Information Fusion 91 (2023) 134–148

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

GAN review: Models and medical image fusion applications

Tao Zhou [a,b], Qi Li [a,*], Huiling Lu [c,*], Qianru Cheng [a], Xiangxiang Zhang [a]

[a] School of Computer Science and Engineering, North Minzu University, Yinchuan, 750021, China
[b] Key Laboratory of Image and Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan, 750021, China
[c] School of Science, Ningxia Medical University, Yinchuan, 750004, China

ARTICLE INFO

Keywords:
Generative adversarial network (GAN)
Generative models
Medical image fusion
Probability distribution distance
Multi-generators and multi-discriminators GAN

ABSTRACT

Generative Adversarial Network (GAN) is a research hotspot in deep generative models, which has been widely used in the field of medical image fusion. This paper summarizes GAN models from the following four aspects: firstly, the basic principles of GAN are expounded from two aspects: basic model and training process; secondly, variant GAN models are summarized into three directions (Probability Distribution Distance, Overall Network Architecture, Neural Network Structure), from the methods based on f-divergence, the methods based on IPM, Single-Generator and Dual-Discriminators GAN, Multi-Generators and Single-Discriminator GAN, Multi-Generators and Multi-Discriminators GAN, Conditional Constraint GAN, Convolutional Neural Network structure GAN and Auto-Encoder Neural Network structure GAN are eight dimensions to summarize the typical models in recent years; thirdly, the advantages and application of GAN models in the field of medical image fusion are explored from three aspects; fourthly, the main challenges faced by GAN and the challenges faced by GAN models in medical image fusion field are discussed and the future prospects are given. This paper systematically summarizes various models of GAN, advantages and challenges of GAN models in medical image fusion field, which is very important for the future research of GAN.

1. Introduction

Generative models have been obtained high success in applications, such as image processing, popularity density estimation and image style migration. With the rapid increase of the number and dimensionality of samples, generative models are gradually being replaced by deep generative models that containing multiple hidden layers. Deep generative models have been successfully applied in the fields of computer vision, natural language processing, image generation and semi-supervised learning and a good paradigm is provided for unsupervised learning. There are some common deep generative models, such as Restricted Boltzmann Machines (RBM) [1], Variational Auto-Encoders (VAE) [2], Autoregressive models [3] and Generative Adversarial Network (GAN) [4]. The GAN is proposed by Goodfellow et al [4] in 2014 and it is a current research hotspot in deep generative models. GAN can implicitly estimate the density function of the data distribution and generate the samples set that is consistent with the distribution of the real samples set.

Given the excellent performance of GAN, it has gained wide attention from researchers in the field of medical image fusion. Fu et al [5] proposed a method for anatomical and functional medical image fusion based on a dual-stream attention mechanism, which solves the problems of low efficiency and blurred details in existing multimodal medical image fusion algorithms. Zhan et al [6] proposed a GAN model based on multiscale gate merging (GM), which can automatically learn the weights of different modes at different locations by introducing a GM mechanism and it effectively alleviating the problem of missing or corrupted MRI sequences. Huang et al [7] proposed a Multi-Generators and Multi-Discriminators conditional generative adversarial network based on CGAN [8], which was successfully applied in the field of different types of medical image fusion. Ma et al [9] proposed DDcGAN, a model for multimodal medical image fusion with different resolutions using a Dual-Discriminators architecture. Jiang et al [10] proposed a framework called FA-GAN for generating super-resolution MRI from low-resolution MRI and the scanning time of MRI is reduced. Combining GAN with traditional medical image fusion methods by Wang et al [11], which can effectively suppress the artifacts and distortions and the detail information can be well preserved by the proposed method, as well as the better quantitative performance. Kang et al [12] proposed a method for fusing brain PET and MRI using tissue-aware conditional generation

* Corresponding author.
E-mail addresses: zhoutaonxmu@126.com (T. Zhou), liqi1774646271@163.com (Q. Li), Lu_huiling@163.com (H. Lu), chengqianru5@163.com (Q. Cheng), zxx19990503@163.com (X. Zhang).

https://doi.org/10.1016/j.inffus.2022.10.017
Received 8 September 2022; Received in revised form 7 October 2022; Accepted 17 October 2022
Available online 20 October 2022
1566-2535/

adversarial network, which enhanced the anatomical details of the fused images while effectively retaining the color information of the PET images. Guo et al [13] proposed a multimodal medical image fusion model based on the residual attention mechanism of GAN, which overcomes the drawback of traditional brain image fusion that requires manual setting fusion rules.

