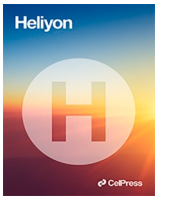
Heliyon 10 (2024) e25250

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Generative adversarial network based on frequency domain data enhancement: Dual-way discriminator structure Copes with limited data

Jian Wei , Qinzhao Wang *, Zixu Zhao

Army Academy of Armored Forces, Beijing, 100071, China

ARTICLE INFO

Keywords:
Generative adversarial network
Limited datasets
Dual-way model
Frequency domain

ABSTRACT

The excellent image-generation ability of generative adversarial networks (GANs) has been widely used. However, training a GAN requires large-scale data support, which hinders in-depth development. Therefore, the research on stable training of GANs under limited data conditions is helpful to further expand the application scenarios. To solve this problem, a new network based on a dual-ways discriminator structure is designed, used to eliminate the problem that a single discriminator model is prone to overfitting under the condition of limited data. Then, the problem that the traditional data augmentation strategy is limited to pixel space and lacks attention to the overall structure and contour of the image is analyzed. An adaptive dynamic data augmentation strategy based on the Laplace convolution kernel is proposed from the frequency domain space, which realizes the purpose of implicitly increasing the training data in the training process. This new designed module improves the performance of the generative adversarial network. Through extensive experiments, it was confirmed that the new network, named FD-GAN, achieved prefer image generation ability, and its Fid score reached 4.58, 12.007, and 10.382 in the AFHQ-Cat, AFHQ-Dog, and TankDataSet datasets, respectively.

1. Introduction

Generative adversarial networks (GANs) are used in art creation [1–6], object detection [7–12], style transfer [13–19], and even adversarial samples [20–24] of rapid development. We note that these powerful capabilities of GAN are supported by various large-scale datasets. Under the data-driven strategy, after a long period of training, GANs based on the deep neural network learns an effective mapping rule, that is, the semantically clear image samples are generated by random vectors [25]. However, as we mentioned, massive amounts of data are necessary to achieve this function. In fact, what is not optimistic is that in some tasks that lack existing massive data, such as medical pathology samples, military target samples, etc., it is difficult and inefficient to manually collect and sort out such massive data [26,27]. Therefore, how to carry out GANs training under limited data conditions and obtain stable training results is worth studying. This helps to further reduce the difficulty of training GANs and expand the application field of GANs.

To reduce the difficulty of training GAN under limited data conditions, researchers have carried out a lot of research. By summarizing their work, it can be summarized into 2 main practices: the first one is to adopt data augmentation strategies, such as DA (data augmentation) [28], ADA (adaptive data augmentation) [29], APA (adaptive pseudo augmentation) [30], FSMR (feature statistics

* Corresponding author.
E-mail addresses: 18513327667@163.com (J. Wei), airy_snow@outlook.com (Q. Wang), aafezzx@163.com (Z. Zhao).

https://doi.org/10.1016/j.heliyon.2024.e25250
Received 20 June 2023; Received in revised form 29 November 2023; Accepted 23 January 2024
Available online 10 February 2024

mixing regularization) [31], GenCo (generative co-training) [32], and adopt diversified data augmentation strategies to provide more training data for the model; The second is to optimize the cost constraint function, such as ICR (improved consistency regularization) [33], FFL (focal frequency loss) [34], LeCam [26] and other models, and achieve stable training GAN model by improving the cost constraint function. In short, to solve the problem of stable training GAN under limited data conditions, the fundamental focus is how to optimize the discriminator to prevent it from falling into overfitting, to achieve the purpose of improving the quality of the generated images.

As a seminal work, we propose a frequency-domain enhanced dual-ways discriminator generative adversarial network (FD-GAN). Firstly, we re-examine the data augmentation strategy, analyze the characteristics of the complete structure and clear contour of the image in the frequency domain space. And then we propose a training method based on the frequency domain data augmentation strategy. Specifically, we use the Laplace convolution kernel to realize the conversion of the image from pixel space to the frequency domain. The designed adaptive conversion intensity parameters to control the richness of image details and edge structures in the data conversion process, useful for dynamically enhance the training data. Then, endeavor to further stabilize the training process of GANs under limited data conditions and prevent single discrimination from falling into overfitting, we design the structure based on a dual-ways discriminator. Through synchronous training of these two discriminators, we can obtain more feature information of the image at the same time, and provide more effective loss gradient information for the generator of GAN after fusing the output results of the discriminator at the end, to improve the quality of the generated image. In the experimental part, we designed and implemented the FD-GAN model based on the typical unconditional generative adversarial network StyleGAN2. Extensive experiments are conducted in the dataset AFHQ [35] and TankDataSet. Experimental results show that our designed FD-GAN gains more stable training and achieves better performance, especially in our self-built TankDataSet dataset[1], where the Fid of FD-GAN reaches 10.382.

In summary, the main contributions of this paper are 4 points:

1 A dual-ways discriminator generative adversarial network is proposed, further expands the research of GANs, realizes the synchronous training mechanism by improving the model structure, and alleviates the overfitting of the discriminator. At the same time, the dual-ways discriminator adopts the loss processing method of end feature fusion, which improves the feature information of the image in the loss and provides a rich gradient for generator training;
2 A data enhancement strategy based on frequency domain space is conducted, taking advantage of the characteristics of complete structure and clear contour of images in frequency domain space. Laplace convolution kernel is proposed for image conversion, adding intensity adaptive control parameters, controlling the degree of detail after image conversion, and provide more data for GAN training;
3 Self-built limited image dataset is handed. By querying the public image website, we refer to AFHQ and FFHQ, and collect and organize a 3888 tank image, called TankDataSet,[1] to expand the research object of GAN.
4 Extensive experiments are conducted to verify the FD-GAN model capability. The results showed that the FD-GAN Fid reached 4.58, 12.007, and 10.382 in the AFHQ-Cat, AFHQ-Dog, and TankDaTaSet datasets, respectively.

2. Related research

2.1. GANs

GANs is a deep neural network model based on game theory, and its main function is to complete semantically explicit images generated by randomization. This model was first proposed by Goodfellow et al. [25] in 2014 and has since developed rapidly. So far, many interesting research contents have been derived based on GANs, such as the most popular Artificial Intelligent Generated Content (AIGC) technology [36,37], which can be used for image editing, art creation, style conversion, etc., behind which are Diffusion GAN [38], StartGAN [3], CUT (contrastive unpaired translation) [39] and other models to the credit. At the same time, many studies are committed to continuously improving the quality of generated images, such as StyleGAN [40] proposed an unconditional generative adversarial network based on the AdaIN (adaptive instance normalization) mechanism [41], which decouples the feature entanglement problem between input vectors, and StyleGAN2 [35] proposed a generative model with convolutional kernel encoding-decoding. Literature [34,42] improves image quality from the frequency domain by adding structure and contour loss functions. Unfortunately, however, these efforts require FFHQ (flickr faces HQ) [40] and CelebA [43]. and other large-scale datasets. Data-driven is at the heart of all models based on deep learning techniques, so GANs also face the challenge that training GANs stably is a challenging task under limited data conditions. In this regard, from the perspective of improving the model structure, we design a new generative model based on a dual-way discriminator, and in addition, without increasing the workload of data collection, a data augmentation technology based on frequency domain space is proposed to achieve the purpose of stable training GANs.

