IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 January 2024, accepted 17 February 2024, date of publication 23 February 2024, date of current version 14 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3369234

RESEARCH ARTICLE

Anemone Image Generation Based on Diffusion-Stylegan2

HUIYING ZHANG, FEIFAN YAO, YIFEI GONG, AND QINGHUA ZHANG
College of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin City, Jilin 132022, China

Corresponding author: Huiying Zhang (yingzi1313@163.com)

This work was supported by the Science and Technology Development Plan of Jilin Provincial Department of Science and Technology under Grant 20220508145RC.

ABSTRACT Given the complexity and uncertainty of the underwater environment, it is of great importance to generate realistic and high-quality images. In this paper, we propose six unconditional generative models based on the Diffusion-Styegan2 generative model, incorporating Wasserstein, R2 regularization terms, and other techniques for anemone image generation. The Wasserstein distance technique is used in the loss part of Diffusion-Styegan2, combined with the back propagation algorithm to compute the gradient in the neural network while retaining the computational map to improve the training efficiency and training stability; the R2 regularization term is used to introduce the r2 hyperparameter, and the L2 regularization technique is used based on the original R1 regularization term to regularize the gradient of the discriminator to improve the training and generation performance of the model; the ADA technique is used based on DWBG-Stylegan2 to further improve the quality and stability of the generated images. In addition, a set of SA datasets (sea anemone datasets) with a resolution of 256*256 is proposed in this paper. The experimental results show that the FID value of Diffusuion-Stylegan2 is 10.31, the value of FID of DWBG-Stylegan2 is 8.32, the value of FID of Diffusion-Stylegan2-R2 is 9.58, and the optimal FID value of this experiment is achieved by DWBG-Stylegan2-ADA with a value of 5.67, which is considerably lower compared to the FID value of Diffusion-Stylegan2. Therefore, techniques such as Wasserstein and R2 regularization terms can effectively generate more realistic images of anemones. Meanwhile, this experiment provides new ideas and methods for the construction of the unconditional generative model.

INDEX TERMS Diffusion-Stylegan2, Wasserstein, R2 regularization term, SA datasets.

I. INTRODUCTION

The study of underwater imagery has gone through several stages of development, from early photographic techniques to modern digital underwater photography and advanced imaging techniques, as well as the emergence of deep-sea exploration and underwater robotics [1]. These advances have had a profound impact on the fields of marine science, environmental protection, resource exploration, and cultural heritage preservation.

Currently, the main research methods for generating marine biology images include: physical model-based methods, image synthesis-based methods, neural network-based methods, and meta-learning [2] based methods. Researchers mainly use GAN (Generative Adversarial Network) to generate marine biology images. Among them, GAN models including Cyclegan [3], Pix2Pix [4], Stargan [5], and Stylegan [6] are widely used in underwater image generation tasks. In addition, some researchers have tried to combine deep learning models with traditional image processing techniques to improve the quality of marine biology images. However, in practice, GAN models often suffer from problems such as non-convergence, unstable training, and pattern collapse. Therefore, a wide range of analyses and improvements have been proposed for GAN, including improving the network architecture, changing the objective function, regularizing the weights or gradients, etc.

GAN and Diffusion are two highly regarded directions in the field of deep learning, representing the frontiers of generative and probabilistic modeling respectively. Combining the

The associate editor coordinating the review of this manuscript and approving it for publication was Jiju Poovvancheri.

two creates a new approach to generative modeling that opens up new possibilities for image generation. The combination of GAN and Diffusion overcomes the limitations of each, and this model is often referred to as Diffusion-GAN (Dif-GAN) or Diffusion-Stylegan. This model utilizes the stability and sampling performance of Diffusion and the generative power of GAN to produce realistic and diverse images.

In this paper, we try to improve the training and generation results of the model by employing the loss function or regularization term in the loss part of Diffusion-Stylegan2 [7]. Employing the Wasserstein loss function in the loss part of the discriminator of Diffusion-Stylegan2 is used to measure the gap between the generated fake image and the real image, thus helping the discriminator to better distinguish between them. Firstly, the logits of the generated fake image and the logits of the real image (from the training datasets) are obtained by using the generator to generate the fake image and the logits of the real image. These logits represent the scores of the fake image and the real image on the discriminator respectively. Then, the Wasserstein distance is calculated by computing the difference in expected value between the two distributions. Here, the Wasserstein distance is obtained by calculating the difference between the expected values of the two logit distributions. Finally, to compute multiple loss terms (e.g. ''loss_Gmain'', ''loss_Gpl'', ''loss_Dgen'', ''loss_Dr1'') gradients, in the back propagation, the relevant nodes in the computational graph are retained so that the parameters of the model can be correctly updated in subsequent steps. In another model, this paper employs the R2 regularization term in the discriminator loss, which changes the training and generation behavior of the original model. The R2 regularization term applies the technique of the L2 regularization term [8], and also penalizes the gradient, which is a sum of squares of the weights, usually multiplied by a hyperparameter known as the regularization factor. At the same time, the R2 regularization term introduces the r2 hyperparameter, which together with the R1 regularization term and the L2 paradigm acts on the discriminator loss part of the model. The R2 regularization term helps to balance the weights to mitigate the gradient explosion problem, promotes the smoothing of the gradient, improves the model's generalization ability, and effectively controls the training process to ensure the stability of the model.

This paper makes the following contributions: 1) six unconditional generative models are proposed based on the Diffusion-Stylegan2 generative model, which are WBG-Stylegan2, DWBG-Stylegan2, Diffusion-Stylegan2-R2, DWBG-Stylegan2-ADA, DWBG-Stylegan2-R2, DWBG-Stylegan2-GP; 2) a large number of experimental results show that the DWBG-Stylegan2 model is more stable in training and produces higher quality images than the Diffusion-Stylegan2 model, and is, therefore, more suitable for underwater work; the Diffusion-Stylegan2-R2 model in a large extent improves the generalization ability of the model and the quality of the generated images is closer to the real images; the DWBG-Stylegan2-ADA model greatly improves the image generation quality of the generated model after adopting the ADA technique; 3) given the scarcity of the marine single-organism datasets, 9235 SA datasets with a resolution of 256*256 are provided.

II. RELATED WORKS

A. COMBINATION OF GAN AND DIFFUSION

A simple technique to stabilize GAN training is to inject suitable instance noise into it. This approach extends the support of the generator and discriminator distributions and prevents overfitting of the discriminator. However, finding suitable noise distributions is difficult. In 2017, Mescheder et al. showed that adding instance noise to high-dimensional discriminators is ineffective and proposed to approximate it by adding a zero-centered gradient penalty to the discriminator [9]. In 2018, Mescheder et al. proved both theoretically and empirically that this approach is convergent. They also showed that adding a zero-centered gradient penalty to unsaturated GANs leads to stable training and better generation quality compared to WGAN-GP [10]. In the same year, NVIDIA released Stylegan for the first time, a model that introduced several innovations, including AdaIN (Adaptive Instance Normalization) [11] and a style transfer mechanism [12]. In 2019, to address the problem of feature artifacts in Stylegan-generated images, NVIDIA used a PLR (Path Length Regularization) new training technique and proposed Stylegan2. In 2020, Ho et al. proposed the DDPM (Denoising Diffusion Probabilistic Models) [13], and the excellent results of DDPM on image generation and denoising in the Deep Learning field caused an extensive research boom. In 2022, Zhendong Wang et al. used a diffusion process to generate Gaussian mixed distributed instance noise and proposed the Diffusion-Stylegan2 model.

The noise of Diffusion-Stylegan2 is obtained by sampling a Gaussian mixture distribution over the diffusion step with the help of a temporal correlation discriminator. The use of Gaussian mixture distribution sampling provides two benefits: firstly, it can stabilize the training by mitigating the gradient vanishing problem, which occurs when the data and generator distributions are too different; secondly, the data can be augmented by creating different noisy versions of the same image, which improves the data efficiency and the diversity of the generators.

B. ANALYSIS OF PRIOR APPROACHES

In recent years, the research on image generation methods has been very hot, this paper focuses on several newer image generation methods for GAN, and summarizes the advantages and disadvantages of the models used in various generation methods, Table 1 gives the advantages and disadvantages of several comparison models used in this paper.

TABLE 1. Advantages and disadvantages analysis of previous methods.

Models	Advantages	Disadvantages
GAN	Laying the groundwork for the development of the AIGC field	Unstable and prone to mode collapse
Diffusion	Greater interpretability	Iterative sampling is slow
Stylegan2	Optimized for artifact problems and texture adhesion effect in Stylegan	High demand for computing resources
Stylegan2-ADA	Dynamically adjusting the data augmentation of the discriminator during the training process	Increased difficulty and complexity of training
Diffusion-Stylegan2	Combining the stepwise noise generation process of the Diffusion model and the powerful feature control of Stylegan2	Increased complexity and inadequate integration of the two mechanisms

C. WASSERSTEIN DISTANCE AND R2 REGULARIZATION TECHNIQUES

In 2017, Martin Arjovsky, Soumith Chintala, and Leon Bottou first proposed the concept of ''Wasserstein GAN'', which introduces the Wasserstein distance into GAN models to measure the difference between the generators and discriminators, as an alternative to the JS dispersion or KL dispersion used in traditional GANs. The key idea of WGAN [14] is to measure the distributional difference between the generators and discriminators by the Wasserstein distance, which leads to more stable training and mitigates the problem of pattern collapse when training GANs, and this approach is an important advancement. Subsequently, many researchers have proposed improved models such as WGAN-GP based on WGAN, introducing asymptotic penalties to promote gradient smoothness and improve the convergence of training.