In the previous work, Wang et al [14] mainly summarizes the application of GAN models in the field of computer vision; Jeong et al [15] mainly introduces the application of GAN in the medical image classification and segmentation fields; Wali et al [16] illustrates the application of GAN models in the field of speech processing. Different from previous work, this paper mainly summarizes the variant models of GAN, analyzes the advantages of GAN models in the field of medical image fusion, introduces the application status of GAN in the field of medical image fusion, and discusses the main challenges faced by GAN and the main challenges faced by GAN models in medical image fusion field. The overview of the structure of this paper is shown in Fig. 1. Firstly, the basic principles of GAN are described in terms of the basic model and the training process. Secondly, the models of GAN are reviewed from three aspects: Probability Distribution Distance, Overall Network Architecture and Neural Network Structure. According to the different metric methods, Probability Distribution Distance is divided into the methods based on f-divergence [17] and the methods based on IPM (Integral Probability Metric) [18]; according to the number of Generator and Discriminator and the external conditional constraints, Overall Network Architecture is divided into Single-Generator and Dual-Discriminators GAN (SGDD GAN), Multi-Generators and Single-Discriminator GAN (MGSD GAN), Multi-Generators and Multi-Discriminators GAN (MGMD GAN) and Conditional Constraint GAN (CC GAN); according to the type of neural network used in the models, Neural Network Structure is divided into Convolutional Neural Network structure GAN and Auto-Encoder Neural Network structure GAN. Thirdly, the advantages of GAN models in the field of medical image fusion are expounded and the application of GAN models in the field of medical image fusion are discussed from three aspects: Single-Generator and Single-Discriminator GAN (SGSD GAN), Single-Generator and Dual-Discriminators GAN (SGDD GAN) as well as Multi-Generators and Multi-Discriminators GAN (MGMD GAN). Fourthly, the main challenges faced by GAN and faced by GAN models in the field of medical image fusion application are summarized and the future directions of GAN are prospected.

As shown in Fig. 1, the remainder of this paper is organized as follows. Section 2 expounds the basic principles of GAN. In Section 3, the various GAN models are summarized. In Section 4, the advantages and application of GAN in medical image fusion field are analyzed. Section 5 discusses the main challenges faced by GAN and the challenges faced by GAN models in medical image fusion field, and prospects the future directions of GAN, and conclusions follow in Section 6.

2. Basic principles

2.1. Basic model

Inspired by the zero-sum game in game theory, GAN treats the generation problem as an adversarial game between the Generator and the Discriminator. The basic model is shown in Fig. 2.

In the standard GAN model, Multilayer Perceptron (MLP) structure is consisted of Generator and Discriminator. The purpose of Generator is to learn the distribution of the real samples x and generate a generated sample $G(z)$ with a high similarity with the real sample's distribution. The input of Generator is a random noise z obeying some simple samples distribution, such as Gaussian distribution. The output of Generator is $G(z)$ with the same dimensions as real samples x.

Discriminator is a binary classifier with input x or $G(z)$ and output discrimination results (Real or Fake), which is used to calculate the objective function and update the network weights by Backpropagation algorithm (BP algorithm) [19]. The purpose of Discriminator is to discriminate x from $G(z)$ as accurately as possible. Discriminator expects the discrimination result to be Real (Label=1) when its input is x. And Discriminator expects the discrimination result to be Fake (Label=0) when its input is $G(z)$, but Generator expects to be able to "cheat" Discriminator in this time. There is competition and adversarial relationship between Generator and Discriminator.

Compared with other generative models, GAN has the following four advantages: firstly, comparing with the Boltzmann machines, the computational processing of GAN only uses the BP algorithm to