2.2. Limited GANs

GANs training under limited data conditions has been of increasing interest to researchers. Studies have shown that the biggest impact of limited data is the overfitting of discriminators, which is the root cause of the difficulty of GAN training [29,30]. Therefore,

[1] The TankDataSet has been uploaded to: www.kaggle.com/datasets/airy975924806/tank-for -DDG.

given this task, the current main practice is to alleviate the gap between the amount of data and the number of model parameters through data enhancement technologies such as geometric transformation, blending, AdaIN, copy-paste, feature fusion, etc., to reduce the possibility of the discriminator falling into overfitting. Based on this observation, DA [28] points out different types of image transformation methods, which play different degrees, and proposes to use geometric transformation methods such as color jitter, translation, and clips that are not easy to "leak". ADA [28] applies dynamic control parameters to traditional geometric transformations, enriches the sample distribution space, and avoids the problem of "leakage" by controlling the application probability of geometric transformations. APA [30] further from the perspective of effective use of generated samples, proposed to input samples to the discriminator with dynamic probability, to reduce the discriminator recognition ability, and also achieve the purpose of preventing the discriminator from falling into overfitting, in addition, FMSR [31] and GenCo [32]. Similarly, from the perspective of data expansion, by adding the image style to transform the model, the degree of overfitting of the discriminator is reduced. On the other hand, some methods based on optimized f divergence are used to limit the performance of the discriminator, such as ICR [33] proposes a loss constraint function based on data self-enhancement technology, which performs a geometric transformation on the real image and the generated image at the same time, and then takes the classification loss of the discriminator as the constraint condition to ensure that different forms of the same target get similar judgment results. Spectral Regularization [44] limits the range of weight parameter updates to a very small range, ensuring that the depth model always conforms to the Lipschitz continuity constraint and avoids gradient vanishing and explosion. LeCam [26] dynamically adjusts the comprehensive loss of the discriminator by adding two weight guidelines to prevent the update parameters from being dominated by unilateral loss. Based on these studies, we propose a cross-spatial data augmentation strategy based on Laplace convolution kernels, which is different from the traditional geometric enhancement, color jitter, etc., the frequency domain more image the structure and edge features of the image, providing more substantial training samples for GANs training, and also achieving the effect of alleviating model overfitting.

2.3. Multi-branch structure

The multi-branched model structure is not a completely new design, and this idea has been excellent in many studies [45,46]. For example, the model based on the Transformer attention mechanism [47,48] increases the feature extraction ability of the model by setting 2 or more branch modules. In Ref. [47], a GAN model of a dual-channel generator is designed to balance the false alarm rate and missed detection rate of the infrared small target detection process. In Ref. [48], a dual-ways data stream is adapted to implement a contrastive learning mechanism. Similarly, DivCo (diverse conditional) [43] even adopted a three-ways branching model. However, there are relatively few studies on image-generation tasks. Combined with limited data, the advantages and disadvantages of the discriminator have a more direct effect on the final performance of the model. Specifically, we improve the model based on the typical unconditional generative adversarial network, only need to add a new discriminator branch based on the original structure, without major adjustments to the model structure. The advantages of this are a training condition does not require special attention and can

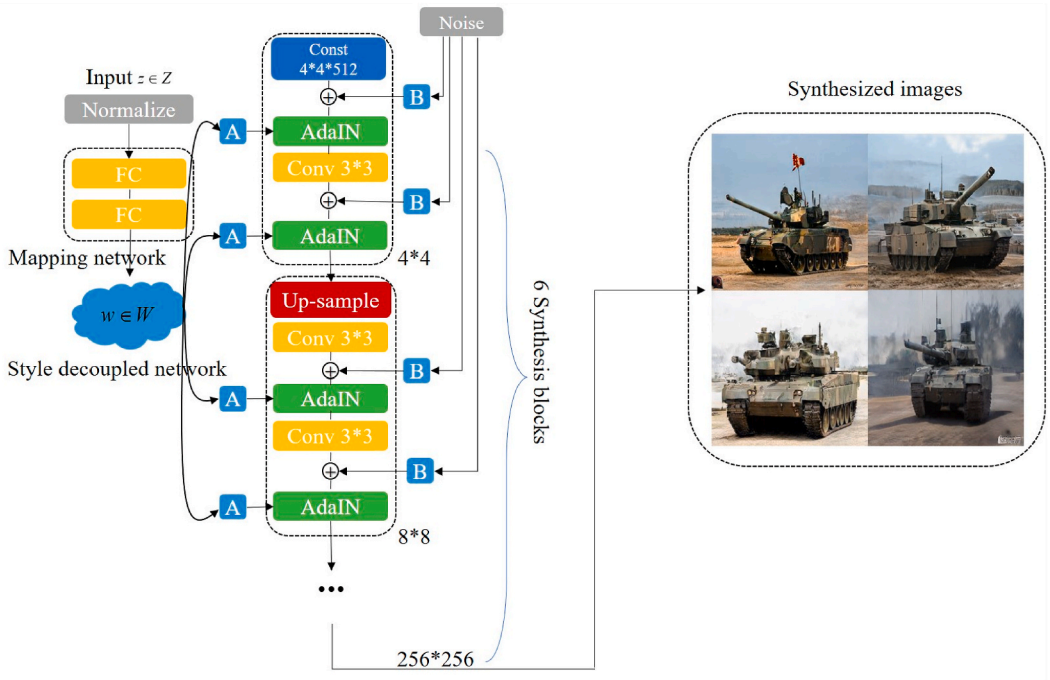


Fig. 1. The structure of Generator (G). FC (full connection) is used for multiple layers perceptron, W is the decoupled feature, A and B is the mean and variance separately, const is one random scalar in each iteration.

follow the design scheme of typical models such as ADA [29], and APA [30]; The second is to ensure that FD-GAN can be fairly compared with many models that are also based on StyleGAN2. The results show that this structural design is easy to deploy and has higher performance, and Fid exceeds the single-branch ADA model by 0.49 and 7.006 in the AFHQ-Cat, and AFHQ-Dog, respectively.

3. Preliminary

Our FD-GAN model is developed from the SOTA (state of the art) model named StyleGAN2. In this part, we first introduce the model structure and training strategy of StyleGAN2.

3.1. Instruction of StyleGAN2

StyleGAN2 [35] is one of the SOTA networks used for image generation. we illustrate its architecture in the following part. First, the network consists of one style-based generator and a discriminator. Regarding the model structure of StyleGAN2, we recommend checking the literature [35]. Below we explain the detailed composition of the FD-GAN model generator and discriminator.

3.1.1. Generator

The generator (G) used in the experiment in the manuscript mainly completes the task of generating 256*256 images, so considering the amount of calculation and the quality of the generated images, we set 2 fully connected decoupled modules, 1 style decoupled module, 6 synthesis modules, and the model structure is shown in Fig. 1.

The working process of the model can be described as follows: by randomly sampling the normal distribution, we set the model input to a noise vector of 512 dimensions. If the input is not sampled from the standard normal distribution, we can still convert it to a standard normal distribution with a mean of 0 and a variance of 1 by normalization, which allows the model to generate more diverse images. Then, the two-layer fully connected layer realizes the transformation of random noise vectors into the hidden space of the feature vector, that is, the space *W*. It has been confirmed in StyleGAN2 that the combination of multiple layers of fully connected layers uses the characteristics of nonlinear transformation of neural networks. By decoupling the interaction between the dimensions of the input vector, the mapping network avoids the feature coupling effect caused by the image generation of the noise vector, and then achieve the accurate matching between the values of each dimension and the target features. The Synthesis module is the core of the model, using the mechanism of AdaIN. The initial image is guided under the Style feature vector, from the small size target image of