In 1943, Andrey Tikhonov et al. first proposed L2 regularization technology, and it has been widely used in statistics and linear regression to control the complexity of the model parameters. In 1970, Rokem and Kay applied L2 regularization technology to ridge regression [15] to solve the problem of multicollinearity [16]. In 1992, David H. Wolpert et al. mentioned L2 regularization in the paper ''Stacked Generalization''. This technique is widely used in machine learning and deep learning to reduce the risk of model overfitting. With the rise of deep learning, L2 regularization has been widely adopted to stabilize the training of neural networks.

III. UNCONDITIONAL IMAGE GENERATION MODEL

A. ARCHITECTURE ANALYSIS OF DIFFUSION-STYLEGAN2

Diffusion-Stylegan2 is a unique generative model for GANs that uses a forward diffusion chain to generate Gaussian mixed distributed instance noise, allowing for a better combination of the Diffusion Model and GAN.

1) INJECTION OF INSTANCE NOISE

Diffusion models are latent variable models that use Markov chains [17] mapped to a latent space. Fundamentally, the diffusion model works by reducing the quality of the training data by continuously adding Gaussian mixture distributed noise in iterations, and then recovering the data by inverse denoising, the diffusion process of the model is shown in Fig.1.

To make the generator more robust and diverse, Gaussian mixture distributed instance noise is injected into the generated sample x_g at each step of the diffusion process. Starting from the original sample x, Gaussian noise z is continuously added to the original image, gradually erasing its information until an isotropic normal distribution N(0,1) is reached after T steps. The mixing distribution q(y|x) models the noise samples acquired at any step of the diffusion process, with a mixing weight π_t at each step t. The mixing component q(y|x,t) is Gaussian distributed with a mean proportional to x and a variance that depends on the noise level at step t. The diffusion-induced mixing distribution can be expressed as follows:

$$x \sim p(x), y \sim q(y|x), q(y|x) := \sum_{t=1}^{T} \pi_t q(y|x,t) \tag{1}$$

$$x_g \sim p_g(x), y_g \sim q_g(y|x), q(y_g|x_g) := \sum_{t=1}^{T} \pi_t q(y_g|x_g,t) \tag{2}$$

where q(y|x) is a T-component mixture distribution and the mixture weights π_t are non-negative and sum to 1. The mixture components q(y|x,t) are obtained by diffusion and are denoted as:

$$q(y|x,t) = \mathcal{N}(y;\sqrt{\overline{\alpha}_t}x, (1-\overline{\alpha}_t)\sigma^2 I) \tag{3}$$

Samples of this mixture can be plotted as $t \sim p_\pi :=$ Discrete$(\pi_1, \pi_2, \ldots, \pi_T)$, $y \sim q(y|x,t)$, one can sample y from this mixture distribution to obtain real and generated samples with different noise levels.

2) GENERATE TRAINING

Diffusion-Stylegan2 trains its generator and discriminator by solving the min-max game, which is given in Eq.4 after adding Gaussian mixed distributed instance noise to the original GAN:

$$\begin{aligned} V(G,D) = {} & \mathbb{E}_{x\sim p(x), t\sim p_\pi, y\sim q(y|x,t)}[\log(D_\phi(y,t))] \\ & + \mathbb{E}_{z\sim p(z), t\sim p_\pi, y_g\sim q(y|G_\theta(z),t)}[\log(1-D_\phi(y_g,t))] \end{aligned} \tag{4}$$

For any diffusion step t, the objective function in Eq.4 encourages the discriminator to assign a high probability to the perturbed real data and a low probability to the perturbed generated data. On the other hand, the generator tries to generate samples that can deceive the discriminator at any diffusion step t.

3) ADJUSTMENT OF DIFFUSION INTENSITY

The discriminator is optimized by the strength of the diffusion process, adding noise to y and yg. When the diffusion step size t is large, the noise-to-data ratio is higher and the

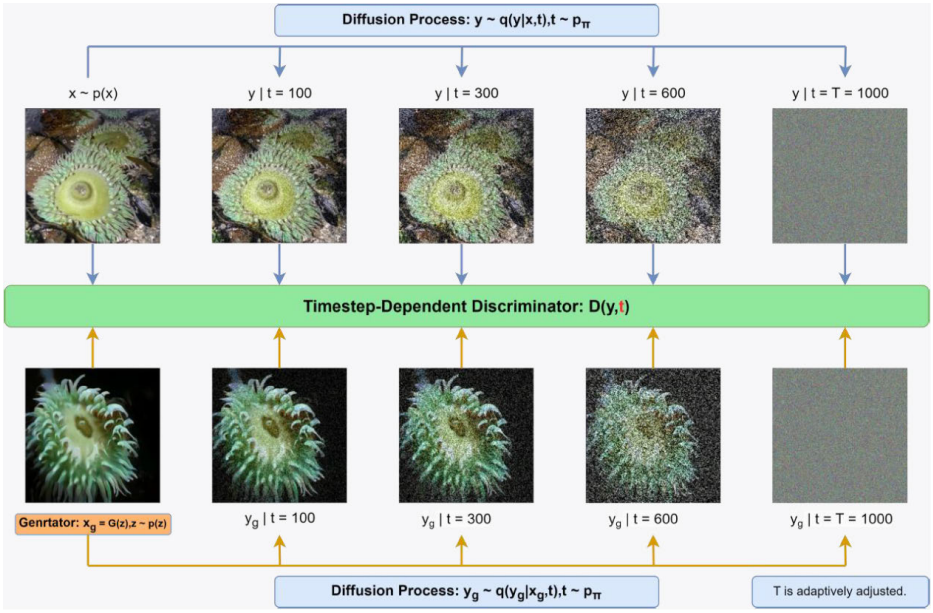


FIGURE 1. The flowchart of Diffusion Stylegan2. The top row image represents the forward diffusion process of the real image, while the bottom row image represents the forward diffusion process of the generated fake image.

likelihood of completing the task is lower. Use $1 - \overline{\alpha}_t$ to measure the diffusion strength, which increases as t grows. To control the diffusion strength, the maximum number of steps T is adaptively modified. The discriminator first learns from the original data samples, and then progressively increases the difficulty by providing it with samples of larger t. For this purpose, a schedule is customized for T, which depends on the discriminator and a metric rd for the degree of data overfitting:

$$r_d = \mathbb{E}_{y,t\sim p(y,t)}[sign(D_\phi(y,t) - 0.5)], T = T + sign(r_d - d_{target}) * C \quad (5)$$

Every four small batches rd is computed and T is updated, with two options for the distribution p_π used to sample t in the diffusion process:

$$t \sim p_\pi := \{^{uniform:Discretre(\frac{1}{T},\frac{1}{T},...,\frac{1}{T})}_{priority:Discretre(\frac{1}{\sum_{t=1}^{T}t},\frac{1}{\sum_{t=1}^{T}t},...,\frac{1}{\sum_{t=1}^{T}t})} \quad (6)$$

The ''priority'' option gives more weight to larger t, which means that the discriminator will see more new samples from new steps as T increases.

B. ARCHITECTURAL DESIGN OF DWBG-STYLEGAN2

1) INTRODUCING WASSERSTEIN AND PRESERVING COMPUTATIONAL GRAPHS

The architecture of DWBG-Stylegan2 is based on the architecture of Diffusion-Stylegan2 using the Wasserstein distance loss function, while the computational graph is retained to perform additional back propagation calculations for multiple gradient calculations. DWBG-Stylegan2 is trained several times on the SA datasets to obtain the optimal pre-trained model, the structure of which is shown in Fig.2.

In a Generator, Diffusion is used to improve the quality and stability of the generated image. The main architecture of the Generator is Stylegan2, in this part of the architecture combined with Diffusion, further noise addition and denoising of the generated image is carried out to get the generated image of higher quality. The architectural diagram of the Generator is shown in Fig.3.

In particular, the Mapping network serves to perform feature untangling, consists of 8 fully connected layers, and has the same size of output and input layers. The middle part is the architecture of Stylegan2, which removes some redundant operations from Stylegan's architecture, moves the addition of b and B outside the active region of style, and adjusts only the standard deviation of each feature mapping.