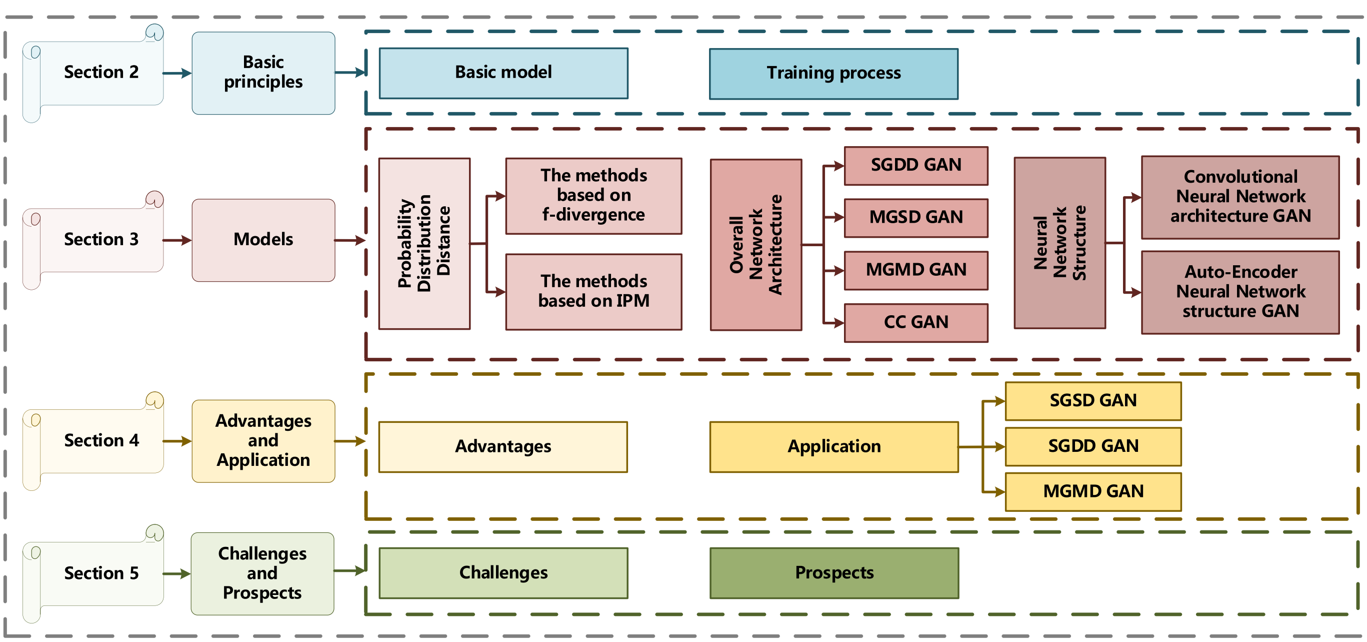

Fig. 1. An overview of the structure of this paper from Sections 2 to 5. SGDD GAN expresses Single-Generator and Dual-Discriminators GAN; MGSD GAN expresses Multi-Generators and Single-Discriminator GAN; MGMD GAN expresses Multi-Generators and Multi-Discriminators GAN; CC GAN expresses Conditional Constraint GAN; SGSD GAN expresses Single-Generator and Single-Discriminator GAN; SGDD GAN express Single-Generator and Dual-Discriminators GAN.

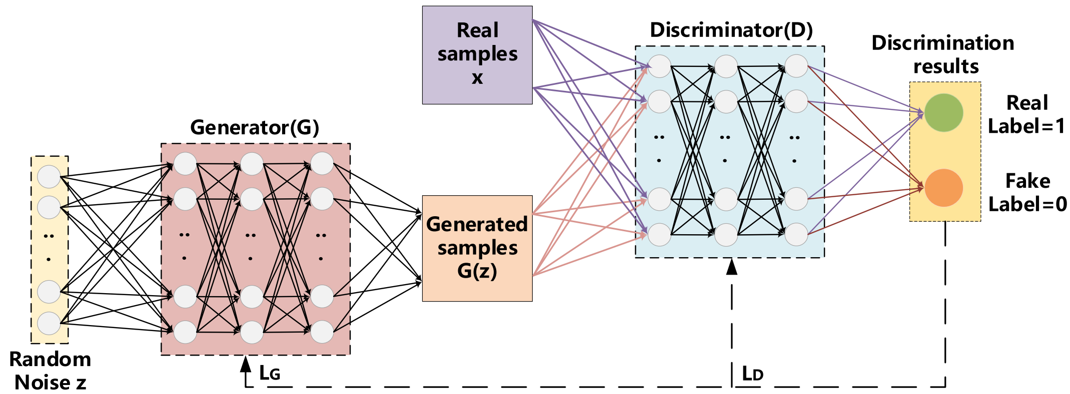

Fig. 2. Basic model. L_G expresses the loss function of Generator; L_D expresses the loss function of Discriminator.

calculate the gradient, hence its computational speed is faster and the learning processing of GAN does not require approximate reasoning; secondly, comparing with the VAE, GAN does not have a variational lower bound and it also can generate clear images without using any bias; thirdly, comparing with the PixelRNN which can only generate a pixel at one time when generating image, GAN can generate data samples in parallel and the time required to generate samples is shorter; fourthly, comparing with the Independent Component Analysis (ICA), the Generators of GAN do not have any restrictions on the size of the inputting data and GAN are able to train various generative networks with a flexible framework.