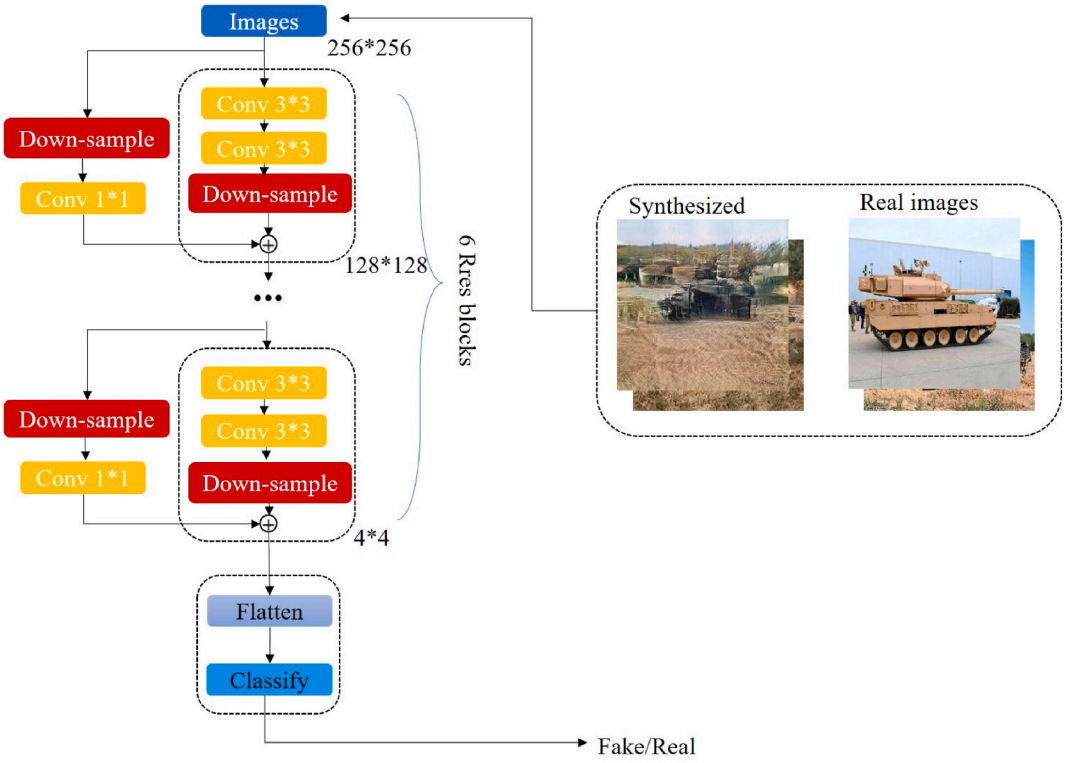


Fig. 2. The structure of Discriminator (D). The discriminator is mainly composed of 6 blocks, each block adopts the residual connection, which effectively prevents the loss of features in the forward propagation process. The discriminator ultimately outputs a dichotomous cross-entropy loss of the input image.

4*4, gradually generating a tank image of 256*256 size specified in this paper. The AdaIN mechanism is represented by a formula as follow:

$$\mathrm{AdaIN}(x) = \frac{x-\mu}{\sigma}\gamma + \beta \tag{1}$$

where x is the image feature map, μ, σ are the mean and standard deviation of the x feature map, γ, β is the mean and standard deviation of the style vector, respectively. Through this normalization and demodulation, the transformation of the feature style can be realized. In particular, γ and β is a set of learnable parameters during training, determined by the style vector.

At this point, we realize that the generative model based on StyleGAN2 is still essentially generating target images with clear meanings from noise vectors. At the same time, to further increase the robustness of the model and the diversity of generated images, some additional noise disturbances are added to the Synthesis module. So that the target image shows more diverse background

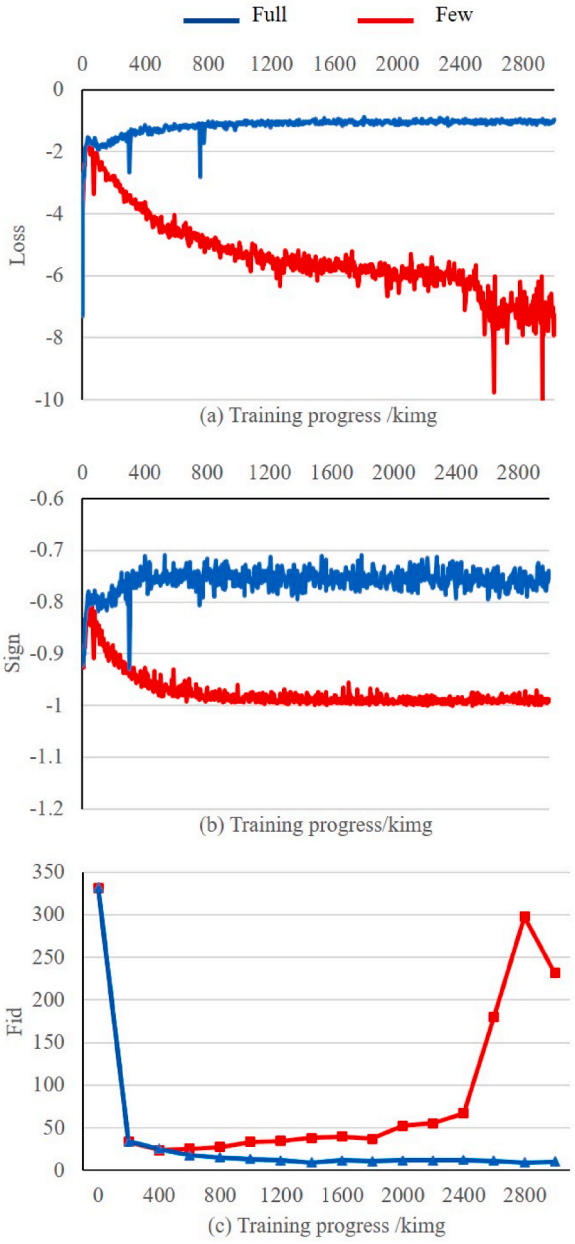


Fig. 3. The training curve of the model in the TankDataSet. (a) the discriminator discriminant loss of the discriminator on the true and false targets, (b) the probability of the true and false targets being correctly recognized by the discriminator (using positive and negative numbers to represent the true and false sample recognition results), (c) the change of the Fid of the image quality evaluation index.

details, weather, texture, color and other details with the help of slight disturbances of noise based on maintaining the main content. Combined with the set image size, one synthesis module completes a 2x up-sampling, so a total of 6 Synthesis modules needs to be superimposed from the initial 4*4 image.

3.1.2. Discriminator

In the manuscript, the discriminator is only used to binary classify the input image. Therefore, for the 256*256 size image, the discriminator also needs 6 convolutional blocks, 1 flatten layer and 1 classification layer (the detailed is shown in Fig. (2). Furthermore, there are three possible connection structures of the discriminator (D) are compared experimentally, and here we show a discriminator model based on the residual connection method.

As shown in Fig. 2, the final discriminator designed in this paper is based on a typical residual module, where the input to the model is an image of 256*256 size, and they come from a real dataset or generator. The workflow of the discriminator is as follows: for the input 256*256 image, 6 layer-by-layer convolutional blocks extract the features of the image successively, which is worth noting, in the residual module, the input image is extracted by two branches, one of which is composed of two convolutional layers with 3 convolution kernel and a 2x down-sampling layer, the other is composed of 2x sampling and 1*1 convolutional layer, and finally, the two features are summed and input to the next residual module. After layer-by-layer processing of the image, at the end of the discriminator, the model extracts the core features of the image. Before the binary classification judgment, drawing on the idea of Patch-GAN, the feature map is kept at a size of 4*4, and then the feature map containing image features is converted into a 16-dimensional feature vector through Flatten. Finally, the judgment is carried out by the binary cross-entropy loss function. Thus, the discriminator completes the judgment of the true or false attributes of the input image, and at the same time, obtains the error loss that is critical to updating the generator parameters.

3.2. Training strategy

The training metric of Vanilla GAN is the maximum minimization cost function, and the formula is expressed as follows:

$$\min_G \max_D V(G,D) = E_{x\sim p_x}[\log(D(x))] + E_{z\sim p_z}[1-\log(D(G(z)))] \tag{2}$$

where $V(G,D)$ is the loss function of GAN, E is the expected value of the loss， $x \sim p_x$ represents the training data set extracted from the true probability distribution, $z \sim p_z$ is a random vector sampled from the standard normal distribution, $G(z)$ is the generated image from random input vector , log is the sigmoid nonlinear activation function， $D(\cdot)$ is the result of the input image. In the actual training process, G and D adopt the method of asynchronous training update, and their cost functions L_D are L_G expressed as follows:

$$L_D = E_{x\sim p_x}[\log(D(x))] + E_{z\sim p_z}[1-\log(D(G(z)))] \tag{3}$$

$$L_G = E_{z\sim p_z}[1-\log(D(G(z)))] \tag{4}$$

When the training data is insufficient, D quickly learns how to distinguish between real and fake targets. At this time, because D can distinguish between true and false images with high confidence, and then the discrimination loss tends to stabilize, resulting in the disappearance of the loss gradient, G cannot obtain effective feedback information. Thus, to satisfy the constraints of the maximum minimized loss function, G is forced to generate several images of the fixed model. G occurs mode collapse phenomenon. As shown in Fig. 3, the training process curve of StyleGAN2 under different levels of TankDataset is shown, and the changing trend of the loss curve in (b) shows that StyleGAN2 gradually falls into complete overfitting, resulting in the gradual divergence of the corresponding Fid curve, indicating that the quality of the generated image becomes worse and the model performance decreases.

