Article

BaMSGAN: Self-Attention Generative Adversarial Network with Blur and Memory for Anime Face Generation

Xu Li [1,*][ID], Bowei Li [2], Minghao Fang [3], Rui Huang [4] and Xiaoran Huang [5]

1 Department of Computer, Central South University, Changsha 410083, China
2 School of Telecommunication Engineering, Xidian University, Xi'an 710126, China; 20012100063@stu.xidian.edu.cn
3 Zhejiang University-University of Illinois at Urbana-Champaign Institute, Zhejiang University, Haining 314400, China; minghao.22@intl.zju.edu.cn
4 School of Earth Sciences, Zhejiang University, Hangzhou 310000, China; hrhr@zju.edu.cn
5 School of Software Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; b22041704@njupt.edu.cn
* Correspondence: 8208200711@csu.edu.cn

Abstract: In this paper, we propose a novel network, self-attention generative adversarial network with blur and memory (BaMSGAN), for generating anime faces with improved clarity and faster convergence while retaining the capacity for continuous learning. Traditional self-attention generative adversarial networks (SAGANs) produce anime faces of higher quality compared to deep convolutional generative adversarial networks (DCGANs); however, some edges remain blurry and distorted, and the generation speed is sluggish. Additionally, common issues hinder the model's ability to learn continuously. To address these challenges, we introduce a blurring preprocessing step on a portion of the training dataset, which is then fed to the discriminator as fake data to encourage the model to avoid blurry edges. Furthermore, we incorporate regulation into the optimizer to mitigate mode collapse. Additionally, memory data stored in the memory repository is presented to the model every epoch to alleviate catastrophic forgetting, thereby enhancing performance throughout the training process. Experimental results demonstrate that BaMSGAN outperforms prior work in anime face generation, significantly reducing distortion rates and accelerating shape convergence.

Keywords: anime face generation; self-attention generative adversarial network; blur dataset; memory replay; generative adversarial network; self-attention

MSC: 68T07

1. Introduction

In the ever-evolving landscape of modern computer vision, image generation has emerged as a prominent and extensively explored research domain. Drawing upon the capabilities of artificial intelligence and machine learning, the field of image generation is dedicated to crafting images that exhibit realistic appearances and a rich diversity of features. Particularly noteworthy is the pervasive influence of Japanese anime across various domains, igniting profound interest in anime face generation. Within the expansive realm of image generation, anime face generation has carved out a distinctive niche, capturing substantial attention from the realms of academia and industry alike. The allure of breathing life into anime-inspired visages has ignited a fervent pursuit of innovation and excellence in this specialized subfield.

The realm of machine learning and neural networks has witnessed unprecedented growth, with numerous applications such as asset allocation [1] and non-negative matrix factorization [2]. These cutting-edge technologies have found utility in diverse domains, from natural language processing to medical diagnostics. The profound impact of AI and

machine learning has been especially pronounced in image generation, where they play a pivotal role in creating images with lifelike qualities and intricate features. The fusion of AI and machine learning has unleashed a wave of innovation, propelling the field of image generation into new horizons.

With the ongoing advancements in deep learning, particularly with the emergence of generative adversarial networks (GANs) [3], significant progress has been achieved in image generation. Among the various GAN architectures, deep convolutional GANs (DCGANs) [4], based on deep convolutional networks [5], have demonstrated exceptional image-processing capabilities and have succeeded in generating high-quality images. However, despite their achievements, DCGANs still face certain limitations, particularly when handling long-range dependencies that necessitate passing through several convolutional layers due to their local receptive fields.

To address these concerns and further enhance the quality of image generation, self-attention GANs (SAGANs) [6] were developed. SAGANs incorporate the self-attention mechanism [7], enabling it to effectively capture long-range dependencies while retaining the advantages of DCGANs. This fusion of self-attention architecture with DCGANs has demonstrated promising potential for various image generation tasks. Nevertheless, when applied specifically to the generation of anime faces, SAGANs encounter certain challenges.

Notably, the generated anime images still exhibit certain shortcomings, such as blurry edges and distorted facial features. This indicates the need for further improvements to achieve even higher image quality in anime face generation. Additionally, although SAGANs propose some techniques to mitigate issues like mode collapse, a well-known problem in GAN training, where the model collapses to generating a limited variety of samples, there is still room for improvement to ensure stable and diverse image generation results. Furthermore, during the training process, the performance of the SAGAN tends to deteriorate. One possible explanation is the phenomenon of catastrophic forgetting, which describes a sharp decline in the model's performance on previously learned tasks as soon as a new task is introduced.

To address the challenge of generating more detailed anime images, researchers have developed CartoonGAN [8]. This model constructs its own blurry dataset using methods from digital image processing (DIP), which includes the canny edge-detection algorithm and Gaussian blur. Additionally, it combines the traditional loss function with the edge-promoting adversarial loss function to encourage the model to avoid generating images with blurry edges. Moreover, AnimeGAN [9] employs additional techniques to create fake datasets containing a broader range of negative samples. However, all these anime-related works primarily focus on Image style transfer [10], and there is a lack of powerful methods for generating anime faces from random noise. To tackle the issues of mode collapse and catastrophic forgetting, the closed-loop memory GAN [11] introduces a memory structure that assists the model in learning from past data, thereby improving the variety of generated pictures. This approach shows promise in addressing these challenges effectively.

