation type	Time
15 000	1000kimg	3	~22 h
7000	1000kimg	3	~12 h
3000	4000kimg	6	~10 h

The first training step proceeded for 1000kimg (1 million iterations), and the ADA method incorporated pixel blitting (blit), geometric transformation (geom), and color techniques. The training process for image generation took approximately 22 h. Fig. 3 represents the image results obtained from the first training. The training results included a significant number of green images, and distorted images were observed due to the application of special effects.

Fig. 3: Example of the first experiment result

In the second training, the training proceeded for 1000K img (1 million iterations), and the ADA method incorporated pixel blitting (blit), geometric transformation (geom), and color techniques. Following the first training, a refinement process was conducted to remove green images and images containing special effects or text. The second training was conducted using 7000 images that had undergone refinement, and the training process took approximately 12 h. Fig. 4 presents the results of the generated images. The phenomenon of distortion in the appearance of faces and bodies occurred due to the simultaneous training of both two- dimensional (2D) and three-dimensional (3D) images. Additionally, images were generated in rotated forms.

Fig. 4: Example of the second experiment result.

In the final training, the training proceeded for 4000Kimg (4 million iterations), and the ADA method included pixel blitting (blit), geometric transformation (geom), color augmentation, image space filtering (filter), noise insertion (noise), and cutout techniques. These additions were made to correct the distortion problems observed in the second training and produce images with different colors. Additionally, to reconstruct the training dataset, we created a dataset of 3000 images by removing blue images, 3D images, and those that did not fit the intended pose. Due to the reduction in the size of the training dataset, it was expected that the training time would decrease despite increases in training volume and ADA. Thus, training options expanded. The training process took approximately 10 h.

Fig. 5: Example of the final experiment result.

Fig. 5 represents the results of the final experiment. The images are generated with a quality sufficient for use by businesses, organizations, and similar entities.

Specifically, we incorporated the ADA technique, resulting in the generation of diverse styles of images with accurate representations of facial shapes and positions of facial features. Furthermore, these generated images maintain the same form as the initial training images. However, they are transformed into character images with distinguishable styles, demonstrating the utility of these creatively generated images.

IV. STYLEGAN2-ADA EXPERIMENT RESULTS

Fig. 6-8 depicts the step-by-step process of image generation observed during the training procedure. Fig. 6. is an image labeled Fake0000 and shows a noise image applied during training.

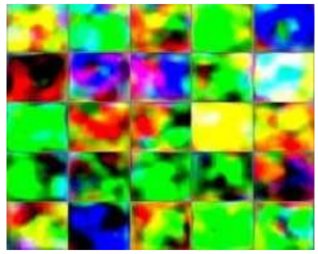
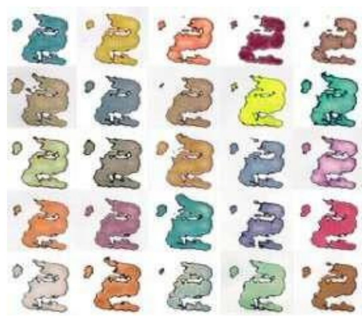


Fig. 6: Initial Random Noise (Fake0000)

Fig. 7 intermediate images generated during the training process are similar to the images labeled Fake0100 and Fake2000. The images generated after the training of 4000Kimg are depicted in Fig. 8. The images generated by the training model are in a format suitable for use as mascot images.



(Fake0200)

(Fake2000)

Fig. 7: Intermediate stages of image generation

Fig. 8: Final Stage of image generation process (Fake4000)

In this paper, we utilized discriminator detection scores based on loss rates and the FID (Frechet Inception Distance) metric to evaluate the performance of the model.

The GAN model is an algorithm in which the generator and discriminator compete with each other. The loss value changes, and after the initial training, it converges to the target value within a specific range. Adversarial training is conducted through this process [11].

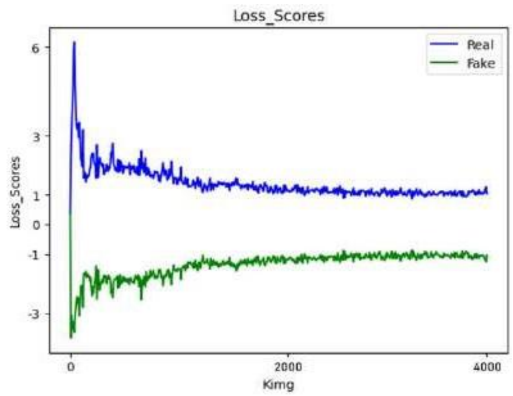


Fig. 9: Real and fake score graphs

Fig. 9 visualizes the Loss/Score values, which are obtained by dividing the losses of the generator and discriminator by the discriminator's detection score in the StyleGAN2-ADA model. The Real_score indicates the extent to which the discriminator accurately identifies input images as real images, with higher values indicating superior performance of the discriminator. On the other hand, the Fake_score represents the degree to which fake images are incorrectly classified as real images. Therefore, lower Fake_score values indicate better performance of the discriminator. This indicates that as the denominator, Real_score, increases, the Loss/Real_score values decrease. On the contrary, it can be observed that the Loss/Fake_score values increase. This signifies that as the denominator, Fake_score, decreases, the Loss/Fake_score values increase. These results indicate that the discriminator and generator of the model are effectively learning.