If KL divergence and JS divergence are used as a measure of the difference between the two probabilities, the most critical point is that if the support sets of the two probabilities do not overlap, it is impossible to make that parameterized, moveable probability distribution slowly move over to fit the target distribution. So in this paper, we use Wasserstein distance as a loss function in Loss, and also as a measure of the distance between two probability distributions, to calculate the difference between the generated data and the real data, which is defined as follows:

$$W(\mathbb{P}_r, \mathbb{P}_\theta) = \sup_{||f||_L \leq 1} \mathbb{E}_{x\sim\mathbb{P}_r}[f(x)] - \mathbb{E}_{x\sim\mathbb{P}_\theta}[f(x)] \quad (7)$$

where $W(\mathbb{P}_r, \mathbb{P}_\theta)$ is the probability distribution, $\mathbb{P}_r$ and $\mathbb{P}_\theta$ denote the distributions of the real data and the generative model respectively, which represent the lower bound on the

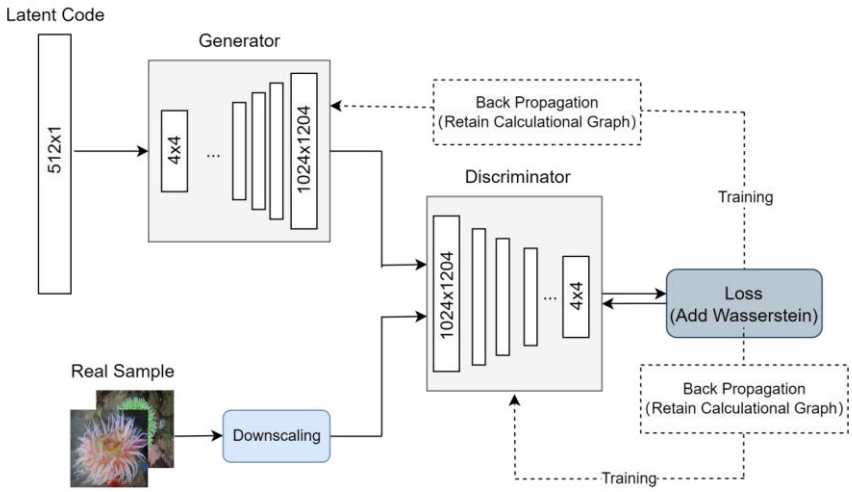


FIGURE 2. Generating adversarial flowchart for DWBG-Stylegan2.

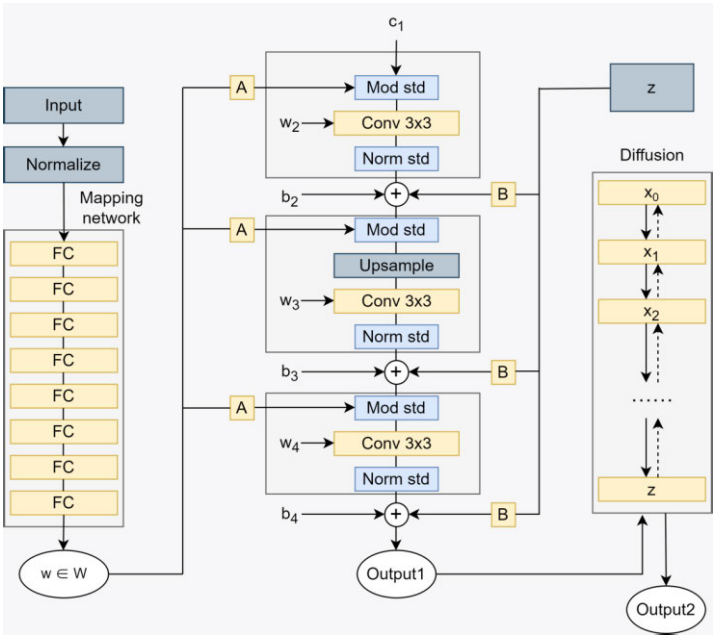


FIGURE 3. The generator structure of DWBG-Stylegan2.

mean of the difference in the paradigms of two random variables x, y on the same space. f(x) is a function in the set of functions $||f||_L \leq 1$, $||f||_L \leq 1$ denoting the set of functions that satisfy the 1-Lipschitz condition.

2) BASIC PRINCIPLES AND TECHNICAL EXPLANATIONS

Traditional GANs use cross-entropy losses or mean-square error losses, which can lead to unstable training, especially when there is an imbalance in power between the generator and the discriminator. In Stylegan2 combines Diffusion's reconstruction loss and latent loss, the reconstruction loss ensures that the denoising process recovers a high-quality image, and the latent loss encourages the generated samples to follow a simple prior distribution, these losses add a certain amount of training stability, especially when dealing with complex and high-resolution images. The Wasserstein loss function, on the other hand, serves to provide a smoother gradient, which helps to reduce oscillations in the training process, thus making training more stable. In addition, using the Wasserstein loss function, the generator can produce samples that are closer to the true data distribution. This is because the Wasserstein distance directly measures the "distance" between the generated samples and the real samples, thus providing a more direct optimization objective.

The computational map is retained for the fact that in some cases, calculating the Wasserstein distance requires

calculating the cost of pairing between the generated and real samples. Preserving the computational map can help calculate these costs more efficiently, thus speeding up the training process.

C. ARCHITECTURAL DESIGN OF DIFFUSION-STYLEGAN2-R2

1) CALCULATION OF THE L2 PARADIGM

The L2 regularized loss function is based on the original loss function with an L2 paradigm penalty term for the weight parameter, and the L2 regularized loss function and weight update can be expressed as:

$$\widehat{J}(w) = J(w) + \lambda ||w||_2^2 = J(w) + \frac{\alpha}{2} w^T w \tag{8}$$

$$w = (1 - \eta \cdot \alpha) w - \eta \cdot \nabla_w J(w) \tag{9}$$

where $J(w)$ is the original loss function, λ is the regularization strength hyperparameter, w is the weight parameter, $\nabla_w J(w)$ is the gradient of the original loss function, and η is the learning rate. Compared to the weight update of the original loss function, the difference between the weight update with the addition of L2 regularization and the weight update of the original loss function is only the addition of the coefficient of $1 - \eta \cdot \alpha$ in front of w. When $0 < 1 - \eta \cdot \alpha < 1$, the w weights are scaled down for each weight update, i.e., for each learning of the gradient descent method.

To find the expression between the maximum value w_i for which L2 regularization is added and the maximum value w_i^* for the original loss function. The L2 regularization is added to the original loss function to give a Taylor expansion, preserving only the quadratic terms, i.e., the quadratic approximation expression:

$$\widehat{J}(w) \approx J(w^*) + \nabla_w J(w^*)(w - w^*) + \frac{1}{2}(w - w^*)^T H(w - w^*) + \frac{\alpha}{2} w^T w \tag{10}$$

Simplifying and deriving the above equation yields the gradient of the loss function with the addition of L2 regularization as:

$$\nabla_w \widehat{J}(w) = H(w - w^*) + \alpha \cdot w \tag{11}$$

When $w = w$, find w and convert to the form of $H = Q \Lambda Q^T$ to get:

$$w = Q(\Lambda + \alpha \cdot I)^{-1} \Lambda Q^T \cdot w^* \tag{12}$$

Get the expression between w_i and $w_i{}^*$:

$$w_i = \frac{\lambda_i}{\lambda_i + \alpha} \cdot w_i^* \tag{13}$$

From the above equation, it follows that when $\alpha = 0$, the two are equal; when $\alpha > 0$, it corresponds to a magnification of $w_i{}^*$; and when $\alpha > 0$, it corresponds to a reduction of $w_i{}^*$.

2) ANALYSIS OF THE R2 REGULARIZATION TERM

To reduce the overfitting phenomenon of the generated image, this paper proposes the R2 regularization term applied in the discriminator loss, which retains the R1 regularization term in Diffusion-Stylegan2 and introduces the new r2 hyperparameter. The structure of the Diffusion-Stylegan2-R2 is shown in Fig.4.

The computation of the R2 regularization term added to the discriminator loss consists of: 1. computing the sum of squares of the gradients: in the loss section, "r1_grads" is used to denote the gradients, "r1_grads.squares()" compute the squares of the gradients for each parameter, and then ".sum([1], [2], [3])" is used to sum the squares of all the parameters; 2. Calculate the L2 parameter: the L2 parameter of the sum of the squares of the gradients is calculated as part of the regularization term. In discriminator loss, this is calculated by "r2_penalty=(r1_grads**2).sum([1], [2], [3])", where "r1_grads**2" represents the square of the gradient, and then L2 norms for each parameter are calculated using ".sum([1], [2], [3])". Taken together, the formula for R2 regularization is:

$$\widehat{J}(w) = J_{r1}(w) + \frac{\gamma_{r2}}{2m} ||w||_2^2, \text{ where } w = \nabla_x D_{real}(x) \tag{14}$$

The above equation can be noted as:

$$J = J_0 + \frac{\gamma_{r2}}{2m} ||\nabla_x D_{real}(x)||_2^2 \tag{15}$$

where γ_{r2} are hyperparameters used to control the strength of the R2 regularization terms, $||\nabla_x D_{real}(x)||$ denotes the L2 paradigm of the gradient of the discriminator for the real image, and m denotes the batch size. This formula describes the R2 regularization term, which is used to smooth the gradient, prevent certain weights from being too large, and improve the generalization of the model.

3) BASIC PRINCIPLES AND TECHNICAL EXPLANATIONS

The R2 regularization term is a technique used to prevent neural network overfitting by adding a penalty term to the loss function to discourage large weights, which can improve the training stability and overall quality of generative models such as Stylegan2. By penalizing large gradients and encouraging smoother decision boundaries, the R2 regularization term helps prevent overfitting and promotes the generation of high-quality, realistic images.