It can be seen from Fig. 3 (b) that with the improvement of the resolving ability of the StyleGAN2 model discriminator, the accuracy

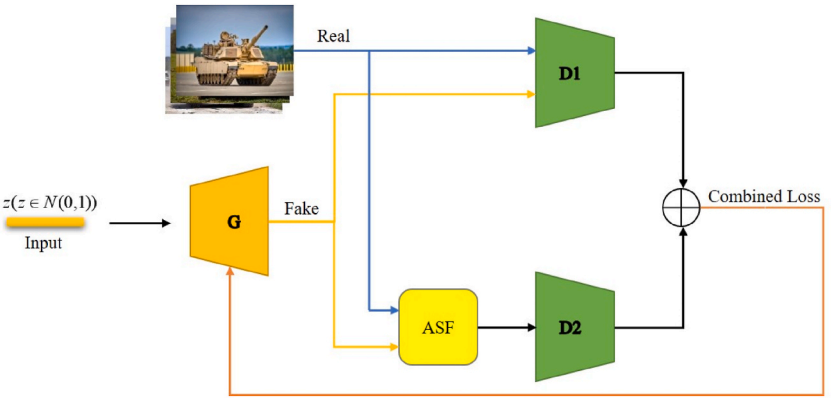


Fig. 4. The structure of the FD-GAN model. The core of FD-GAN is 2 homogeneous discriminators, 1 dynamic filter module, and the generator performs parameter update under the loss after fusion.

of correctly classifying true and false targets increases, and the loss generated becomes less, resulting in the gradual disappearance of the loss gradient, which is manifested in (a) as G due to the lack of effective guidance, so that the loss of the generated image gradually increases, and at the same time, in (c), the image quality index Fid gradually diverges, proving that the difference between the generated image and the real sample is increasing.

4. FD-GAN

The FD-GAN solution to the problem of overfitting caused by limited training data consists of two parts, namely the dual-ways discriminator structure and the frequency domain data augmentation strategy. This section describes them in detail.

4.1. Model structure

As shown in Fig. 4, the G of FD-GAN uses the generator in Fig. 1 to modulate the input vector into a generated sample through a feature decoupling network and a style encoding-decoding network $G(z)$; D1 and D2 adopts the discriminator in Fig. 2, and calculates the binary cross entropy loss of the input image. Mean fusion of the output from D1 and D2 is used to guide the update the parameters of G. The Adaptive Structure Filter (ASF) adopts an edge feature extraction operator based on the Laplace convolution kernel to adaptively extract the target structure features during the training process to assist D2 training.

Based on the traditional GANs model, FD-GAN independently realizes the training process of mutual supervision by adding a discriminator branch. Upon the discriminator's learning ability, FD-GAN avoids repeated adjustment of control parameters. The loss functions of D1 and D2 in Fig. 4 are:

$$L_{D_1} = E_{x\sim p_x}[\log(D_1(x))] + E_{z\sim p_z}[1 - \log(D_1(G(z)))] \tag{5}$$

$$L_{D_2} = E_{x\sim p_x}[\log(D_2(ASF(x)))] + E_{z\sim p_z}[1 - \log(D_2(ASF(G(z))))] \tag{6}$$

where $ASF(\cdot)$ represents the dynamically extracted sample structure feature operator. The functional structure diagram is shown in Fig. 5. In summary, the total classification loss of the FD-GAN model is expressed as:

$$L_G = \frac{1}{2}\left[E_{z\sim p_z}[1 - \log(D_1(G(z)))] + E_{z\sim p_z}[1 - \log(D_2(ASF(G(z))))]\right] \tag{7}$$

4.2. Adaptive structure filter

Traditional data augmentation techniques, such as geometric transformation, feature interpolation, color dithering, etc., pay more attention to the representation of images in pixel space [49,50]. The transformation strategy of pixel space makes GANs more inclined to pay attention to the characteristics of image texture and brightness, resulting in the structure of the image generated by traditional GAN is not clear enough. Different from this, the frequency domain space pays more attention to the edge and structural characteristics of the image, which can make up for the feature information missed by GAN. Therefore, the purpose of the dynamic structure filter is to use the spatial transformation method to use the Laplace convolution kernel with an edge filtering effect to complete the image structure feature extraction task with the intensity parameters.

In the traditional image processing operator, the isotropic derivative operator has the characteristics of response independent of direction, so it has the characteristics of maintaining the structural information of the original map. The Laplace convolution kernel is the simplest isotropic derivative operator, which is essentially a high-frequency filter. GenCo adopts a similar idea, proposes RFCF. RFCF randomly deletes the frequency component through a multi-layer loop structure, removes a large number of detailed features in the image, and tries to retain the edge information of the image. However, the calculation efficiency of the multi-loop strategy is slow, and the quality of the reconstituted image is low, and the target structure is not clear enough (see Fig. 6). Therefore, this paper proposes to use a 3*3 Laplace convolution kernel to dynamically process images, considering the image structure characteristics and computational complexity. The convolution kernel weights of ASF are defined as follows:

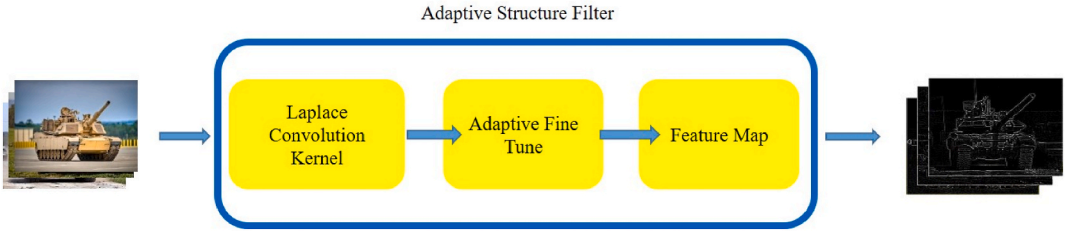


Fig. 5. The diagram of ASF. The ASF consists of a Laplace convolution kernel and a dynamic weight parameter fine-tuning operator. The image undergoes a dynamic filtering module, and the output frequency component is adaptively enhanced structural image.

$$\text{Kernel} = 2\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \Big/ e^{2p} \tag{8}$$

$$p = E[sign(D(x))] \tag{9}$$

In the formula p is a dynamic parameter, which indicates the probability of the discriminator correctly distinguishing the real image, which dynamically changes with the degree of overfitting of the discriminator during the training process. $p = 0$ indicates that it is not overfitting, and $p = 1$ indicates fully overfitting, $sign(\cdot)$ is a sign function that counts the number of positive and negative loss results.

As shown in Fig. 6, the image edge structure features processed by the RFCR policy in GenCo are relatively extensive, and a lot of detailed information about the tank is lost. On the contrary, ASF extracts the overall structure and detail features of the target more accurately through dynamic convolution kernels.

5. Experiments

5.1. Datasets

5.1.1. TankDataSet

we use the retrieval of the public image website to obtain raw armored vehicle image. Then manually delete the target image with poor image definition, inconsistent size, and too small target proportion, and complete the cleaning of the original data. Finally, a dataset containing 3888 armored vehicle images (TankDataSet), whose image size is unified to 256*256, is obtained.

5.1.2. AFHQ

AFHQ [35] consists of 3 target images, which are composed of 5153 Cat, 4739 Dog, and 4738 Wild images. In this paper, we select AFHQ-Cat and AFHQ-Dog as experimental subjects.

5.1.3. Platform

This experiment is based on the PyTorch deep learning framework, and the hardware platform is configured as Windows 10 operating system, Intel i7-9700 CPU, and Nvidia RTX 3090 GPU.

5.2. Metrics

5.2.1. Fr'echet Inception Distance

Fr'echet Inception Distance (Fid) [48] is based on the feature similarity evaluation of the Inception-V3 model, measuring the relative distance between the generated image and the real image in the feature space. The lower Fid indicates higher quality of the generated image, which means that the better corresponding model, vice versa. In practical applications, Fid is more in line with the observation results of the human eye. It is widely used to evaluate the performance of image generation models; the calculation formula is:

$$Fid = \|\mu_r - \mu_g\|^2 + Tr\left(\sum\nolimits_r + \sum\nolimits_g - 2\left(\sum\nolimits_r \sum\nolimits_g\right)^{\frac{1}{2}}\right) \tag{10}$$

where μ_r is the feature mean of the real picture, μ_g is the feature mean of the generated picture, $\sum_r$ is the covariance matrix of the real picture, $\sum_g$ is the covariance matrix of the generated picture, Tr is the trace of the matrix.