At the same time, GAN also has the following four drawbacks: the first is that it is difficult to achieve the status of convergence and Nash equilibrium [20]. Only in the case of the objective function is convex function, the gradient descent method can be used to ensure that the model reaches the Nash equilibrium state and it is difficult to achieve a synchronous state between Generator and Discriminator during the training process. The second is the mode collapse. When the generator can generate real samples under certain parameters, its learning ability will decline, resulting in the lack of diversity of generated samples. The third is gradient disappearance. After sufficient training, the Discriminator can always distinguish real samples and generated samples. In this case, the gradient is zero and the Generator cannot continue to learn the distribution of the real samples, resulting in the disappearance of the gradient. The fourth is that GAN is uncontrollable. In the standard GAN model, the input of the Generator only uses random noises and the model cannot be controlled by constraint conditions to generate samples with specified characteristics.

2.2. Training process

The training process of GAN is the binary maximin game process between Generator and Discriminator. The objective function is defined as:

$$\underset{G}{min}\underset{D}{max}V(G,D) = E_{x\sim P_{data}(x)}[logD(x)] + E_{z\sim P_z(z)}[\log(1-D(G(z)))],$$

where *V (G, D)* is a binary cross-entropy function, *E(*)* represents the expected value of the samples distribution function, *x* represents the real samples and $P_{data}(x)$ represents the real samples distribution, *z* represents the random noise which input into Generator and $P_z(z)$ is a random noise distribution, *G(z)* represents the generated samples of Generator, *D(x)* is a probability value that represents the probability that *x* came from a real sample rather than a generated sample. *D(G(z))* is also a probability value that represents the probability that Discriminator will discriminate *G(z)* to be a real sample. The training process of GAN is an alternate iterative training of these two models, with one of the models is fixed during training and then updating the parameters of the other model. First, Discriminator learns the distribution of *x*. When Discriminator has some knowledge of the distribution of *x*, Discriminator is used to discriminate the real-fake of *G(z)*. Then, the ability of Generator to generate samples is improved during the discrimination process of Discriminator; the ability of Discriminator to distinguish is increased in the processing of learning the distribution of *x*. Through continuous adversarial training, the probability value of Discriminator to discriminate real-fake samples are maximized and the similarity degree between *G(z)* and *x* is maximized. Eventually, Generator and Discriminator reach the state of Nash equilibrium. At this time, it can be assumed that Generator learns the distribution of the real samples.

3. Models

Although GAN is a pioneering deep generative model and it breaks the limitation that previous models must learn parameters by the method of maximum likelihood estimation, its training process is usually difficult. Therefore, ever since GAN was proposed, there are many researchers who have conducted in-depth studies on GAN from different aspects. This paper summarizes the representative models since 2014 and introduces them in three aspects: Probability Distribution Distance, Overall Network Architecture and Neural Network Structure, as shown in Fig. 3. The models within each dashed box in the Fig. 3 are the same category models and all the models are arranged in the order of the year above.

3.1. Probability distribution distance

The purpose of GAN is to minimize the distance between the generated samples distribution and the real samples distribution by the processing of adversarial training. Therefore, choosing an appropriate method of probability distribution distance metric has an important impact on the performance optimization of models. The standard GAN model measures the distance between the real and the generated samples distribution through JS-divergence (Jensen-Shannon divergence), which is prone to the problem of gradient disappearance, resulting in unstable training process of the models and poor quality of the generated images. For this reason, different methods of probability distribution distance metric have been proposed by researchers in recent years. The commonly used methods are divided into two categories: the methods based on f-divergence [17] and the methods based on IPM [18], as shown in Fig. 4.

As seen in Fig. 4, the methods based on f-divergence include JS-divergence, PersonX^2 divergence and total variance. The methods based on IPM include Wasserstein Distance [21], MMD (Maximum Mean Discrepancy) [22] and Fisher IPM [23].