D. ADAPTIVE DISCRIMINATOR ENHANCEMENT TECHNOLOGY

Adaptive Discriminator Augmentation (ADA) technique [18] can be used on top of DWBG-Stylegan2 to get the best generated images, the architecture of DWBG-Stylegan2 is described in detail in B. Based on the PCR (balanced consistency regularization) method, Stylegan2-ADA does image enhancement for the discriminator along with image enhancement for the generator and removes the consistency regularity term which is enhanced in the loss. The network architecture of DWBG-Stylegan2-ADA is shown in Fig.5.

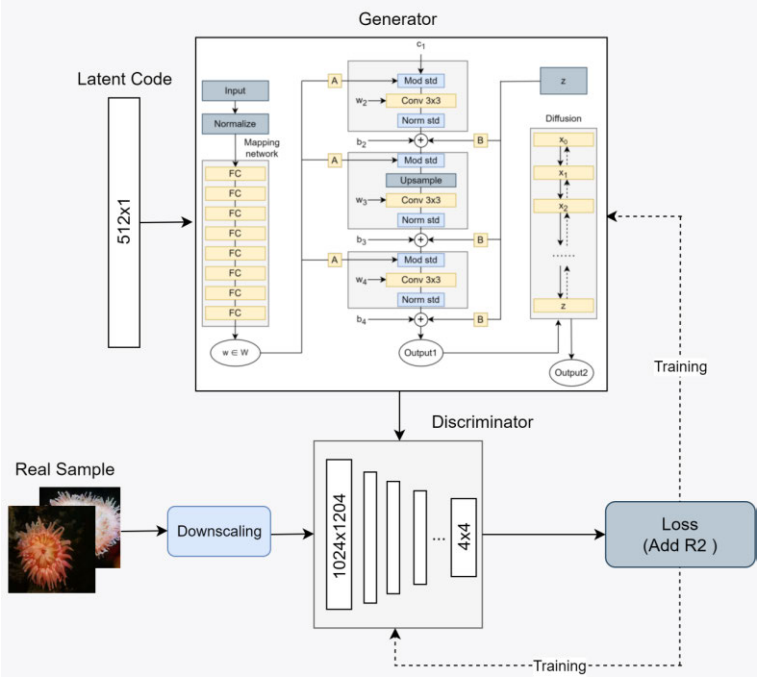


FIGURE 4. The generative adversarial flowchart of Diffusion-Stylegan2-R2.

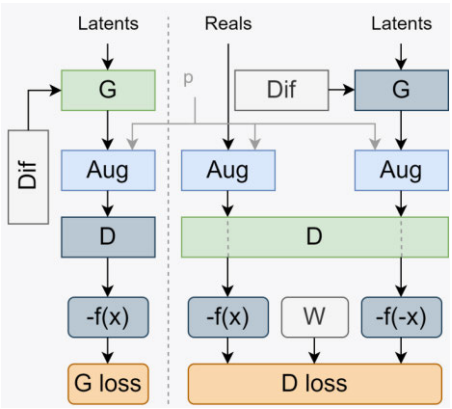


FIGURE 5. Network structure diagram of DWBG-Stylegan2-ADA.

ADA solves the problem mainly by getting the probability p of data enhancement through network adaptation, and deciding whether to do image enhancement on the image or not with the probability p. Here p is generally a human-set hyperparameter, and its value can have a great impact on the generated results. In practice, the weights of the samples or the training difficulty of the discriminator can be adjusted manually, or the parameters can be adjusted automatically by some automatic adjustment methods such as gradient descent and Adam.

In a word, the training stability of DWBG-Stylegan2 proposed in this paper far exceeds that of Diffusion-Stylegan2, has more reasonable gradient flow, generates higher-quality anemone images, and this model applies to a wider range of tasks and datasets; the Diffusion-Stylegan2-R2 proposed in this paper introduces the R2 module, which is a new module combining the R1 regularization term and the L2 paradigm in the loss function of Diffusion-Stylegan2, which generates images of higher quality than Diffusion-Stylegan2, with better training stability, and at the same time, it can accelerate the convergence speed of the training; with the introduction of the ADA, DWBG-Stylegan2-ADA can generate the highest quality anemone images among all comparison models.

IV. SEA ANEMONE DATASETS

The datasets of underwater images are more difficult to obtain than other datasets, so the research on underwater images is largely limited by the problems of insufficient basic information, imperfect equipment, etc. In 2020, Md Jahidul Islam et al. publicly released the EUVP [19] datasets on the web through marine field data collection and related information search, which provided relatively abundant large-scale marine underwater image datasets for the subsequent underwater image researchers. Other underwater image datasets include SQUID [20], LSUI [21], UIEB [22], etc. However, datasets about marine underwater single organisms are very scarce, and the research in their related directions is largely limited. Therefore, in this paper, 9235 SA datasets with a resolution of 256*256 were collected and produced, and some examples are provided in Fig.6.

The SA datasets contains a large number of single-individual biological images of anemones with good perceptual quality, which are mainly derived from the EUVP datasets, Flickr, and YouTube. High-quality images after data

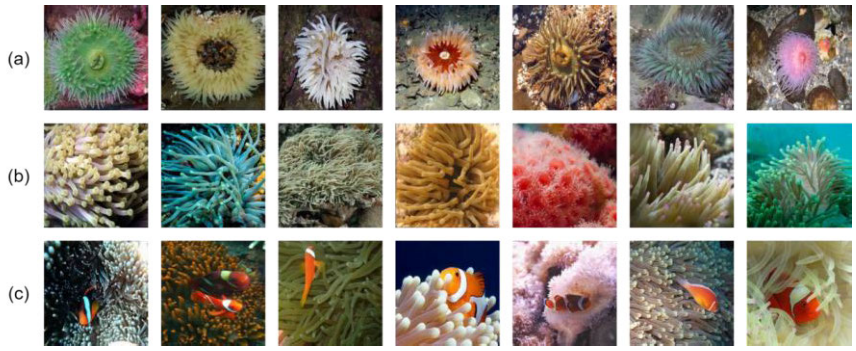


FIGURE 6. Displayed some example images from the SA datasets: (a). Individual small sea anemone images; (b). Large-scale images of sea anemones and their communities; (c). Images of sea anemones and clownfish in symbiosis.

augmentation by the FUnIE-GAN [19] model were selected one by one in the paired and unpaired datasets of the publicly released EUVP datasets. Meanwhile, on Flickr and YouTube, this paper collects a large number of better-quality anemone images publicly posted by users and bloggers. Through the image cropping technique, 9235 anemone images were uniformly cropped to a resolution size of 256*256 to facilitate the experimental operation.

Compared with existing underwater image datasets, the anemone image dataset provided in this paper belongs to a medium-sized single-individual biological image dataset, which is suitable for unconditional image generation, and is the first medium-sized anemone dataset for underwater image generation. The dataset contains multiple species of large and small anemones, and at the same time, the use of anemone images with a resolution of 256*256 can effectively shorten the training time and reduce the training cost of the model. However, since anemones and clownfish are symbiotic organisms in marine ecosystems, clownfish may be present in some anemone images. In this paper, we have collected this dataset of marine underwater single-organism images and minimized the impact of clownfish on image generation as much as possible. There may be some shortcomings in the dataset that subsequent researchers will hopefully be able to remedy.

V. EXPERIMENTAL RESULTS AND ANALYSIS

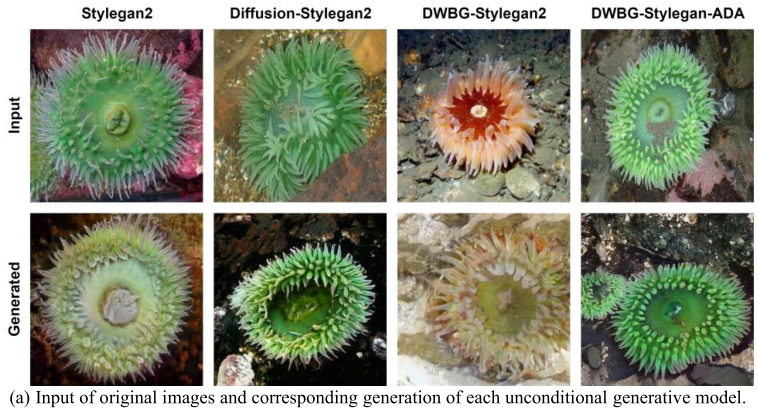
The Pytorch library was used to implement the construction of the Diffusion-Stylegan2 model and trained on 9235 SA datasets with a resolution of 256*256. The training was performed using an NVIDIA GeForce GTX 4090 graphics card, all models were trained with 8K iterations, the cfg was auto, the batch was 16, the lr was 0.0025, the target was 0.6, and the optimizer was Adam.