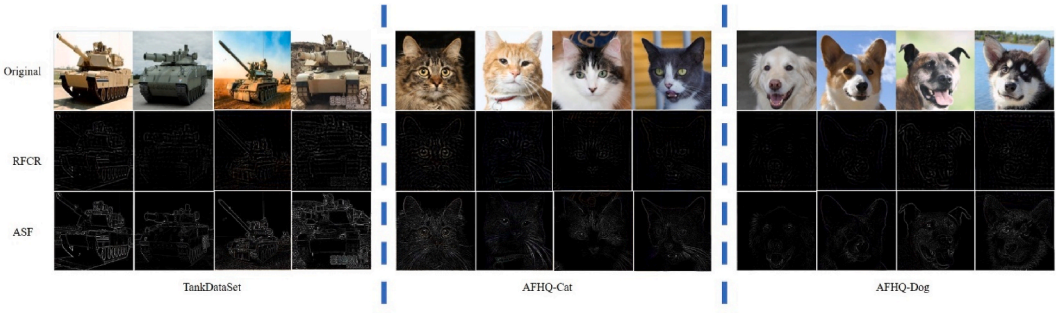


Fig. 6. The comparison of methods in the frequency domain ($p = 0.6$). Compared with the RFCF strategy, ASF can provide clearer images on different datasets.

5.2.2. F1 score

The F1 score is an indicator to evaluate the binary classification ability of the model, reconcile the accuracy and recall of the model, avoid bias in the model evaluation, and achieve a more objective evaluation. Its calculation formula is:

$$F1 = \frac{\text{precision} \times \text{recall}}{\text{recall}} \tag{11}$$

5.3. Comparison models

To evaluate the ability of the FD-GAN model of generating complex images, this paper selects StyleGAN2 [35], ADA [29], APA [30], FFL [34], and GenCo [32] as comparision. Among them, StyleGAN2 is the unconstrained generative model with the best current performance in large-scale datasets. ADA is an improved model of the StyleGAN2 model for limited training datasets, which adopts traditional data augmentation strategies to process data. Based on the StyleGAN2 model, APA enriches the training datasets with the data generated by the generator in real-time. FFL adopts frequency domain constraint, adds the corresponding frequency domain second-order loss to the cost loss function, and improves the image quality generated by the model through synchronous guidance of pixel space and frequency domain space. GenCo adopts a dual-ways model structure to randomly delete frequency components in the frequency domain space, to increase the model training data and improve the model's ability to learn image features. The following is an analysis of the aspects of generating images, visual performance, and quantitative indicators, respectively.

5.4. Results in TankDataSet

Fig. 7 shows the randomly generated images of each model in the TankDataSet dataset, which intuitively reflects the image generation capability of each model. From Fig. 7, it can be seen that the images generated by StyleGAN2 and ADA models have obvious oblique noise; The APA model generates images that lack image texture information. FFL has insufficient ability to distinguish between background and foreground, and even produces deformation; GenCo-generated images are also significantly deformed; FD-GAN integrates image texture and structural features well, and the generated tank image maintains good structural integrity and diversity.

It can be seen from Fig. 8 that in the initial stage of training process, the curves of each model rapidly decrease, indicating that the generator learns effectively under the guidance of gradient information and quickly updates the model parameters in the direction of generating realistic images. However, after 500 kimg iterative training, the curves of StyleGAN2 and APA began to diverge, indicating that their discriminators gradually fell into overfitting and could not continue to provide effective gradient information. The rest of the

Fig. 7. The generated images of different models. Each row is generated by the same random seed, and the main structure of the image generated by different models is very different, and poorly trained models cannot produce clear detailed structures, and vice versa.

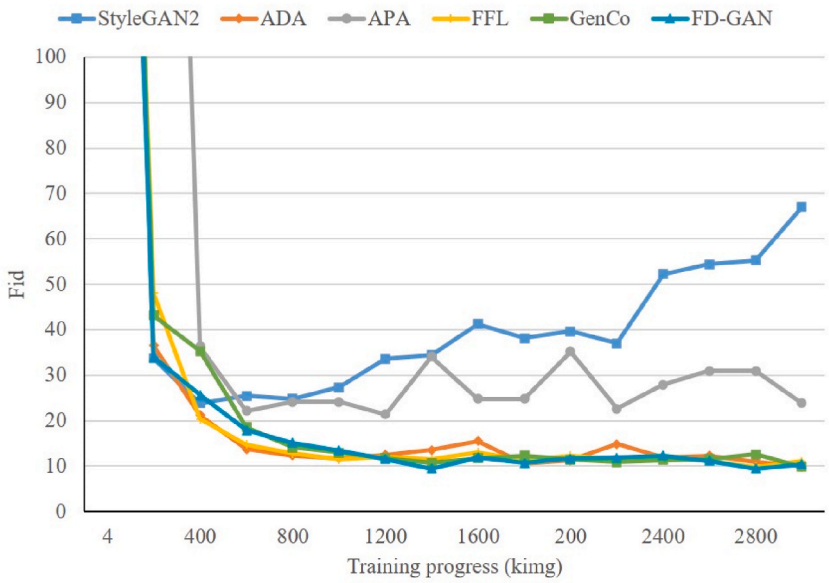


Fig. 8. The diagram of the training process. The FID curve can clearly show the training process of the corresponding model, the curve divergence indicates that the model training failed, and the curve fluctuation indicates that the model converges unstable. Kimg is an abbreviation for 1000 training images, indicating the training time.

models maintained a prefer parameter update gradient and guided the quality of the generated images to steadily decrease. Specially, the FD-GAN maintained a more stable generation process under the joint action of the dual-ways discriminator, and took the lead in obtaining smaller Fid indicators around 1500 kimg. Furthermore, Fig. 8 reflects FD-GAN shows smaller curve fluctuations compared to others.

Comparing the Fid scores in Table 1, it can be seen that FD-GAN achieves better performance with the dual-ways model structure and diversified data enhancement methods, and achieves a performance improvement of 16.95% compared with the single-channel optimal model ADA. Compared with GenCo, the optimal model with a dual-ways structure, the performance improvement is 12.15%. Section 3.2 analysis points out that reasonable data augmentation can effectively enrich the data distribution space and improve the model training effect, but the training data processed by RFCF lacks effective features of the target (Fig. 6), resulting in the discriminator cannot effectively learn the structural features of the tank, resulting in lower Fid score. FFL forces the model to fit the frequency domain distribution features of the image by directly adding the frequency domain loss function; it has an advantage in highly aligned datasets with simple target structures such as CelebA and FFHQ with low image diversity, but not in TankDataSet, which has complex structural features and diverse morphological expressions. The frequency domain loss function will cause the model to be unable to find a more accurate update direction.

5.5. *Results in AFHQ*

We also verify the performance of FD-GAN in the AFHQ dataset, and we focus on the change of loss curve during FD-GAN training. As shown in Fig. 9, the classification loss of true and false images and the degree of model overfitting in the process of model training are shown respectively. First of all, the classification loss curve of the discriminator for true and false targets in (a) shows that there is a fluctuating process in the initial stage of training of each model, indicating that the discriminator gradually learns how to distinguish between true and false targets in the error loss. However, with the gradual enhancement of the generator's generation ability, the

Table 1
The results in TankDataSet in terms of Fid, Precision, Recall, and F1.

Methods	Fid (↓)	Precision (↑)	Recall (↑)	F1 (↑)
APA[a]	36.442	0.5527	0.0289	0.0550
ADA[b]	12.501	**0.6936**	0.1281	0.2163
FFL[c]	12.731	0.6525	0.1287	0.2150
GenCo[d]	11.818	0.6167	0.2199	0.3243
FD-GAN	**10.382**	0.6299	**0.2432**	**0.3509**

[a] https://github.com/EndlessSora/DeceiveD.
[b] https://github.com/NVlabs/stylegan2-ada-pytorch.
[c] https://github.com/EndlessSora/focal-frequency-loss.
[d] https://github.com/jxhuang0508/GenCo.