3.1.1. The methods based on f-divergence

The f-divergence is a way to measure the differences between the distribution of two samples using a special convex function *f*. f-GAN [17]

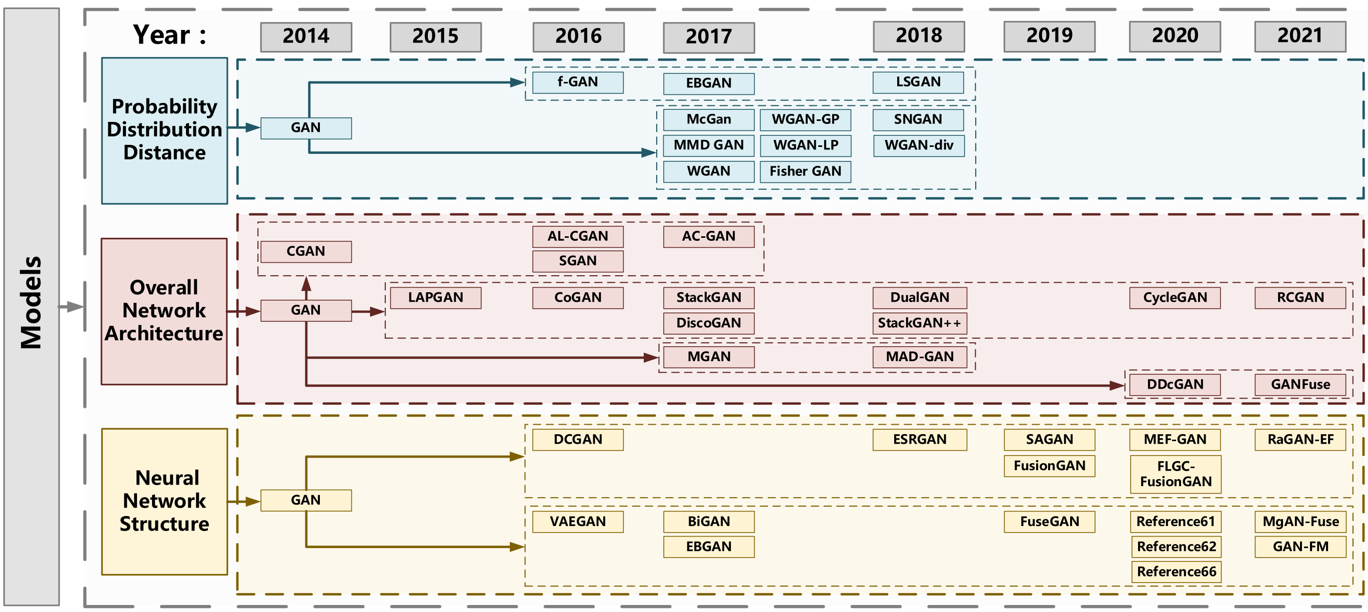

Fig. 3. Models.

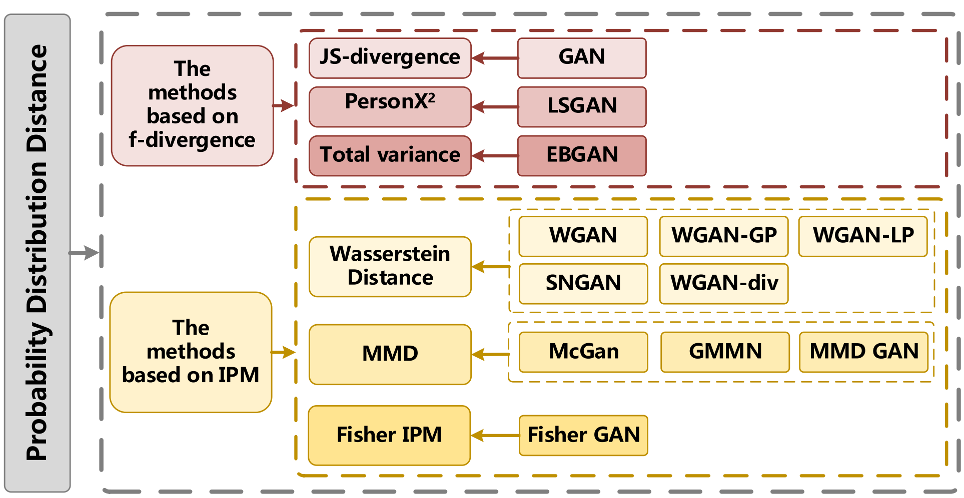

Fig. 4. Probability distribution distance.