The FID (Frechet Inception Distance) is an evaluation metric used to compare the quality between real images and images generated by the model [12]. This metric calculates the distance between the distributions of images generated using GAN and real image data, quantifying the similarity of images as a distance metric. FID is used to measure and compare the performance of generative models like GAN.

As training progresses, a decrease in the FID value indicates a reduction in the distance between the distributions of real and generated images, signifying higher quality generated images.

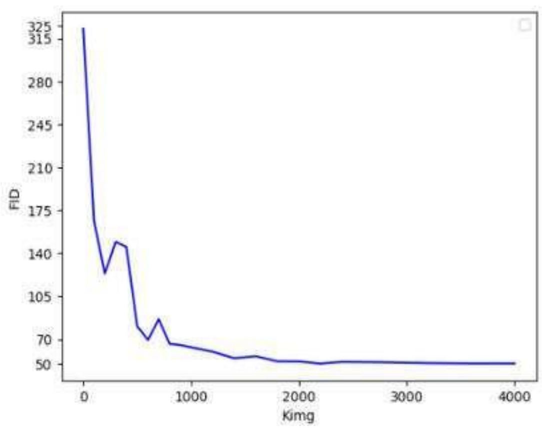


Fig. 10: FID value graphs

Fig. 10 visualizes the FID values of the StyleGAN2-ADA model. The vertical axis represents the FID values, while the horizontal axis represents the training units of the model, denoted as kimg for StyleGAN2-ADA. Through Fig. 10, it can be observed that as the training progresses, the FID values decrease. This indicates that the training of the generator is proceeding normally, and the final FID value of this model is 50.34.

Fig. 11 represents the mascot images generated in this study. The generated images emphasize the face and distinctly depict facial features, showcasing their characteristic traits. However, shortcomings exist regarding representing the intricate details and special effects of mascot images. Nevertheless, through these generated images, we can provide ideas and initial design drafts to novice artists and the public.

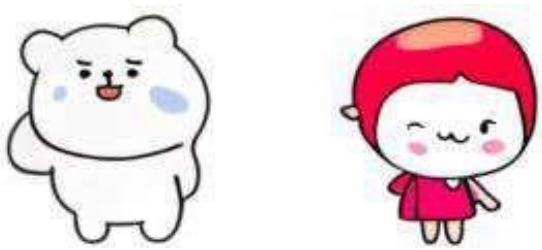


Fig. 11: Examples of generated images

V. CONCLUSIONS

In this paper, we conducted an experiment to automatically generate creative mascot images using a GAN model. By refining the training data, we were able to produce mascot images that accurately represent the organization's meaning. Future advancements could involve harnessing advanced Generative Adversarial Network (GAN) models to generate emoji images from various perspectives, including frontal views. These advancements hold the potential to transform the image-making industry, enhancing its dynamism and versatility.

The results of this study, in particular, can be applied to support systems for people with special needs. To quickly adapt developmentally disabled individuals to the use of certain systems, it is necessary to increase their interest in the devices and encourage frequent access to them. To achieve this goal, we plan to develop applications using generated AI images to enhance the concentration of people with special needs on various IT devices.

ACKNOWLEDGMENT

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5C2A07091326).

REFERENCES

[1] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks.arXiv preprint arXiv:1511.06434.

[2] Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X., & Wang, F. Y. (2017). Generative adversarial networks: introduction and outlook.IEEE/CAA Journal of Automatica Sinica,4(4), 588-598.

[3] Cui, M., Kim, M., Choi, S., & Lee, S. (2022). The Usage and Impact of GAN in Graphic Design.Archives of Design Research,35(4), 285-307.

[4] Liao, C., Sawayama, M., & Xiao, B. (2022). A Perceptual Evaluation of the StyleGAN2-ADA Generated Images of Translucent Objects.Journal of Vision,22(14), 3642-3642.

[5] Lu, Y. (2023). Style-based Image Manipulation Using the StyleGAN2-Ada Architecture.Applied and Computational Engineering.

[6] Tero Karras, Samuli Laine, Timo Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[7] Ramesh Chandra, G., Golla, R. C., Gangireddy, T., Nagula, A., & Bala, D. S. (2023, March). Emoji Creation from Facial Expression Detection Using CNN. In Proceedings of Third International Conference on Advances in Computer Engineering and Communication Systems: ICACECS 2022 (pp. 303-312). Singapore: Springer Nature Singapore.

[8] Pixbay, [Online]. https://pixabay.com/ko. [access: January 31, 2024]

[9] Freepik, [Online]. https://www.freepik.com. [access: January 31, 2024]

[10] T. Karras and J. Hellsten, "StyleGAN2-ADA-Pytorch," [Online]. https://github.com/NVlabs/stylegan2-ada-pytorch/. [access: January 31, 2024]

[11] Talib, D. A., & Abed, A. A. (2023). Real-Time Deepfake Image Generation Based on Stylegan2- ADA. Revue d'Intelligence Artificielle, 37(2). [13] M. Heusel, H. Ramsauer, T. Unterthiner and B. Nessler, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," Neural Information Processing Systems, pp. 6629-6640, 201

[12] M. Heusel, H. Ramsauer, T. Unterthiner and B. Nessler, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," Neural Information Processing Systems, pp. 6629-6640, 201