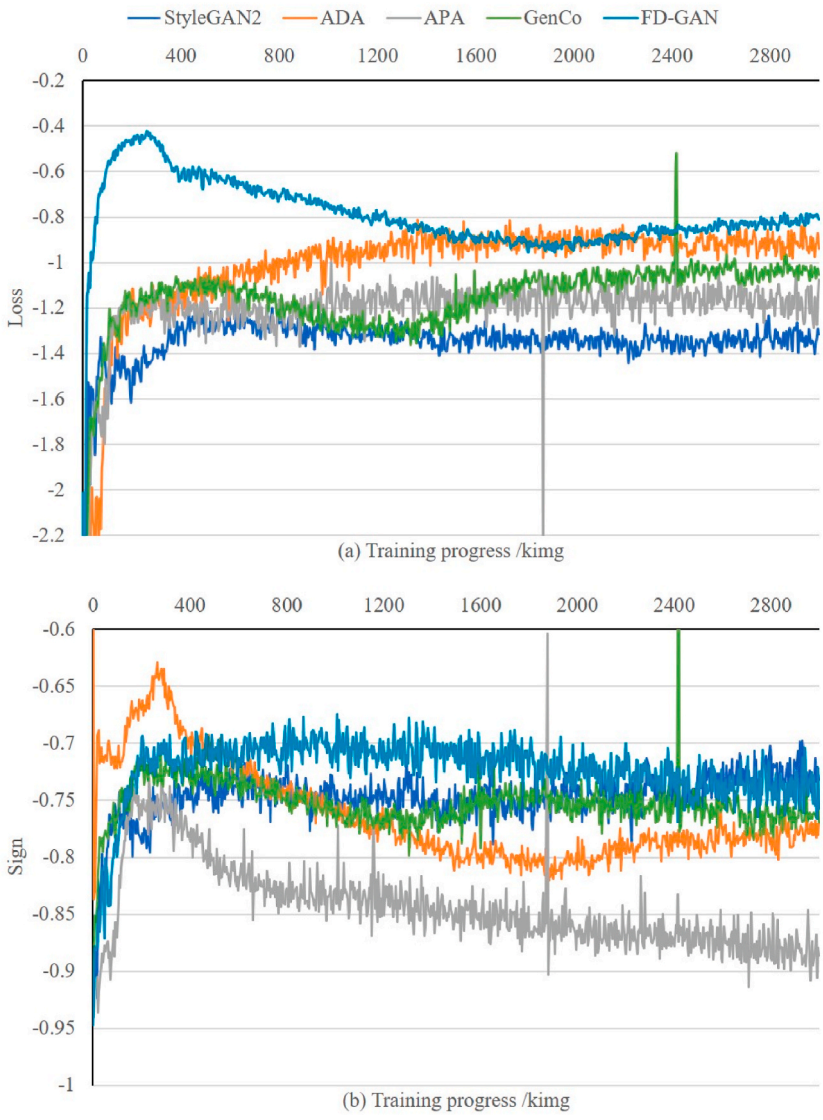


Fig. 9. Training progress in AFHQ-Cat. (a) The discriminator's binary-classification loss curve response of false images during the training process, representing the discriminator's classification ability; (b) The proportional change curve of the discriminator correctly classifies false targets, representing the overfitting of the discriminator.

discriminator is increasingly unable to effectively distinguish between true and false images, which is manifested by the gradual increase of the loss curve. Finally, under the training strategy of game confrontation, the discriminator and the generator reach a balance, that is, the curve gradually tends to be stable, and its numerical size reflects the discriminator's ability to distinguish between true and false targets. (b) Taking the proportion of correct distinguishing between true and false images as the standard, the degree of overfitting of the discriminator is more clearly displayed, in which according to ADA, APA, etc., the overfitting upper line is set $p = 0.6$. The upper part of the abscissa axis shows that each model can better follow this optimization setting in the real image, but the ability to identify false images is different. FD -GAN is more able to maintain a more stable degree of overfitting and continue to provide an effective loss gradient for the generator, while the APA discriminator exceeds the limit value and gradually falls into complete overfitting, losing the ability to guide the generator. This result is more directly represented in the image quality indicator curve in Fig. 10.

Fig. 10 shows the Fid curve of each model during the AFHQ-Cat training process, as we analyzed above, each model has not learned how to provide an effective loss gradient for the generator at the initial stage, resulting in generally poor image quality. Subsequently, after the discriminator learns the effective error loss, the generator capability is rapidly improved, and the quality of the generated image is rapidly improved. Based on Figs. 9 and 10, we cashback, the degree of overfitting of the discriminator is closely related to the quality of the generated image, within a reasonable overfitting range, the discriminator can improve the ability of the generator to a limited extent, and if the discriminator gradually falls into overfitting like APA, the image quality will gradually decline.

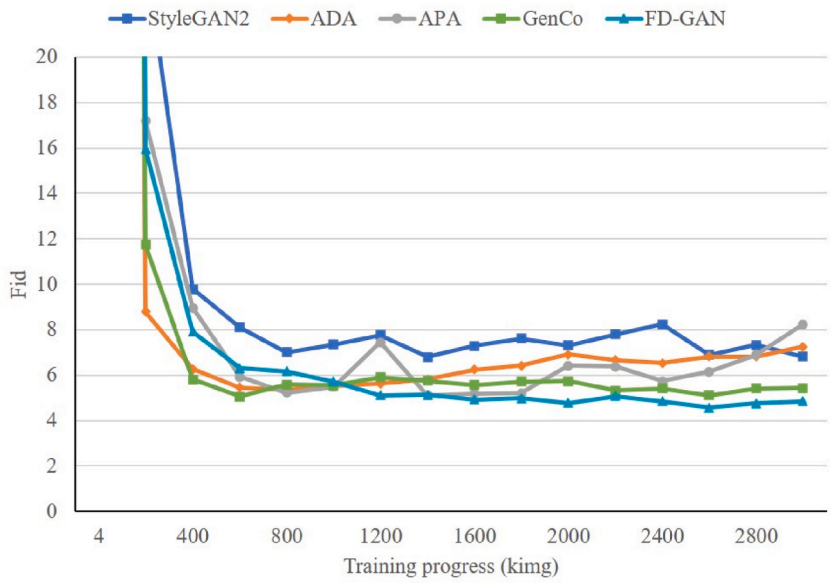


Fig. 10. Training Fid in AFHQ-Cat. Under the condition of sufficient training data, all models had good convergence.

Fig. 11 shows the image generation capability of each model in the AFHQ dataset. Specifically, in (a), there is a certain leakage of the images generated by StyleGAN2, APA, and ADA, and there are certain unreasonable textures in the images, GenCo and FD-GAN perform better, the generated images are of higher quality, and the intuitive feeling is more natural. Similarly, (b) trained in AFHQ-Dog observes a similar situation.

As shown in Table 2, comparing the evaluation indexes horizontally, we find that FD-GAN has reached a higher performance level. Specifically, in the AFHQ-Cat and AFHQ-Dog datasets, Fid reaches 4.580 and 12.007, respectively. Both exceeded 4 comparison models, indicating that the structural design of the dual-ways discriminator effectively stabilized the training process of FD-GAN. In addition, we found that the Precision index of FD-GAN did not reach the optimal in the two datasets, but the comprehensive evaluation index F1 could achieve the best score, indicating that the FD-GAN model found a better balance between precision and recall, thereby

Fig. 11. Images generated from AFHQ. Comparing the training results of the model in 2 benchmarks, each column of images came from the same randomly generated seed, and the clarity of image details and structural integrity response corresponded to the performance of the model.

Table 2
The results in AFHQ in terms of Fid, Precision, Recall, and F1.

Methods	AFHQ-Cat				AFHQ-Dog			
	Fid (↓)	Precision (↑)	Recall (↑)	F1 (↑)	Fid (↓)	Precision (↑)	Recall (↑)	F1 (↑)
StyleGAN2	6.767	0.6167	0.2030	0.3054	25.133	0.5837	0.2577	0.3575
APA	5.353	**0.6713**	0.2603	0.3752	15.240	**0.6813**	0.3330	0.4474
ADA	5.373	0.6390	0.3290	0.4344	19.013	0.6330	0.4447	0.5224
GenCo	5.227	0.6480	0.3387	0.4448	14.697	0.6670	0.4540	0.5403
FD-GAN	**4.580**	0.6270	**0.4133**	**0.4982**	**12.007**	0.6420	**0.5127**	**0.5701**

ensuring a more stable generation level.