defines generalized forms of various divergences, which are defined under the f-divergence framework. Given any two distributions P and Q, the continuous density functions of these two distributions are $p(x)$ and $q(x)$, respectively. Based on the ratio between the two distributions, f-divergence can be defined as:

$$D_f(P \parallel Q) = \int_x q(x) f\left(\frac{p(x)}{q(x)}\right) dx,$$

where f is a convex function defined on the domain X, the lower part of the function image is continuous and satisfies $f(1)=0$. The f-GAN [17] model considers the JS-divergence used by the standard GAN model as a special case of f-divergence and the objective function of GAN defined with f-divergence can be expressed by the Fenchel covariance function of the convex function, which is defined as:

$$\min_G \max_D E_{x\sim P_{data}(x)}[D(x)] + E_{z\sim P_z(z)}[\log(1 - D(G(z)))].$$

As a result, different probability distribution distance metric methods are obtained by selecting f-function that satisfy different conditions. In this way, various GAN models can be obtained and the stability of the training process can be improved. The typical models include LSGAN [24] and EBGAN [25].

LSGAN [24] uses PersonX^2 divergence for probability distribution distance metric, which is a method under the f-divergence framework. The model replaces the cross-entropy loss function used in the standard GAN model with a least-squares loss function to effectively alleviate the problem of gradient disappearance. The training process of LSGAN is more stable and it also can generate higher quality images. EBGAN [25] uses the total variance based on the method of f-divergence for the probability distribution distance metric of two distributions. The model views the discriminator as an energy function that provides high energy for the generated samples and low energy for the real samples, which can directly discriminate the reconstruct ability of the input images. It also can effectively improve the quality of the generated images and the stability of the training process.

3.1.2. The methods based on IPM

The methods based on IPM [18] can get different models by choosing different evaluation function $f(x)$. It first defines a set of real-valued functions named F, whose values are all real numbers and then it looks for a function $f(x)$ in the set F that maximizes the difference between the two distributions which is defined as:

$$D_f\left(P_{data} \parallel P_g\right) = \sup_{f \in F} E_{x\sim P_{data}(x)}[f(x)] - E_{x\sim P_g(x)}[f(x)],$$

which is the distance between the two distributions. The methods based on IPM are independent of the data dimensionality and always converge to the true distance between the distribution of two samples. The main models based on IPM methods include Wasserstein Distance [21], MMD [22] and Fisher IPM [23].

(1) *Wasserstein distance*

Wasserstein Distance [21], also known as the EM distance, is used to indicate the degree of similarity between the distribution of two samples. It measures the minimum value of the average distance required to move the data from the real data distribution P_r to the generated data distribution P_g. Wasserstein Distance can alleviate the problem of gradient disappearance and it is formalized as:

$$W\left(P_r, P_g\right) = \inf_{\gamma \epsilon \Pi\left(P_r, P_g\right)} E_{(x,y)\sim\gamma}[||x-y||],$$

where $\prod(P_r, P_g)$ denotes the joint probability distribution of all edge distributions as P_r and P_g. Typical models based on Wasserstein Distance include WGAN [21], WGAN-GP [26], WGAN-LP [27], WGAN-div [28] and SNGAN [29].

WGAN [21] uses the EM distance to measure the difference between the distribution of two samples instead of the JS distance used in the standard GAN model. Furthermore, it uses the method of weight clipping to impose Lipschitz constraints on the discriminating function. When the Discriminator is trained to its best, it is able to calculate the difference between P_r and P_g. The training process of WGAN is more stable than the standard GAN model and it can effectively alleviate the problem of gradient disappearance, but the capability to discriminate of the Discriminator is reduced. WGAN-GP [26] replaces the method of weight clipping with gradient penalties method, which is defined as:

$$Gradient\ penalty = \lambda E\widehat{x}\left[\left(||\nabla_{\widehat{x}} D(\widehat{x})||_2 - 1\right)^2\right],$$

which does not affect the capabilities of the discriminator, but the computational cost of this model is higher.