In this work, we take a two-fold approach to address the challenges in anime face generation. First, we introduce a regularization parameter into the optimizer of SAGANs to mitigate mode collapse. Secondly, we present SAGANs with blur and memory, abbreviated as BaMSGAN. This innovative model incorporates both a blur dataset and a memory structure into the self-attention GAN framework. The self-attention structure plays a crucial role in enabling the model to capture features from all positions and learn long-range dependencies. Additionally, we introduce a blurry dataset, randomly sampled from the original dataset and processed using digital image processing (DIP) techniques. This dataset is used as negative data in the discriminator to encourage the model to avoid generating images with blurry edges. The memory structure is designed to store previously generated images, which are then utilized to assist the discriminator in learning from the model's historical output. Extensive experiments were conducted on an anime face dataset [12] comprising 63,566 images at a resolution of 64×64 pixels. The results confirm the effectiveness of the proposed model, BaMSGAN, which significantly outperforms

previous methods in anime face generation. It generates more detailed images at a faster pace, with a noticeable reduction in distortion rates after each epoch.

2. Related Work

2.1. CartoonGAN

In recent years, there has been a growing interest in creating cartoon and anime-style images using style transfer techniques [13,14]. One notable approach is the use of "CartoonGAN", as proposed by researchers [8]. They have presented a solution for transforming real-world scenes into cartoon-style images by training on unpaired photos and cartoon images. However, there are a few limitations to consider when using this technique. Firstly, achieving high-quality results requires a significant amount of time for training and substantial computing resources. Secondly, while CartoonGAN primarily focuses on converting landscapes into anime-style representations, it often lacks emphasis on transforming human faces into cartoon-like versions. The work in the area of style transfer for cartoon and anime-style images has demonstrated promising results, yet there remains ample room for improvement and further exploration.

2.2. StyleGAN and ProGAN

StyleGAN, introduced by Karras et al. [15] and Progressive GAN (ProGAN), proposed by Karras et al. [16], have made significant strides in synthesizing high-quality images, including anime faces. StyleGAN's adaptive synthesis of styles and ProGAN's progressive refinement of image resolution have paved the way for more realistic and diverse anime-style images. These models emphasize the preservation of anime-specific features, such as distinct eye shapes, colorful hair, and various facial expressions. The success of StyleGAN and ProGAN highlights the importance of fine-grained control and the ability to capture nuanced details in anime characters.

2.3. CycleGAN and Pix2Pix

CycleGAN, developed by Zhu et al. [17], and Pix2Pix, designed by Isola et al. [18], have earned acclaim for their versatility in image-to-image translation tasks, including anime face style transfer. These models offer artists and designers the ability to apply the style of one image to another while preserving the structural integrity of the content. This innovation allows for the conversion of real-world photographs into anime-style images, vastly broadening the creative possibilities for anime-inspired content. The capacity to seamlessly transform photographs into anime-style renderings has become a valuable tool for content creators and enthusiasts, opening new avenues for artistic expression and generating unique visual experiences.

2.4. DCGAN

Deep convolutional GANs, or DCGANs, are foundational models in the field of image generation. Proposed by Radford et al. [4], DCGANs are celebrated for their effectiveness in generating high-quality images. The architecture employs a deep convolutional neural network for both the generator and discriminator. DCGANs revolutionized the field by addressing key challenges in GAN training, such as mode collapse and slow convergence. They excel in capturing local and global image features, making them an ideal choice for a wide range of image synthesis tasks, including anime face generation. DCGANs serve as a benchmark and point of reference for evaluating the capabilities and advancements of subsequent GAN architectures.

2.5. WGAN

The Wasserstein GAN, or WGAN, introduced by Arjovsky et al. [19], tackles issues related to training stability and mode collapse, which are common challenges in GANs. WGAN introduces a novel loss function based on the Wasserstein distance, also known as the Earth mover's distance. This distance metric provides a more reliable and interpretable

measure of the quality of generated images. The WGAN is known for its ability to produce high-quality images with better training dynamics, making it an important advancement in the GAN landscape. In the context of anime face generation, the WGAN's contributions include enhanced training stability, enabling the generation of more coherent and diverse anime-style faces.

2.6. SAGAN

The emergence of generative adversarial networks (GANs) [20] has resulted in remarkable progress in this field, notwithstanding the existence of several unresolved issues [21]. GANs that leverage deep convolutional networks [4,22] have particularly achieved significant success. The DCGAN represents the first successful utilization of convolutional neural networks within GANs [4]. Subsequently, several other GAN variants have been created, such as the Wasserstein GAN [19] and SNGAN [23]. Self-attention generative adversarial networks (SAGANs) [6] distinguish themselves from traditional convolutional GANs by incorporating a self-attention mechanism within the GAN framework, effectively capturing long-range dependencies. By integrating self-attention into both the generator and discriminator modules, the SAGAN enables the generation of detailed images by leveraging information from all feature locations. This comprehensive analysis of highly detailed features across distant parts of the image ensures their consistency with each other. However, in our experiments involving the generation of anime faces using the original SAGAN model, we encountered specific challenges. Notably, we observed a high degree of distortion and blurred edges in the generated anime faces. In response, we developed our own model to address these issues and enhance the overall quality of the generated anime faces.