A. QUALITIVE EVALUATION

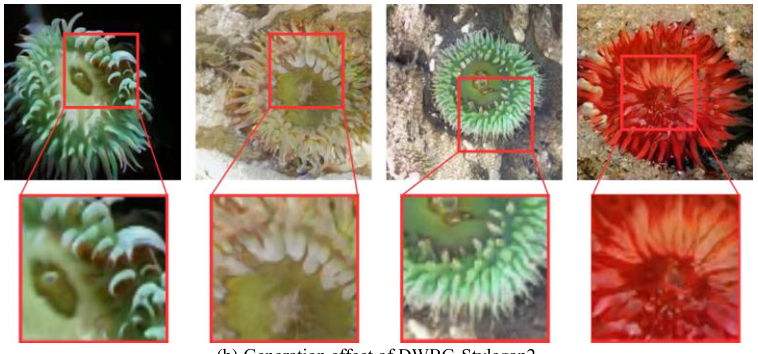
The similarity between the images generated by Diffusion-Stylegan2 and real samples is first qualitatively analyzed to measure the performance of the generated models. At the same time, several models proposed in this paper are pre-trained for model generation and the images with relatively good generation results are demonstrated. Some examples of generation are given in Fig.7.

As shown in Fig.7(a), the original anemone image with a resolution of 256*256 is input, and the anemone image is output through each generative model respectively. Stylegan2 [23] and Diffusion-Stylegan2 are the original models, the DWBG-Stylegan2 model is the model with the best generation in this paper, and the last model is based on DWBG-Stylegan2 using ADA. As can be seen, DWBG-Stylegan2 and DWBG-Stylegan2-ADA have significantly improved the distortion of the image over Stylegan2 and Diffusion-Stylegan2, and the backgrounds are generated with more diversity than the original model. In addition, as shown in Fig.7(b), (c), and (d), the image generation effects of DWBG-Stylegan2, Diffusion-Stylegan2-R2, and DWBG-Stylegan2-ADA are demonstrated, where the DWBG-Stylegan2-ADA has been largely close to the real image in generating anemone images.

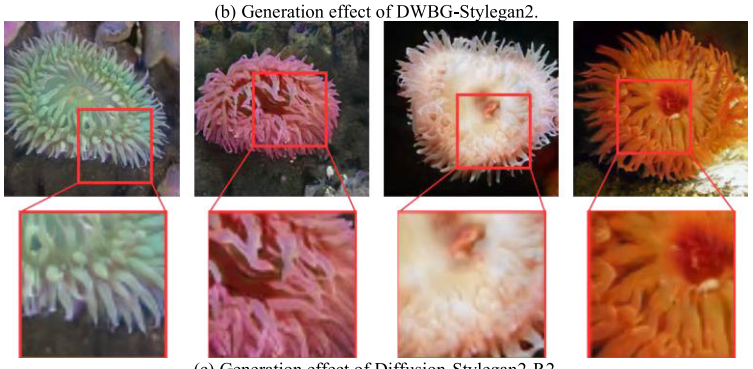
This paper provides a qualitative comparison of Diffusion-Stylegan2 and related models. Eleven deep learning-based generative models are considered:(I) GAN [24]; (II) Stylegan2; (III) Diffusion-Stylegan2; (IV) Stylegan2 with ADA (Stylegan2-ADA) [25]; (V) Stylegan2 with Wasserstein and back propagation preserving computational maps (WBG-Stylegan2); (VI) Diffusion-Stylegan2 with Wasserstein and back propagation preserving computational maps (DWBG-Stylegan2); (VII) Diffusion-Stylegan2 with the addition of R2 (Diffusion-Stylegan2-R2); (VIII) Diffusion-Stylegan2 with ADA (Diffusion-Stylegan2-ADA); (IX) Diffusion-Stylegan2 with ADA, Wasserstein, and back propagation preserving computational maps (DWBG-Stylegan2-ADA); (X) Diffusion-Stylegan2 with R2, Wasserstein and back propagation preserving computational maps (DWBG-Stylegan2-R2); (XI) Diffusion-Stylegan2 with the addition of GP, Wasserstein and back propagation preserving computational maps (DWBG-Stylegan2-GP). These models use technical tools such as regularization and introduce the concept of style migration from Stylegan2 onwards. The training was performed on the SA datasets using the same setup (the aug



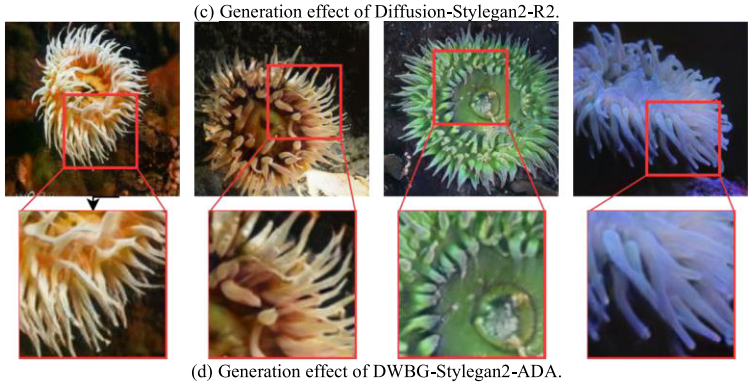
(a) Input of original images and corresponding generation of each unconditional generative model.



(b) Generation effect of DWBG-Stylegan2.



(c) Generation effect of Diffusion-Stylegan2-R2.



(d) Generation effect of DWBG-Stylegan2-ADA.

FIGURE 7. Demonstrates the generation effect of sea anemone images and the three improved models in this article.

needs to be added as ada for the ADA experimental part of the training). In addition, Diffusion-Stylegan2 was considered as a benchmark for comparing the performance of DWBG-Stylegan2 and Diffusion-Stylegan2-R2 and qualitatively evaluated using generated images with 1K resolution of 256*256. Fig.8 shows some example comparisons.

As shown in Fig.8, the generation effect of GAN is relatively limited in terms of the quality of the raw generated

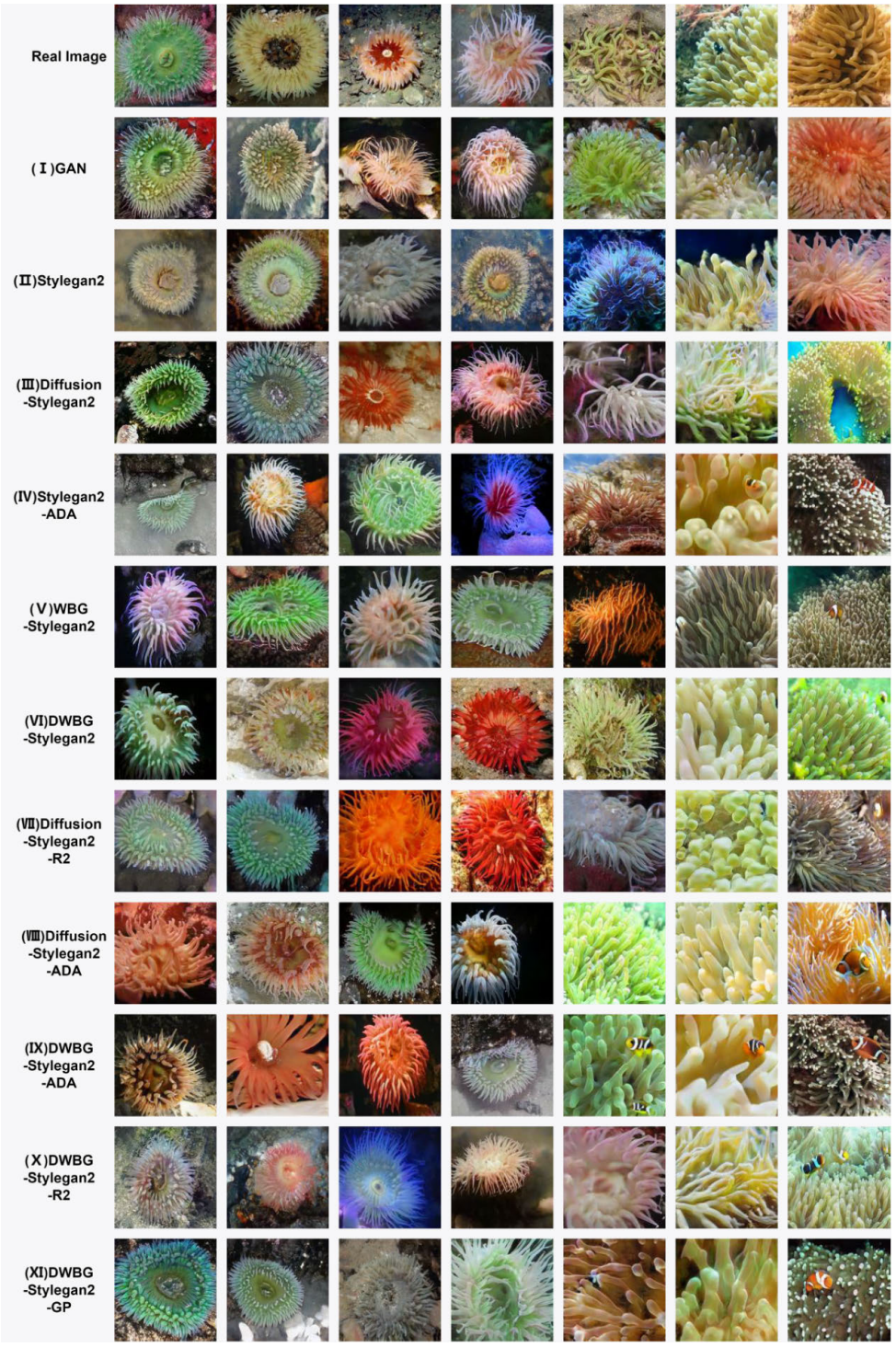


FIGURE 8. Qualitative performance comparison of generative models based on deep learning (250% zoom for optimal viewing effect).