As we know, GAN is a non-zero-sum game process in which discriminator and generation carry out adversarial training, and the sign of training completion is that they reach a local saddle point. This means that when the GAN training is completed, the discriminator and the generator reach a consistent judgment result. That is, they cannot successfully distinguish between real and fake images. For the discriminator, it means that it cannot distinguish between real and fake images more effectively, so the probability of false judgment increases. In a word, the recall of the discriminator increases. At the same time, F1 is a comprehensive evaluation indicator of the discriminator, which is used to quantify the degree of overfitting, and it examines the precision and recall of the discriminator at the same time. As we pointed out earlier, the abundant training data can effectively alleviate the risk of overfitting the discriminator, and the improvement of F1 also proves that the ASP frequency domain enhancement module proposed by us has played a significant role.

5.6. Ablation experiment

To comprehensively test the reliability of the FD-GAN model, this section conducts self-ablation comparison experiments on the regular constraint coefficient, and quantifies the influence on the quality of the generated images. Considering the model training effect and time consumption, all experiments were carried out in the TankDataSet dataset with a training time of 3000 kimg.

In this paper, FD-GAN is adopted from StyleGAN2, and the regularized loss function based on the R1 constraint is adopted. That is the model weight is compressed by punishing the gradient of the discriminant loss of the real image to achieve the purpose of alleviating overfitting, and the calculation formula is:

$$R1 = \frac{\lambda}{2} E_{x\sim p_x}\left[\|\nabla D(x)\|^2\right] = \frac{\lambda}{2} E_{x\sim p_x}\left[\left\|\frac{\partial D(x)}{\partial x}\right\|^2\right] \tag{12}$$

Among them, the R1 constraint weight coefficient, the λ numerical size affects the sparsity degree of the model, that is, the change range of the control weight parameters. To analyze the influence of the regular constraint loss function on the model performanceλ, grid searching between 10, 1 and 0.1 is used for experiments.

The curves in Fig. 12 (a) and (b) reflect the training process of FD-GAN, and it can be seen that FD-GAN converges stably under all the weight coefficients. However, different coefficients have different performances in constraint weight sparsity. That is a larger constraint coefficient can loosen the weight matrix faster, but excessive sparsification will lead to a decrease in the learning ability of the model, which is reflected by insufficient ability to learn image details, resulting in a decrease in image quality. As shown in (c), the quality curve of the resulting image is relatively poor when $\lambda = 10$ and $\lambda = 0.1$.

Table 3 shows that the λ does have a significant impact on the quality of the generated images. That is to say when the λ increase, the weight parameters fluctuate in a smaller range, and the purpose of thinning can be achieved faster; However, as λ decreases, irrelevant gradients will accumulate and cannot be effectively eliminated, increasing the risk of overfitting the model. The overly controlled weight constraints will lead to a decrease in the quality of the generated images. Based on the above factors, this article adopts the condition setting of $\lambda = 1$.

6. Conclusion

The current GAN cannot adapt to the training conditions of limited data, resulting in low model performance. Through experimental analysis, we pointed out that the overfitting problem of the discriminator is the root cause. To this end, the dual-ways FD-GAN is redesigned, who is composed of two discriminators to alleviate the overfitting problem of the discriminator by improving the model structure. At the same time, in order to better enrich the training data, a dynamic filtering module is proposed with significant structural enhancement based on frequency space. Furthermore, a limited image of a military target is specifically collected to verify the performance of FD-GAN in the special scenario. In addition, comparative experiments and ablation experiments on 2 publicly available limited datasets is handled. A large number of experimental results show that FD-GAN achieves excellent performance with the assistance of dual-ways discriminator structure and adaptive structure filtering module.

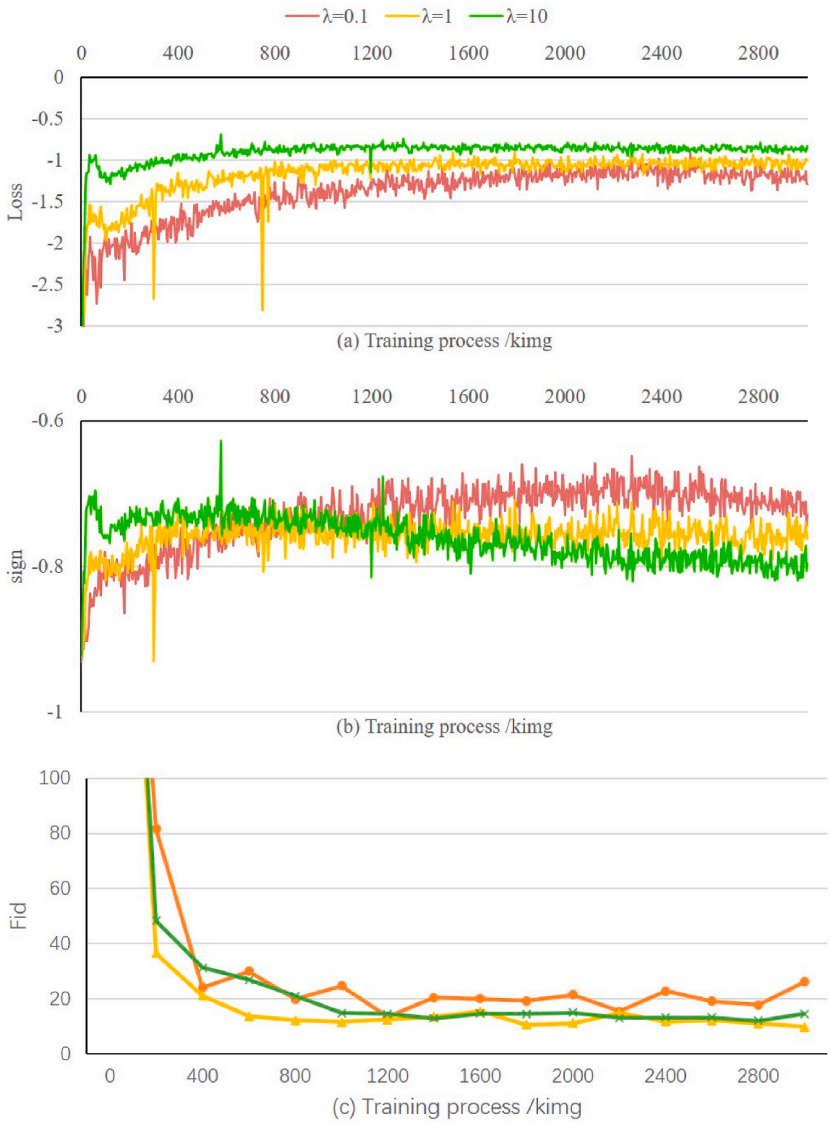


Fig. 12. The curves of the different regular constraint coefficients. The R1 loss function has a significant impact on the overall loss of the model, and the performance of the model can be fine-tuned by controlling its weight parameters.

Table 3
The impact of different regular constraint coefficients.

Metrics	$\lambda = 10$	$\lambda = 1$	$\lambda = 0.1$
Fid (↓)	17.7004	**10.3821**	16.9980
Precision (↑)	0.5750	0.6299	**0.7339**
Recall (↑)	0.1513	**0.2432**	0.0481
F1 (↑)	0.2396	**0.3501**	0.0904

Funding statement

The authors received no specific funding for this study.

Data availability statement

The full data will be available on request.