2.7. Memory

Catastrophic forgetting [24,25] is a common problem in neural networks. When a trained network is retrained on a new task, its performance on the original task typically deteriorates. For the generative adversarial network (GAN), each training phase is equivalent to training on a new dataset, making GANs susceptible to catastrophic forgetting.

Several methods have been proposed to address the problem of forgetting in GAN learning. The MeRGAN [26] employs two anti-forgetting mechanisms, joint training and replay alignment, to mitigate the forgetting issue in GANs. While it provides some relief from forgetting in sequential learning, it introduces additional memory recall mechanisms, increasing computational complexity and storage requirements.

CloGAN [11] uses a dynamic memory unit to store generated samples and filters them with an embedded classifier. This dynamic memory reduces memory costs and enhances sample diversity but necessitates additional training of new models, thereby elevating computational complexity.

GAN memory for lifelong learning [27] tackles memory issues in lifelong learning through sequential style adjustment. However, this method can be time-consuming when dealing with large-scale datasets.

CAM-GAN [28] addresses the forgetting problem through feature map transformation and residual bias in feature space, requiring fewer parameters. However, the success of this method depends on selecting suitable base tasks; otherwise, it may affect the model's performance.

3. BaMSGAN

A typical generative adversarial network comprises two convolutional neural networks: the discriminator $\mathcal{D}$ and the generator $\mathcal{G}$. The generator generates fake images from random noise, while the discriminator classifies whether the input image comes from the real dataset or is generated by the generator (i.e., real or fake). As training progresses, the discriminator $\mathcal{D}$ becomes more proficient at distinguishing between real and fake im-

ages, while the generator $\mathcal{G}$ improves its ability to deceive the discriminator by producing high-quality images. Obviously, our final goal is to solve this problem:

$$(\mathcal{G}^*, D^*) = \arg \min_{G} \max_{D} \mathcal{L}(\mathcal{G}, \mathcal{D})$$

Several earlier modules, such as Cartoon GAN and Anime GAN, have achieved great success in style transfer tasks. However, tasks like generating anime images from noise and anime face generation have been largely overlooked. Building upon the foundation of self-attention generative adversarial networks, we have adapted our module to suit these tasks. We formulate a generating function that takes the original anime manifold $\mathcal{A}$, the edge-blurred anime manifold $\mathcal{B}$, and eventually generates a set of fake anime manifold $\mathcal{F}$ in the same style. The generating function is trained using the original anime dataset $\mathcal{S}_{data}(a) = \{a_i | i = 1 \dots N\}$ and the edge-blurred anime dataset $\mathcal{S}_{blur}(b) = \{b_i | i = 1 \dots M\}$. In Section 4, we will present in detail the four key components and the overall structure of our architecture.

3.1. SAGAN with Regularization

SAGAN, or self-attention generative adversarial network, is an advanced architecture within the realm of generative adversarial networks. It introduces self-attention mechanisms to enhance the consistency and quality of fine details in generated images. By employing self-attention, SAGAN gains the ability to grasp global structures and dependencies within images during the generation process, resulting in the production of more lifelike and nuanced images. Given these capabilities, we opted to apply SAGAN to the task of anime face generation.

However, our initial application of the standard SAGAN architecture to anime face generation revealed a pressing issue: mode collapse. In this situation, the generator begins to produce images with patterns that are strikingly similar, causing a significant reduction in diversity among the generated samples. In response to this challenge, we embarked on a journey to regularize our optimization process, ultimately aiming to prevent overfitting and reinvigorate the generation of anime face images with a rich variety of patterns.

Through numerous iterations and experiments, we arrived at a regularization parameter of 0.0001 as an effective solution. This regularization procedure is a critical component in the enhancement of image generation in our model. Below, we provide visual representations of the anime face samples before and after the application of regularization.

Mode collapse, a recurrent issue in GANs, occurs when the generator becomes fixated on generating a restricted set of patterns or samples, thereby limiting the diversity in the generated images. Regulation optimization techniques [29], such as adding regularization terms to the loss function [30], offer a means to counteract mode collapse. Methods like weight decay (L2 regularization) or gradient penalties serve to encourage smoother behaviors in both the generator and discriminator networks. As a result, this regularization fosters a more stable training process, better convergence, and a marked reduction in mode collapse occurrences. The samples of original SAGAN (mode collapse) and samples after regularization are shown in Figure 1.