images compared to several other models. Stylegan2 and WBG-Stylegan2 can generate roughly anemone images, but the distortion is high and the generation effect is relatively poor. However, it is a great breakthrough compared to GAN. Diffusion-Stylegan2 has improved the quality of anemone images compared to Stylegan2 and WBG-Stylegan2, the distortion is reduced, and the contours and details of the generated anemone images are improved. DWBG-Stylegan2-GP and DWBG-Stylegan2-R2, which are two models with a relatively complex structure, are more effective in terms of

the details and textures of the images compared to Stylegan2 and WBG-Stylegan2 are significantly improved, but there is no substantial improvement over Diffusion-Stylegan2. The DWBG-Stylegan2 proposed in this paper has much lower distortion and more visible details compared to the backbone model Diffusion-Stylegan2. Diffusion-Stylegan2-R2 has a smoother model compared to the backbone model Diffusion-Stylegan2, which makes the generated image more realistic and natural and the quality of the generated image is enhanced. From the data enhancement point of view, Stylegan2-ADA, Diffusion-Stylegan2-ADA, and DWBG-Stylegan2-ADA significantly outperform the other generated models in terms of clarity, diversity, and distortion. Especially, DWBG-Stylegan2 with ADA augmentation generates anemone images that are more realistic than real images in terms of color and texture. All in all, the performance of DWBG-Stylegan2-ADA has improved significantly in underwater image generation.

B. QUANTITATIVE EVALUATION

1) EVALUATION METRICS

In this paper, seven metrics are considered to measure the performance of the generative model, which are Frechet Inception Distance (FID) [26], Kernel Inception Distance (KID) [27], Perceptual Path Length (PPL) [28], Inception Score Mean (IS_mean), Inception Score Standard Deviation (IS_std), Precision (P) [29] and Recall (R) [11], to quantitatively compare the model performance.

FID indicates the difference between the distribution of the generated image and the real image, smaller FID value means higher quality of the produced image. The formula for FID is given below:

$$\mathrm{FID} = ||\mu_r - \mu_g||^2 + \mathrm{T_r}(\Sigma_r + \Sigma_g - 2\sqrt{\Sigma_r \Sigma_g}) \tag{16}$$

where μ_r denotes the feature mean of the real image, μ_g denotes the feature mean of the generated image, Σ_r denotes the covariance matrix of the real image, and Σ_g denotes the covariance matrix of the generated image.

KID measures the difference between two sets of samples by calculating the square of the maximum mean difference between the Inception representations. A lower KID value indicates a smaller difference between the distribution of the generated image and the distribution of the real image. The formula for KID is given below:

$$\mathrm{KID(x, y)} = ||\mathrm{m_x} - \mathrm{m_y}||^2 + \mathrm{T_r}(\mathrm{C_x} + \mathrm{C_y} - 2(\mathrm{C_x C_y})^{\frac{1}{2}}) \tag{17}$$

where $\mathrm{m_x}$ and $\mathrm{m_y}$ are the centers of the feature representations of the true image distribution x and the generated image distribution y, and $\mathrm{C_x}$, $\mathrm{C_y}$ are the covariance matrices of the feature representations of the true image distribution x and the generated image distribution y.

The idea of PPL is to give two random noises z_1 and z_2 to find the length of the perceptual paths of the two points, the idea of differentiation is used to subdivide the interpolated paths of the two noisy points into small segments, to find the length of each segment, and then averaged. A lower value of PPL indicates a smoother and more continuous transition between the generated images. The formula for PPL is given below:

$$\mathrm{PPL} = \mathbb{E}[\frac{1}{\epsilon^2} \mathrm{d}(\mathrm{G}(\mathrm{slerp}(z_1, z_2; t)), \mathrm{G}(\mathrm{slerp}(z_1, z_2; t + \epsilon)))] \tag{18}$$

where ϵ denotes subdivided segments, slerp denotes spherical linear interpolation, and t denotes interpolation parameters obeying a uniform distribution.

IS is a metric used to evaluate the quality of images generated by a generative model. IS_mean denotes the mean value of the Inception Score [30], this metric is usually used to measure the diversity of the generated images, and a higher mean value of IS implies that the generated images are more diverse. IS_std denotes the standard deviation of the Inception Score, a smaller value means that the generated image is more consistent and stable across multiple samples. The formula for IS is given below:

$$\mathrm{IS_{mean}(G)} = \exp(\mathbb{E}_x[\mathrm{D_{KL}}(\mathrm{p(y|x)||p(y)})]) \tag{19}$$

$$\mathrm{IS_{std}(G)} = \sqrt{\mathbb{E}_x[\mathrm{D_{KL}}(\mathrm{p(y|x)||p(y)})^2] - (\mathbb{E}_x[\mathrm{D_{KL}}(\mathrm{p(y|x)||p(y)})])^2} \tag{20}$$

where p(y|x) denotes the category probability distribution predicted by the Inception-v3 [31] network given an image x, and p(y) denotes the average category probability distribution over all images.

P and R are an evaluation metric based on the precision-recall curve used to assess the performance of a generative model. P is the proportion of the generated samples that match the real data. R is the proportion of the real data that is successfully matched. Higher precision and recall indicate that the generated samples are both diverse and similar to the real data. The formulas for P and R are given below:

$$P = \frac{TP}{TP + FP} \tag{21}$$

$$R = \frac{TP}{TP + FN} \tag{22}$$

where TP denotes the number of samples in which the survival samples match the real data, FP denotes the number of samples in which the generation samples do not match the real data and FN denotes the number of samples in the real data that have not been successfully matched.

2) ABLATION EXPERIMENT

In this paper, three sets of ablation experiments were performed:(1) Stylegan2, WBG-Stylegan2, Diffusion-Stylegan2, DWBG-Stylegan2, DWBG-Stylegan2-R2, DWBG-Stylegan2-GP; (2) Stylegan2, Diffusion-Stylegan2, Diffusion-Stylegan2-R2, DWBG-Stylegan2-R2; (3) Stylegan2, Stylegan2-ADA,

TABLE 2. Quantitative comparison of wasserstein distance series training on FID, KID, PPL, P and R metrics.

Basic	Wasserstein	Diffusion	R2	GP	FID↓	KID↓	PPL↓	P↑	R↑
✓					23.84	0.014694	40.8930	**0.5257**	0.0510
✓	✓				22.95	0.013619	**37.6626**	0.5005	0.0583
✓		✓			10.31	0.003873	56.9142	0.5140	0.2195
✓	✓	✓			**8.32**	**0.002270**	60.9655	0.5009	**0.2658**
✓	✓	✓	✓		10.32	0.004248	56.3985	0.5056	0.2009
✓	✓	✓		✓	10.16	0.003353	61.3546	0.5242	0.2244

TABLE 3. Quantitative comparison of R2 series training on FID, KID, P and R metrics.

Basic	Wasserstein	Diffusion	R2	FID↓	KID↓	PPL↓	P↑	R↑
✓				23.84	0.014694	**40.8930**	**0.5257**	0.0510
✓		✓		10.31	0.003873	56.9142	0.5140	0.2195
✓		✓	✓	**9.58**	**0.003216**	56.0007	0.4896	**0.2246**
✓	✓	✓	✓	10.32	0.004248	56.3985	0.5056	0.2009

TABLE 4. Quantitative comparison of ADA series training on FID, KID, IS_mean, IS_std, P and R metrics.

Basic	Wasserstein	Diffusion	ADA	FID↓	KID↓	IS_mean↑	IS_std↓	P↑	R↑
✓				23.84	0.014694	**2.5649**	0.014659	0.5257	0.0510
✓			✓	6.21	0.001416	2.1499	**0.014566**	0.5285	0.3631
✓		✓	✓	6.16	0.001283	2.1192	0.019430	**0.5404**	0.3610
✓	✓	✓	✓	**5.67**	**0.001073**	2.2287	0.025000	0.5377	**0.3793**

Diffusion-Stylegan2-ADA, DWBG-Stylegan2-ADA. The experimental results are shown in Tables 2, 3 and 4.