CRediT authorship contribution statement

Jian Wei: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Qinzhao Wang:** Resources, Project administration. **Zixu Zhao:** Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Z. WuHebrew, D. Lischinski, E. Shechtman, Stylespace analysis: disentangled controls for stylegan image generation, in: Proc. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2021, pp. 12864–12873.
[2] Y. Choi, M. Choi, M. Kim, et al., Stargan Unified Generative Adversarial Networks for Multidomain Imagetoimage Translation, 2018, pp. 8789–8797.
[3] Y. Choi, Y. Uh, J. Yoo, et al., Stargan v2: diverse image synthesis for multiple domains, in: Proc. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020, pp. 8188–8197.
[4] Y. Alaluf, O. Tov, R. Mokady, et al., Hyperstyle: stylegan inversion with hypernetworks for real image editing, in: Proc. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2022.
[5] Y. Alaluf, O. Patashnik, D. Cohen-Or, Restyle a residual-based stylegan encoder via iterative refinement, in: Proc. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2021, pp. 6711–6720.
[6] R. Abdal, Y. Qin, P. Wonka, Image2stylegan++ how to edit the embedded images, in: Proc. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020.
[7] H. Xia, H. Zhao, Z. Ding, Adaptive Adversarial Network for Source-free Domain Adaptation, IEEE, 2021, pp. 8990–8999.
[8] Z. Wang, L. Zhao, H. Chen, et al., Diversified arbitrary style transfer via deep feature perturbation, in: Proc. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020, pp. 7789–7798.
[9] Y. Wang, A.G. Garcia, D. Berga, et al., Minegan ：effective knowledge transfer from gans to target domains with few images, in: Proc. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020, pp. 9332–9341.
[10] F. Pizzati, P. Cerri, R. de Charette, Comogan: continuous model-guided image-to-image translation, in: Proc. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2021, pp. 14288–14298.
[11] T. Park, A.A. Efros, R. Zhang, et al., Contrastive Learning for Unpaired Image-To-Image Translation, arxiv, 2020.
[12] D. Kotovenko, A. Sanakoyeu, A content transformation block for image style transfer, in: Proc. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019, pp. 10032–10041.
[13] K. Kim, S. Park, E. Jeon, et al., A style-aware discriminator for controllable image translation, in: Proc. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2022, pp. 18240–18249.
[14] C. Xiao, B. Li, J. Zhu, et al., Generating adversarial examples with adversarial networks, in: Proc. The 27 International Joint Conference on Artificial Intelligence Main Track, 2018, pp. 3905–3911.
[15] Wu, D., Wang, Y., Xia, S., et al.:'Skip connections matter: on the transferability of adversarial examples generated with resnets', ArXiv, 2020,abs/2002.05990.
[16] A. Sayles, A. Hooda, M. Gupta, et al., Invisible perturbations: physical adversarial examples exploiting the rolling shutter effect, in: Proc. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2021, pp. 14666–14675.
[17] Wang, D., Lin, J., Wang, Y.:'Query-efficient adversarial attack based on Latin hypercube sampling', ArXiv, 2022,abs/2207.02391.
[18] J. Zhang, B. Li, J. Xu, et al., Towards efficient data free black-box adversarial attack, in: Proc. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2022.
[19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., Generative adversarial networks, Commun. ACM 63 (11) (2014) 139–144.
[20] H. Tseng, L. Jiang, C. Liu, et al., Regularizing generative adversarial networks under limited data, in: Proc. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2021, pp. 7921–7930.
[21] C. Yang, Y. Shen, Y. Xu, et al., Improving gans with a dynamic discriminator, in: *Book* Improving Gans with a Dynamic Discriminator, Series Improving Gans with a Dynamic Discriminator, 2022.
[22] S. Zhao, Z. Liu, J. Lin, et al., Differentiable augmentation for data-efficient gan training, in: Proc. 34th Conference on Neural Information Processing Systems (NeurIPS2020), IEEE, 2020.
[23] T. Karras, M. Aittala, J. Hellsten, et al., Training generative adversarial networks with limited data, in: Proc. 34th Conference on Neural Information Processing Systems (NeurIPS2020), IEEE, 2020.
[24] L. Jiang, B. Dai, W. Wu, et al., Deceive d: adaptive pseudo augmentation for gan training with limited data, in: Proc. 35th Conference on Neural Information Processing Systems (NeurIPS2021), IEEE, 2021.
[25] J. Kim, Y. Choi, Y. Uh, Feature statistics mixing regularization for generative adversarial networks, in: Proc. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2022.
[26] K. Cui, J. Huang, Z. Luo, et al., Genco: generative co-training for generative adversarial networks with limited data, in: Proc. 36th AAAI Conference on Artificial Intelligence, 2022, pp. 499–507.
[27] Z. Zhao, S. Singh, H. Lee, et al., Improved Consistency Regularization for Gans, arXiv, 2020, pp. 2002–4724.
[28] L. Jiang, B. Dai, W. Wu, et al., Focal frequency loss for image reconstruction and synthesis, in: Proc. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2021, pp. 13919–13929.
[29] T. Karras, S. Laine, M. Aittala, et al., Analyzing and improving the image quality of stylegan, in: Proc. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020, pp. 8110–8120.
[30] Y. Cao, S. Li, Y. Liu, et al., A comprehensive survey of ai-generated content (aigc): a history of generative ai from gan to chatgpt, in: *Book* A Comprehensive Survey of Ai-Generated Content (Aigc): a History of Generative Ai from gan to Chatgpt, Series A Comprehensive Survey of Ai-Generated Content (Aigc): a History of Generative Ai from gan to Chatgpt, 2023.
[31] M. Xu, H. Du, D. Niyato, et al., Unleashing the power of edge-cloud generative ai in mobile networks: a survey of aigc services, in: *Book* Unleashing the Power of Edge-Cloud Generative Ai in Mobile Networks: a Survey of Aigc Services, Series Unleashing the Power of Edge-Cloud Generative Ai in Mobile Networks: a Survey of Aigc Services, 2023.
[32] Z. Wang, H. Zheng, P. He, et al., Diffusion-gan: training gans with diffusion, in: *Book* Diffusion-gan: Training Gans with Diffusion, Series Diffusion-gan: Training Gans with Diffusion, 2022.
[33] T. Park, A.A. Efros, R. Zhang, et al., Contrastive learning for unpaired image-to-image translation, in: Proc. Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16, Springer, 2020, pp. 319–345.
[34] T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, in: Proc. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019, pp. 4401–4420.
[35] X. Huang, S. Belongie, Arbitrary style transfer in real-time with adaptive instance normalization, arXiv:1703.06868 (2017).

[36] M. Cai, H. Zhang, H. Huang, et al., Frequency domain image translation more photo-realistic, better identity-preserving, in: Proc. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2021, pp. 13930–13939.
[37] T. Karras, T. Aila, S. Laine, et al., Progressive growing of gans for improved quality, stability, and variation, arXiv:1710.10196 (2018).
[38] K. Liu, W. Tang, F. Zhou, et al., Spectral regularization for combating mode collapse in gans, in: Proc. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2019, pp. 6382–6391.
[39] M. He, Y. Wang, J. Wu, et al., Cross domain object detection by target-perceived dual branch distillation, in: Proc. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2022, pp. 9570–9579.
[40] N. Yu, G. Liu, A. Dundar, et al., Dual contrastive loss and attention for gans, in: Proc. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 3742–6731.
[41] A. Prakash, K. Chitta, A. Geiger, Multi-modal fusion transformer for end-to-end autonomous driving, in: Proc. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2021, pp. 7077–7087.
[42] S. Xu, J. Gu, Y. Hua, et al., Dktnet: dual-key transformer network for small object detection, Neurocomputing 525 (3) (2023).
[43] Y. Kittenplon, I. Lavi, S. Fogel, et al., Towards weakly-supervised text spotting using a multi-task transformer, in: Proc. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2022, pp. 4604–4613.
[44] D.J. Chen, H.Y. Hsieh, T.L. Liu, Adaptive image transformer for one-shot object detection, in: Proc. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2021, pp. 12247–12256.
[45] H. Wang, L. Zhou, L. Wang, Miss detection vs. False alarm: adversarial learning for small object segmentation in infrared images, in: Proc. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2021, pp. 8509–8518.
[46] M. Qi, L. Hsin-Ying, T. Hung-Yu, et al., Mode Seeking Generative Adversarial Networks for Diverse Image Synthesis, CoRR, 2019. abs/1903.05628.
[47] R. Liu, Y. Ge, C.L. Choi, et al., Divco: diverse conditional image synthesis via contrastive generative adversarial network, in: Proc. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2021, pp. 16377–16386.
[48] M. Heusel, H. Ramsauer, T. Unterthiner, et al., Gans trained by a two time-scale update rule converge to a local nash equilibrium, in: *Book* Gans Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium, Series Gans Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium, 2017, pp. 6629–6640.
[49] G. Shekar, S. Revathy, E.K. Goud, Malaria detection using deep learning, in: 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI), 2020.
[50] J. Chaki, M. Woźniak, Deep learning for neurodegenerative disorder (2016 to 2022): a systematic review, Biomed. Signal Process Control 80 (2023) 104223–104232.