(**a**)

Figure 1. *Cont.*

(**b**)
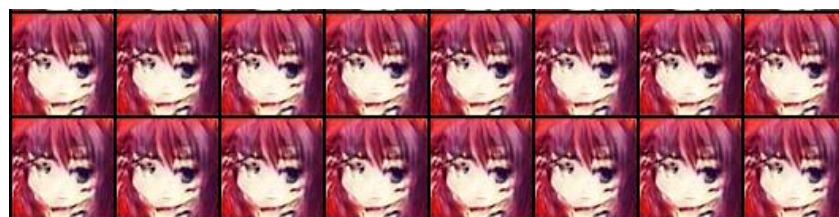
Figure 1. A comparison between generated images before and after regularization. (**a**) A sample generated before adding regularization with serious mode collapse. (**b**) A sample generated after we add regularization containing varieties of faces.

3.2. Edge Blur

To enhance image clarity, we drew inspiration from the Cartoon GAN's architecture. We incorporated a GAN module that would transform real images into a cartoon style. Our strategy involved blurring the edges of the input images and integrating them into our loss function. The intention was for the discriminator, during training, to classify these edge-blurred images as 'fake', thus encouraging the generator to produce sharper images. This approach was a resounding success.

In the preprocessing phase, we selected Gaussian blur in conjunction with Canny edge detection to address the challenge of generating more distinct anime pictures and improving overall image clarity. A subset of the training dataset underwent a blurring preprocessing step. This step enabled the discriminator to better recognize edge-blurred images as "fake", ultimately instructing the generator to create clearer images.

To create a database of edge-blurred images for training, we employed Canny edge detection to outline image edges and applied Gaussian blur to them. This database was then utilized to train the discriminator. The primary goal of this approach was to compel the model to avoid generating images with blurry edges, resulting in anime faces with significantly improved clarity. The algorithm of Edge Blur is shown in Algorithm 1 as pseudo code. An example of Edge Blur is shown in Figure 2a,b.

Algorithm 1 Algorithm of Edge Blur

Input: *Image*, $t1$, $t2$, k
Output: $Image_{blur}$
 1: $Image_{blur} \longleftarrow Image$
 2: $Image_{canny} \longleftarrow Canny(Image, t1, t2)$ {Use canny to get the outline of the edges}
 3: **for** each $i \in height$ **do**
 4:     **for** each $j \in width$ **do**
 5:         **if** $Image_{canny}=0$ **then**
 6:             $Image[i,j] \longleftarrow [0,0,0]$
 7:         **end if**
 8:     **end for**
 9: **end for**
10: $dst = GaussianBlur(Image, k)$ {Apply GaussianBlur to the edges outlines previously,k is the kernal size}
11: **for** each $i \in height$ **do**
12:     **for** each $j \in width$ **do**
13:         **if** $Image_{canny} \mathrel{!}= 0$ **then**
14:             $Image_{blur}[i,j] \longleftarrow dst[i,j]$ {Replace the original edge with the GaussianBlur version}
15:         **end if**
16:     **end for**
17: **end for**
18: **return** $Image_{blur}$

(**a**)　　　　　　　　　　　　　(**b**)

Figure 2. By applying Gaussian blur transformation to the edges outlined by Canny, we could get edge-blurred images from the original database. (**a**) Original image; (**b**) edge-blurred image.

3.3. Memory

We have established a memory repository with a predefined capacity to optimize storage efficiency. In the GAN training process, we follow a deliberate approach. Initially, we allow the GAN to train until it reaches a state of convergence. This strategy is employed because the quality of images generated by the GAN at the outset of training tends to be subpar. These initial images lack the quality necessary to significantly challenge the discriminator, and memory allocation consumes valuable time and resources. Thus, we have a stipulation that, once the GAN has reached a stage where it can generate high-quality images, our memory repository commences the storage of a selection of these generator-produced images.

When the memory repository reaches its maximum capacity, we adopt a policy of random sampling and deletion to manage its contents. This periodic removal of images ensures that the memory repository remains a dynamic and adaptive source of historical data. The images stored in this memory repository are subsequently utilized as fake data in the training process, provided to the discriminator. In doing so, we integrate historical data into the training of the GAN, effectively improving its overall performance.

This memory repository approach helps address issues related to mode collapse and catastrophic forgetting while also contributing to the training of the GAN using a more diverse and comprehensive dataset. Additionally, it allows for the generation of anime faces with improved clarity and fidelity.

3.4. Loss Function

Our loss function comprises three key components. The first component is the adversarial loss, a standard feature in adversarial generative networks. Here, the generator G is tasked with producing anime avatars $G(z)$ from random noise z, which is drawn from a Gaussian distribution $P_z(z)$. The goal is for these generated avatars to deceive the discriminator D, which, in turn, must accurately distinguish between anime avatars in the real anime face dataset $S_{data} = \{x_i | i = 1 \ldots N\}$ and those created by the generator.

The second component involves the edge loss. We prepare an edge-blurred dataset $S_{blur} = \{b_i | i = 1 \ldots M\}$ in advance, where M represents the total number of blurred images. These blurred images are used as fake inputs, allowing our GAN to place a stronger emphasis on capturing the edge features, which are particularly relevant to anime avatars.