In Table 2, the values measured by the generated models on the FID, KID, PPL, P, and R evaluation metrics are provided. The results show that DWBG-Stylegan2 performs best on the FID, KID, and R metrics, with a decrease of 1.99 on the FID metric, a decrease of 0.001603 on the KID metric, and an increase of 0.0463 on the R metric, as compared to Diffusion-Stylegan2; WBG-Stylegan2 performed best on the PPL metric, with a reduction of 19.2516 over Diffusion-Stylegan2; Stylegan2 had the best results on the P metric. In a comprehensive comparison, the performance of DWBG-Stylegan2 is significantly improved relative to Diffusion-Stylegan2. In Table 3, the above five metrics were also used to test the generative model, and the results show that Diffusion-Stylegan2-R2 performs the best on the FID, KID, and R metrics, with a reduction of 0.73 on the FID metrics compared to Diffusion-Stylegan2, a reduction of 0.000657 on the KID metrics compared to Diffusion-Stylegan2 and increased by 0.0051 in the R metrics than Diffusion- Stylegan2; Stylegan2 performed best in the PPL and P metrics. In a comprehensive comparison, the performance of Diffusion-Stylegan2-R2 is improved relative to Diffusion-Stylegan2. In Table 4, six metrics, FID, KID, IS_mean, IS_std, P, and R, are selected to evaluate the ADA series of generative models. The experimental results show that DWBG-Stylegan2-ADA achieves the optimal values for FID, KID, and R, and is slightly lower than Stylegan2 in the IS_mean metric and Diffusion-Stylegan2-ADA in the P metric. However, in a comprehensive comparison, the performance of DWBG-Stylegan2-ADA is optimal among the compared models.

3) COMPARISON EXPERIMENT

To comprehensively compare the performance of the proposed models, this paper conducts comparative experiments on 11 generative models using FID, KID, P, and R as the evaluation metrics, and the experimental results are shown in Table 5.

TABLE 5. Comparative data of 11 generative models on FID, KID, P and R metrics.

Models	FID↓	KID↓	P↑	R↑
GAN	55.65	0.044168	0.3899	0.0238
Stylegan2	23.84	0.014694	0.5257	0.0510
Diffusion-Stylegan2	10.31	0.003873	0.5140	0.2195
Stylegan2-ADA	6.21	0.001416	0.5285	0.3631
WBG-Stylegan2	22.95	0.013619	0.5005	0.0583
DWBG-Stylegan2	8.32	0.002270	0.5009	0.2658
Diffusion-Stylegan2-R2	9.58	0.003216	0.4896	0.2246
Diffusion-Stylegan2-ADA	6.16	0.001283	**0.5404**	0.3610
DWBG-Stylegan2-ADA	**5.67**	**0.001073**	0.5377	**0.3793**
DWBG-Stylegan2-R2	10.32	0.004248	0.5056	0.2009
DWBG-Stylegan2-GP	10.16	0.003353	0.5242	0.2244

The experimental results show that DWBG-Stylegan2 performs better in FID, KID, and R metrics compared to GAN and Stylegan2. Stylegan2-ADA, DWBG-Stylegan2, Diffusion-Stylegan2-R2, Diffusion-Stylegan2-ADA, DWBG-Stylegan2-ADA, and DWBG-Stylegan2-GP outperform Diffusion-Stylegan2 in FID, KID and R metrics. DWBG-Stylegan2 achieves 8.32 in the FID metric, which is 1.99 smaller than the FID value of Diffusion-Stylegan2; in the KID metric, it reached 0.002270, which was 0.001603 less than the KID value of Diffusion-Stylegan2; and in the R metric, it reached 0.2658, which was 0.0463 larger than Diffusion-Stylegan2. In the FID, KID, and R metrics, DWBG-Stylegan2-ADA experimental data was the best generated among the 11 compared models. In the P metric, Diffusion-Stylegan2-ADA reached the optimal value among the 11 compared models. DWBG-Stylegan2-ADA reached 5.67 in the FID metric, which is 2.65 less than the FID value of 8.32 for DWBG-Stylegan2; in the KID metric, it reached 0.001073, which is 0.001197 less than the KID value of DWBG-Stylegan2; and reached 0.3793 in the R metric, which is 0.1135 larger than the R-value of DWBG-Stylegan2. Diffusion-Stylegan2-ADA reached 0.5404 in the P metric, which is 0.0264 larger than the P value of Diffusion-Stylegan2. Overall, DWBG-Stylegan2-ADA generates optimal results on the SA datasets.

Based on the four evaluation indexes of FID, KID, P, and R given in Table 5, the 11 generated models are compared and trained to get the pre-trained model. Evaluation tests are performed on the 11 pre-trained models obtained from training, and the evaluation data of the pre-trained models are measured to obtain the data, which are organized to make Fig.9.

The training of this paper used 8k rounds, and unconditional training of 11 groups of comparison models, mainly on the FID evaluation index with epochs of the data obtained by the recorders setting the recording interval to 200 epochs. Due to some special circumstances, in the training process, retaining the settings of the disconnection reconnection reduces the loss of data due to unexpected circumstances and other problems. Fig.10 gives the change in the value of FID during the training process.

As can be seen from the data given in Fig.10, the GAN training starts to collapse the model at around 2k. Stylegan2 and WBG-Stylegan2 have improved in training time and stability compared to the GAN. Diffusion-Stylegan2, WBG-Stylegan2, Diffusion Stylegan2-R2, DWBG-Stylegan2-GP, and DWBG-Stylegan2-R2 are relatively stable in training. The up and down fluctuations of the data are significantly reduced and stabilized after 1k, and the model starts to slowly collapse after 4k. Stylegan2-ADA, Diffusion-Stylegan2-ADA, and DWBG-Stylegan2-ADA generate the best results, as can be seen from the data in the figure, their FID values are the closest to the real image. However, after adding ADA, the gap between the FID values of the models with the progress of training is not particularly obvious, and the superiority of the improved model cannot be shown intuitively. In a comprehensive comparison, the DWBG-Stylegan2 model has the most obvious effect on SA datasets training generation, the training process is stable, and the training results are also greatly improved based on Diffusion-Stylegan2. Its generation effect is visible in Fig.10(b).

Without considering the ADA series model, the training curve of DWBG-Stylegan2 in Fig.10(b) clearly shows that after the model is trained to 1000 rounds, the generated images have been generally better than the generated images of other models, and the stability of the training process is far better than that of other models. This is closely related to the smooth gradient provided by Wasserstein, which

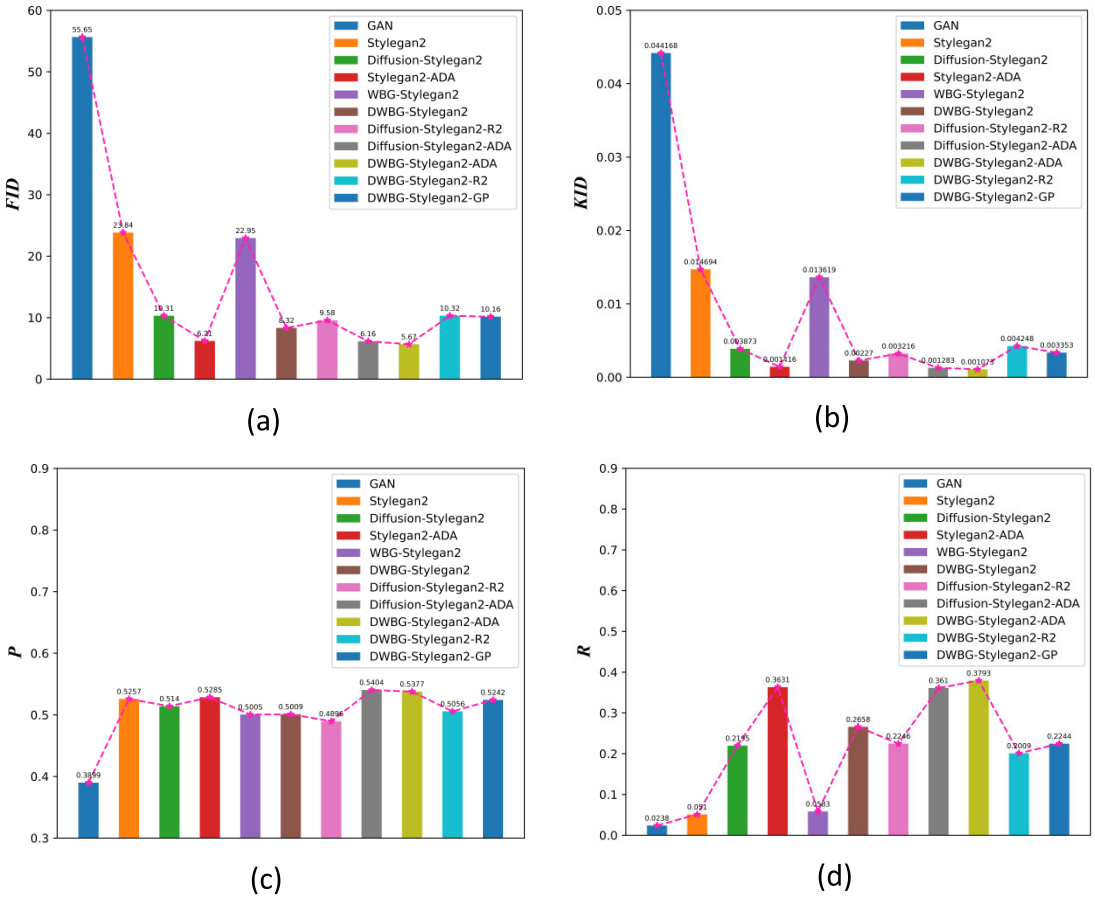


FIGURE 9. Evaluation metrics for underwater image generation: (a). FID evaluation metric; (b). KID evaluation metric; (c). P evaluation metric; (d). R evaluation metric.