The third component is the memory loss. We augment the training process by using anime avatars generated by a segment of the generator as fake inputs and store them in a memory dataset $S_{memory} = \{m_i | i = 1 \ldots P\}$. This dataset captures the historical data, allowing our GAN to learn from previous tasks and avoid deteriorating with extended training periods.

Our loss function is formulated as follows:

$$Loss_D = \mathbb{E}_{x_i \sim S_{data}} log(D(x_i)) + \mathbb{E}_{Z \sim p_z(z)} log(1 - D(G(z))) + \mathbb{E}_{b_i \sim S_{blur}(x)} log(1 - D(b_i)) +$$
$$\mathbb{E}_{m_i \sim S_{memory}(x)} log(1 - D(m_i)).$$

3.5. Our Model

In our model, we harness the power of self-attention mechanisms, a cornerstone of the BaMSGAN, to significantly enhance the generation of anime faces. Self-attention introduces a critical element to our architecture, enabling the model to capture long-range dependencies and intricate spatial relationships within the input images. By implementing self-attention, our BaMSGAN can simultaneously focus on different regions of the image, effectively capturing the global contextual information. This capability empowers both the generator and discriminator to discern complex structures and connections within anime faces, a task often challenging for traditional convolutional architectures.

The integration of self-attention is pivotal for improving the quality of our generated anime faces. It ensures that the model comprehends and synthesizes anime features more cohesively, resulting in sharper and more realistic images. Furthermore, it enhances the discriminator's ability to distinguish between real and generated images, thereby facilitating a more robust and stable training process. The inclusion of self-attention in BaMSGAN underscores our commitment to overcoming the limitations of earlier GAN architectures. It plays a central role in elevating the capabilities of our model, ultimately leading to superior anime face generation and the establishment of new benchmarks in the field.

As for our training approach, we first extract a portion of the samples from the original dataset for edge-blur processing to create a blur dataset. Subsequently, the images from the original dataset serve as positive examples, while the generated images and images from the blur dataset are used as negative examples during training. Once the training converges, the images generated from the sampled portion are preserved as a memory dataset.

During the training process, we continually sample and randomly remove some memory data, which then serves as negative examples to further enhance the training of the BaMSGAN. The whole training process and model architecture are shown in Figure 3.
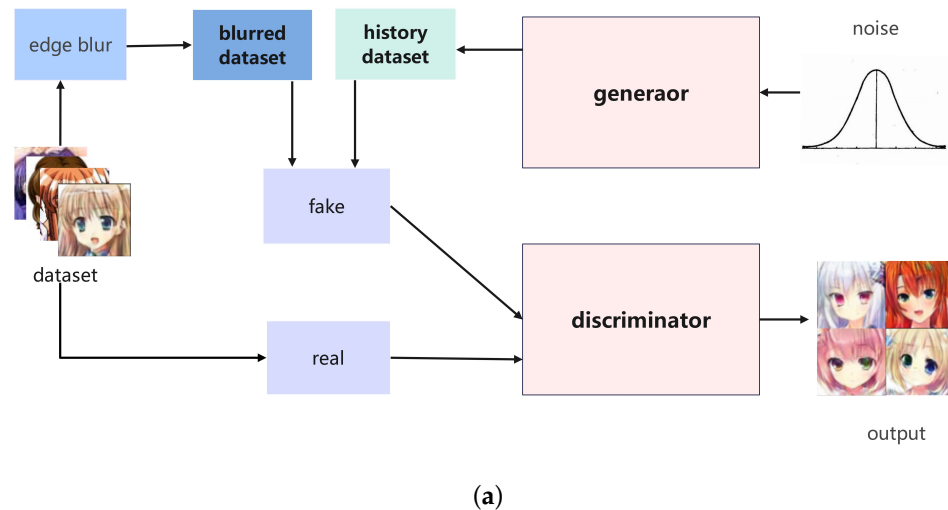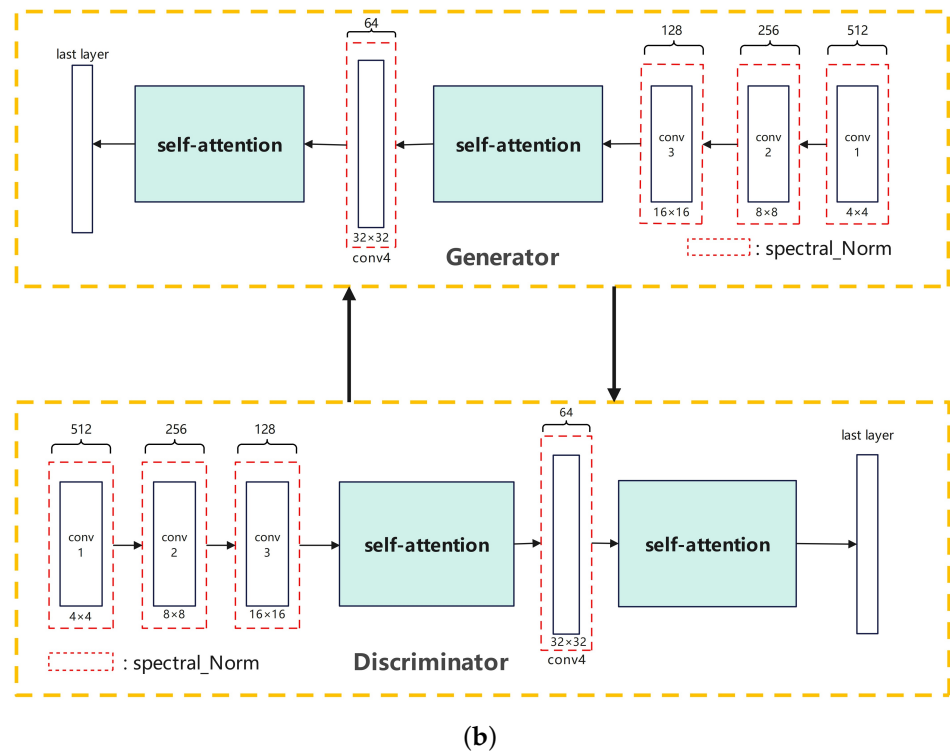

(**a**)

Figure 3. *Cont.*

(**b**)

Figure 3. An overall schematic of our BaMSGAN. (**a**) The overall architecture of our BaMSGAN. (**b**) Our BaMSGAN generator and discriminator hierarchy.

4. Experiments

We trained and implemented our module in Torch and Python. All our experiments were conducted using an NVIDIA RTX 4090 GPU. The BaMSGAN is capable of generating high-quality animated images and does not demand extensive computing resources or time. With less than an hour of training on the RTX 4060, it can produce relatively high-quality animated images. The results generated by the BaMSGAN are displayed in Figure 4.

Our baselines include Wasserstein GANs with gradient penalty (WGAN-GP) and DCGANs. Initially, we will compare the BaMSGAN with the WGAN-GP and DCGAN to showcase the BaMSGAN's performance. Subsequently, we conduct ablation experiments involving a SAGAN with regularization, a SAGAN with regularization and edge-blurred images, and the BaMSGAN to assess the effectiveness of each component within our loss function.

Figure 4. The result of BaMSGAN from epoch 50 to epoch 300.

During the training of these modules, we extracted samples at both 30 epochs and 300 epochs. The results at 30 epochs represent the short-term performance of these modules, while the results at 300 epochs reflect their long-term performance.

4.1. Data

The dataset we use contains approximately 63,565 anime face images obtained from GitHub [12]. All images have been resized to 64 × 64 pixels. The blurred database is one-tenth the size of the dataset, comprising roughly 6000 anime faces with blurred edges.

4.2. Comparison with Prior Work in Anime Face Generation

To confirm the BaMSGAN's superiority in anime face generation, we compare the generated results of the BaMSGAN, DCGAN, and WGAN-GP following training. The results are showcased in Figure 5. In contrast to the DCGAN, our model significantly enhances the diversity of generated anime faces. The results reveal that the DCGAN, both in the short run and the long run, is plagued by a severe problem of mode collapse, producing nearly identical and low-quality faces. Conversely, our model produces clear faces with diverse facial features such as hair color and facial expressions, creating more vivid characters.

Compared to the WGAN-GP, which generates severely distorted and blurred faces, our model produces anime faces with well-defined edges (evident in the hair curves) and natural eyes, noses, and mouths. The inclusion of edge-blurred images in the loss function sharpens the edges, enhancing the clarity of anime faces. Additionally, the memory loss in the loss function contributes to normalizing facial features, as the discriminator learns to recognize the twisted faces generated during earlier epochs as fake images.
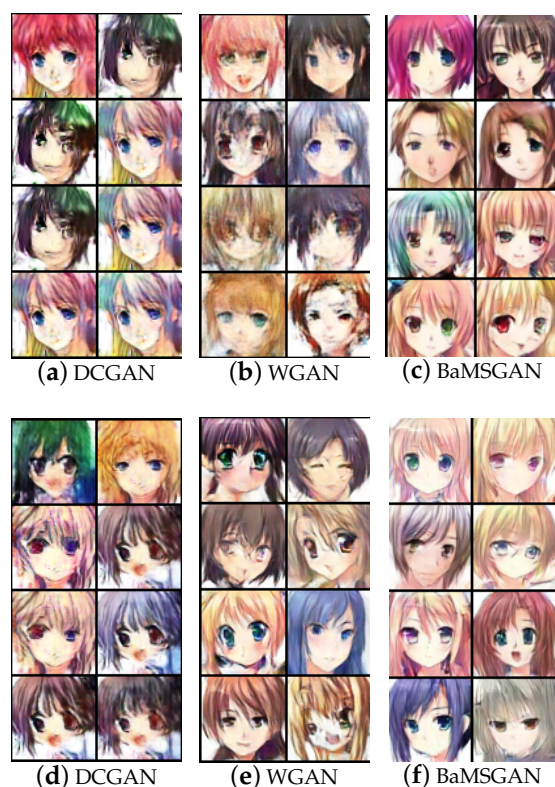

(**a**) DCGAN (**b**) WGAN (**c**) BaMSGAN

(**d**) DCGAN (**e**) WGAN (**f**) BaMSGAN

Figure 5. Comparisons between our model and two baselines ((**a**–**c**) at their 30 and (**d**–**f**) at their 300 epochs).