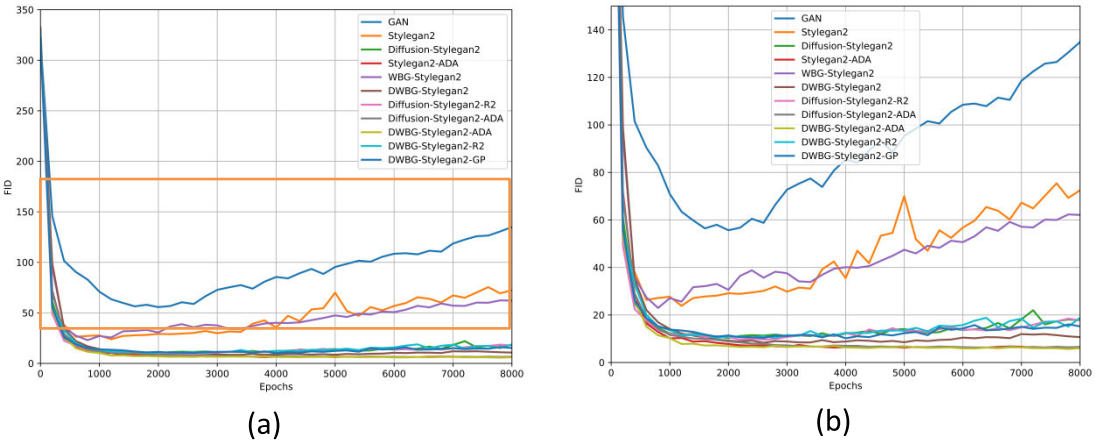


FIGURE 10. The values of FID evaluation metrics during the training process: (a). Changes in FID values during 8k rounds of training; (b). Enlarged line graph of FID between values (0,150) in (a).

reduces unnecessary oscillations in the training process, thus making the model more stable in the training process, while Wasserstein directly measures the "distance" between the generated samples and the real samples to optimizes the quality of the generated images. Therefore, this model achieves the best generation effect in this paper after removing the addition of ADA.

C. GENERATION OF IMAGES OF OTHER MARINE ORGANISMS

In this paper, a simple training generation using the DWBG-Stylegan2 model was performed on existing small-scale fish datasets. The main purpose of this experiment is to demonstrate the generalizability of the model for different underwater image generations, and

FIGURE 11. Fish images generated through 3k rounds of training.

the results of fish monomer generation are shown in Fig.11.

As seen in Fig.11, DWBG-Stylegan2 can effectively generate underwater images of several different species under different datasets. However, the generated images of the fish shown in Fig. 11 differ greatly from the real images except for the fish scales and fins. Analyzing from the algorithm, the reason for such generation results may be improper initialization of parameters or unreasonable setting of hyperparameters in the training process; from the perspective of the dataset, it may be insufficient diversity of data; and from the analysis of the training process, it may be that the training time is too short. Improvements and research on the fish dataset will continue in subsequent experiments.

VI. CONCLUSION

This paper highlights the potential applications of the proposed model in marine biology and related fields by applying the Diffusion-Stylegan2 model across disciplines and proposing six new models based on Diffusion-Stylegan2 for improving the quality of the generated images. A large number of experiments show that the DWBG-Stylegan2 proposed in this paper is more suitable for the generation of complex and variable anemone images than Diffusion-Stylegan2, which reduces the FID by 1.99 based on Diffusion-Stylegan2, and the model training is more stable, and generates higher-quality underwater images with clearer image textures. The FID of Diffusion-Stylegan2-R2 is 0.73 lower than that of Diffusion-Stylgan2 and the quality of the generated image is closer to the real image than that of Diffusion-Stylegan2, which indicates that the R2 regularization term replaces the R1 regularization term to generate a better image. DWBG-Stylegan2-ADA generates the best results, reducing the FID to 5.67, which greatly improves the image generation quality of the generated model. In addition, this paper provides a set of SA datasets with a resolution of 256*256 to solve the problem of insufficient underwater image datasets.

The model in this paper generates higher-quality anemone images compared to previous studies, but some limiting problems also occur. For example, the introduction of ADA technology in the DWBG-Stylegan2 model of this paper, although the model with the best generation effect in this paper is obtained, it still cannot generate the details such as the background effectively, and due to the complexity and uncertainty of the underwater environment, the quality of the generated images is still a certain distance from the realism of the real images. In addition, the operation of retaining the computational graph in DWBG-Stylegan2 in this paper only accelerates the training process to a certain extent, and does not shorten the problem of the long training time of Diffusion-Stylegan2 in the full sense. The next step of the experiment will be to build a more suitable model for underwater image generation for the complex environment of underwater images, and at the same time, propose an effective solution to the problem of long training time.

REFERENCES

[1] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[2] A. Rivolli, L. P. F. Garcia, C. Soares, J. Vanschoren, and A. C. de Carvalho, "Meta-features for meta-learning," *Knowl.-Based Syst.*, vol. 240, Mar. 2022, Art. no. 108101.

[3] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.

[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

[5] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8789–8797.

[6] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4401–4410.

[7] Z. Wang, H. Zheng, P. He, W. Chen, and M. Zhou, "Diffusion-GAN: Training GANs with diffusion," 2022, *arXiv:2206.02262*.

[8] H. Ji, Y. Li, E. Dong, P. Xue, W. Xiong, W. Sun, Z. Tang, D. Zhang, and W. Fang, "A non-rigid image registration method based on multi-level B-spline and L2-regularization," *Signal, Image Video Process.*, vol. 12, no. 6, pp. 1217–1225, Sep. 2018.

[9] L. Mescheder, A. Geiger, and S. Nowozin, "Which training methods for GANs do actually converge?" in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 3481–3490.

[10] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," 2017, *arXiv:1704.00028*.
[11] A. Cotter, H. Jiang, M. Gupta, S. Wang, T. Narayan, S. You, and K. Sridharan, "Optimization with non-differentiable constraints with applications to fairness, recall, churn, and other goals," *J. Mach. Learn. Res.*, vol. 20, no. 172, pp. 1–59, 2019.
[12] M. Li, L. Shen, P. Ye, G. Feng, and Z. Wang, "RFD-ECNet: Extreme underwater image compression with reference to feature dictionary," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 12980–12989.
[13] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.
[14] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.
[15] A. Rokem and K. Kay, "Fractional ridge regression: A fast, interpretable reparameterization of ridge regression," *GigaScience*, vol. 9, no. 12, 2020, Art. no. giaa133.
[16] X. Cheng, B. Khomtchouk, N. Matloff, and P. Mohanty, "Polynomial regression as an alternative to neural nets," 2018, *arXiv:1806.06850*.
[17] P. S. Castro, "Scalable methods for computing state similarity in deterministic Markov decision processes," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 6, 2020, pp. 10069–10076.
[18] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 12104–12114.
[19] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.
[20] T. Xiang, Y. Zhang, Y. Lu, A. L. Yuille, C. Zhang, W. Cai, and Z. Zhou, "SQUID: Deep feature in-painting for unsupervised anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 23890–23901.
[21] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," 2021, *arXiv:2111.11843*.
[22] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.
[23] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8110–8119.
[24] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018.
[25] G. C. Oliveira, G. H. Rosa, D. C. G. Pedronette, J. P. Papa, H. Kumar, L. A. Passos, and D. Kumar, "Which generative adversarial network yields high-quality synthetic medical images: Investigation using AMD image datasets," 2022, *arXiv:2203.13856*.
[26] G. Parmar, R. Zhang, and J.-Y. Zhu, "On aliased resizing and surprising subtleties in GAN evaluation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 11410–11420.
[27] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, "Demystifying MMD GANs," 2018, *arXiv:1801.01401*.
[28] Q. Liu and D. Wang, "Stein variational gradient descent: A general purpose Bayesian inference algorithm," 2016, *arXiv:1608.04471*.
[29] J. Wang et al., "Side-aware boundary localization for more precise object detection," in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*. Springer, 2020, pp. 403–419.
[30] S. Barratt and R. Sharma, "A note on the inception score," 2018, *arXiv:1801.01973*.
[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

HUIYING ZHANG received the Ph.D. degree from Changchun University of Technology, in 2017. She is currently with the School of Information and Control Engineering, Jilin Institute of Chemical Technology, China. Her research interests include spatial optical communication modulation reception technology, visible light communication, and localization technology.

FEIFAN YAO received the bachelor's degree from Weifang Medical University, in 2022. He is currently pursuing the master's degree in engineering with the School of Information and Control Engineering, Jilin Institute of Chemical Technology. His research interests include image processing and generative artificial intelligence.

YIFEI GONG is currently pursuing the bachelor's degree in engineering from the School of Information and Control Engineering, Jilin Institute of Chemical Technology. His research interests include artificial intelligence and computer vision and image processing.

QINGHUA ZHANG received the bachelor's degree from Harbin Institute of Petroleum, in 2022. He is currently pursuing the master's degree in engineering with the School of Information and Control Engineering, Jilin Institute of Chemical Technology. His research interests include image processing, communications, and signal processing.