Moreover, we also compare the DR (distortion rate: the number of distorted faces/the number of all the faces in every epoch) between our model and the SAGAN at epoch 290 to

epoch 298. Compared to the SAGAN, our model can also effectively improves the quality of images in the later stages of training and possesses the ability for continuous learning.

These results underscore the effectiveness of BaMSGAN in overcoming the limitations observed in other GAN architectures, producing anime faces with greater clarity and fidelity. Table 1 presents a comparison of the distortion rate (DR) between our BaMSGAN model and the SAGAN baseline at epochs 290 to 298 during the experiment

Table 1. Comparisons of DR (distortion rate: the number of distorted faces/the number of all the faces in every epoch) between our model and SAGAN at epoch 290 to epoch 298.

Epoch	SAGAN(DR)	BaMSGAN(DR)
290	0.1090	0.0390
291	0.1211	0.0390
292	0.0977	0.0467
293	0.1250	0.0625
294	0.1289	0.0313
295	0.1289	0.0469
296	0.1875	0.0391
297	0.0977	0.0703
298	0.0820	0.0703
Average	0.1163	0.0495

4.3. Ablation Experiment about the Loss Function

To understand the individual functions and roles of each component within the loss function, we conducted a series of ablation experiments. We systematically introduced the edge-blurred image loss and the memory loss into the original GAN loss function (already incorporating regularization), one at a time, and compared the results with the BaMSGAN. The outcomes of these experiments are visualized in Figure 6.
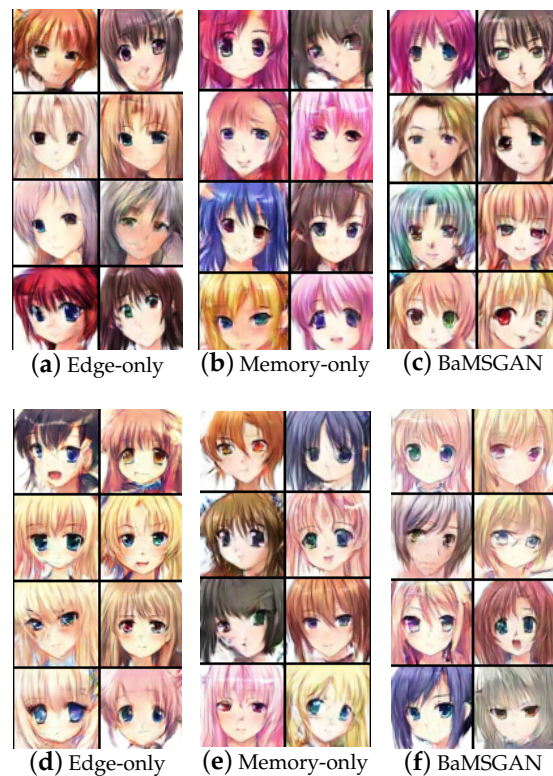

(**a**) Edge-only   (**b**) Memory-only   (**c**) BaMSGAN

(**d**) Edge-only   (**e**) Memory-only   (**f**) BaMSGAN

Figure 6. Ablation experiment result compared to our final model. We add edge-blurred images and memory to the loss function respectively and compare them with our final model ((**a**–**c**) at their 30 and (**d**–**f**) at 300 epochs).

The experimental results clearly demonstrate that the BaMSGAN consistently achieves superior performance, both in the short-term and long-term training, producing anime avatars that closely resemble real characters. In the edge-only group experiment, it's evident that the image quality in the short-term is significantly lower than the BaMSGAN, and there's a noticeable deterioration in image quality over prolonged training. In the memory-only group experiment, it becomes apparent that without the additional training facilitated by the edge-blur loss, the BaMSGAN struggles to generate higher-quality images, with some images appearing missing or blurred.

These ablation experiments highlight the indispensable role played by each component in our loss function, emphasizing their collective contribution to the BaMSGAN's remarkable performance in anime face generation.

5. Conclusions and Future Work

In this paper, we employ a SAGAN enhanced by historical information and edge blurring for anime face generation. Our model introduces regularization operations to the original SAGAN to enhance the stability of SAGAN training. We selected a subset of images from the source dataset, applied edge detection using the Canny algorithm, and used Gaussian blurring to create a blurred dataset. Additionally, we saved a portion of the images generated during the training process to construct a historical dataset. These two datasets are then employed as fake data to bolster training, resulting in improved performance in anime face generation. Experimental results demonstrate that our model outperforms DCGAN, WGAN, and SAGAN, generating clearer, higher-quality images and converging more rapidly.

In our future work, we aim to explore additional methods to enhance GAN performance in specific tasks through modifications to the loss function. We will also test our model on a wider range of datasets to enhance its generalization capabilities. Furthermore, we intend to investigate the potential impact of the blurred dataset on the overall image generation style.

Author Contributions: X.L., B.L., M.F., R.H. and X.H. contributed equally to this work and should be considerd co-first authors. Conceptualization, X.L.; Methodology, X.L. and R.H.; Software, M.F. and R.H.; Validation, M.F. and R.H.; Formal analysis, X.L., B.L., M.F., R.H. and X.H.; Investigation, B.L. and X.H.; Data curation, B.L. and M.F.; Writing—original draft, X.L., M.F., R.H. and X.H.; Writing—review & editing, X.L., B.L., R.H. and X.H.; Visualization, B.L. and X.H.; Supervision, X.L. All authors have read and agreed to the published version of the manuscript.

References

1. Leung, M.F.; Wang, J.; Che, H. Cardinality-constrained portfolio selection via two-timescale duplex neurodynamic optimization. *Neural Netw.* **2022**, *153*, 399–410. [CrossRef] [PubMed]
2. Li, C.; Che, H.; Leung, M.F.; Liu, C.; Yan, Z. Robust multi-view non-negative matrix factorization with adaptive graph and diversity constraints. *Inf. Sci.* **2023**, *634*, 587–607. [CrossRef]
3. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]

4.   Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.

5.   Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

6.   Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-attention generative adversarial networks. In Proceedings of the International Conference on Machine Learning (PMLR 2019), Long Beach, CA, USA, 9–15 June 2019; pp. 7354–7363.

7.   Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.

8.   Chen, Y.; Lai, Y.-K.; Liu, Y.-J. CartoonGAN: Generative adversarial networks for photo cartoonization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 9465–9474.

9.   Chen, J.; Liu, G.; Chen, X. AnimeGAN: A novel lightweight GAN for photo animation. In *Proceedings of the Artificial Intelligence Algorithms and Applications: 11th International Symposium, ISICA 2019, Guangzhou, China, 16–17 November 2019*; Springer: Singapore, 2020; pp. 242–256.

10.   Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.

11.   Rios, A.; Itti, L. Closed-loop memory GAN for continual learning. *arXiv* **2018**, arXiv:1811.01146.

12.   Available online: https://github.com/ANI717/Anime_Face_Generation_with_GANs (accessed on 14 July 2023).

13.   Li, B.; Zhu, Y.; Wang, Y.; Lin, C.; Ghanem, B.; Shen, L. AniGAN: Style-guided generative adversarial networks for unsupervised anime face generation. *IEEE Trans. Multimed.* **2021**, *24*, 4077–4091. [CrossRef]

14.   Hamada, K.; Tachibana, K.; Li, T.; Honda, H.; Uchida, Y. Full-body High-resolution Anime Generation with Progressive Structure-conditional Generative Adversarial Networks. In *Proceedings of the European Conference on Computer Vision (ECCV) 2018 Workshops, Munich, Germany, 8–14 September 2018*; Springer: Cham, Switzerland, 2019.

15.   Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 4401–4410.

16.   Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.

17.   Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2223–2232.

18.   Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.

19.   Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein Generative Adversarial Networks. In Proceedings of the International Conference on Machine Learning (PMLR), Sydney, Australia, 6–11 August 2017; pp. 214–223.

20.   Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.C.; Bengio, Y. Generative adversarial nets. In Proceedings of the 28th Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.

21.   Odena, A. Open Questions about Generative Adversarial Networks. Distill. 2019. Available online: https://distill.pub/2019/gan-open-problems (accessed on 14 July 2023).

22.   Salimans, T.; Zhang, H.; Radford, A.; Metaxas, D.N. Improving GANs using optimal transport. In Proceedings of the 6th International Conference on Learning Representations, ICLR 2018—Conference Track Proceedings, Vancouver, BC, Canada, 30 April–3 May 2018.

23.   Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral normalization for generative adversarial networks. *arXiv* **2018**, arXiv:1802.05957.

24.   French, R.M. Catastrophic forgetting in connectionist networks. *Trends Cogn. Sci.* **1999**, *3*, 128–135. [CrossRef] [PubMed]

25.   Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A.A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A.; et al. Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 3521–3526. [CrossRef] [PubMed]

26.   Wu, C.; Herranz, L.; Liu, X.; van de Weijer, J.; Raducanu, B. Memory replay gans: Learning to generate new categories without forgetting. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 5962–5972.

27.   Cong, Y.; Zhao, M.; Li, J.; Wang, S.; Carin, L. Gan memory with no forgetting. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 16481–16494.

28.   Varshney, S.; Verma, V.K.; Srijith, P.K.; Carin, L.; Rai, P. CAM-GAN: Continual adaptation modules for generative adversarial networks. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 15175–15187.

29.   Yang, X.; Che, H.; Leung, M.F.; Liu, C. Adaptive graph nonnegative matrix factorization with the self-paced regularization. *Appl. Intell.* **2023**, *53*, 15818–15835. [CrossRef]

30.   Pan, B.; Li, C.; Che, H.; Leung, M.F.; Yu, K. Low-Rank Tensor Regularized Graph Fuzzy Learning for Multi-View Data Processing. *IEEE Trans. Consum. Electron.* **2023**. [CrossRef]