WGAN-LP [27] replaces weight clipping method using the method of regularization, which can be defined as:

$$E_{y\sim v}[f(y)] - E_{x\sim\mu}[f(x)] + \lambda E_{\widehat{x}\sim r}\left[\left(\max\{0, ||\nabla f(\widehat{x})|| - 1\}\right)^2\right],$$

and this regularization term implements the Lipschitz constraint by punishing the deviation between the norm of the gradient of the discriminating function and 1, which improves the stability of the training process to some extent.

WGAN-div [28] introduces a method of Wasserstein divergence that does not need to satisfy the Lipschitz condition while maintaining the Wasserstein Distance, solving the problem that the k-Lipschitz constraint is very strict and not easily approximated.

Based on the idea of spectral normalization, SNGAN [29] normalizes the parameter matrix of the neural network to a formula of the spectral norm whose value is constant equal to 1, realizing the Lipschitz constraint on the Discriminator. This method is simple to implement and requires fewer additional parameters.

(2) *MMD*

MMD [22] is a method of measuring the difference between two distributions P and Q. It is given by the super-confide on the function space F, which is about the difference of expected values between two distributions. MMD is formalized as:

$$MMD(F, P, Q) = \sup_{f \in F}(E_{X\sim P}[f(X)] - E_{Y\sim Q}[f(Y)]).$$

Typical models include McGan [30], GMMN [31] and MMD GAN [32]. McGan [30] combined the idea of MMD with the methods of IPM. It embeds data distributions into a feature space of finite dimensions. Based on the mean and covariance characteristics, the model combines and matches first-order and second-order statistics to minimize the meaningful losses and allow for better capture of various distributions.

GMMN [31] replaces the Discriminator of the standard GAN model with a problem of two-sample detection, using a backpropagation algorithm directly on the basic of MMD to minimize the MMD distance and use this as the objective function of the Generator. It does not need to approximate the distance between two probability distributions by training the neural network. But compared to the standard GAN model, it requires a large number of parameters during the training process and the computational efficiency is not ideal. Therefore, MMD GAN [32] introduces the adversarial idea based on GMMN and combines GMMN with GAN. The Discriminator improves the quality of the GMMN-generated samples by learning the kernel function to maximize the mean difference on the Reproducing Kernel Hilbert Space (RKHS) and using the difference value as the distance between the two distributions for the Generator to train.

(3) *Fisher IPM*

Fisher GAN [23] combines the ideas of Fisher's linear discriminant analysis method on the basis of the framework of IPM, which imposes constraints on the second-order matrix of the Discriminator to ensure the boundedness of the measure. Fisher GAN can make stable and time-efficient training without the need for data-independent constraints and without the need for normalization in batches.

3.2. *Overall network architecture*

The standard GAN model generates images through a confrontation game between a Generator and a Discriminator without external condition constraints. With the in-depth study of the overall network architecture of GAN models, researchers found that by increasing the number of Generators and Discriminators, the quality of generated images can be improved and the problem of multiple inputs can be better dealt with. In addition, by adding external condition constraints, the required images can be generated. Therefore, this section will introduce the Overall Network Architecture from the following two aspects: the number of Generators and Discriminators as well as external condition constraints. In the number of Generators and Discriminators, it mainly includes SGDD GAN, MGSD GAN and MGMD GAN. In the study of the external condition constraints, the quality of the generated image can be improved by adding external condition constraints such as category labels and semantic labels. Therefore, this paper divides the research based on Overall Network Architecture into four areas, as shown in Fig. 5. The first class is SGDD GAN, as shown in Fig. 5(A), with typical models DDcGAN [9] and GANFuse [33]. The second class is MGSD GAN, as shown in Fig. 5(B), with typical models MAD-GAN [34] and MGAN [35]. The third class is MGMD GAN, in which the class is divided into Cascade-structure GAN, Parallel-structure GAN and Cycle-structure GAN according to the internal logic structure of Generators and Discriminators, as shown in Fig. 5(C). The typical Cascade-structure GAN models include StackGAN [36], StackGAN++ [37] and LAPGAN [38]; the typical Parallel-structure GAN models are CoGAN [39] and RCGAN [40]; the typical Cycle-structure GAN models include CycleGAN [41], DualGAN [42]and DiscoGAN [43]. The fourth class is Conditional Constraint GAN(CC GAN), as shown in Fig. 5(D) and typical models include CGAN [8], AL-CGAN [44], SGAN [45] and AC-GAN [46].

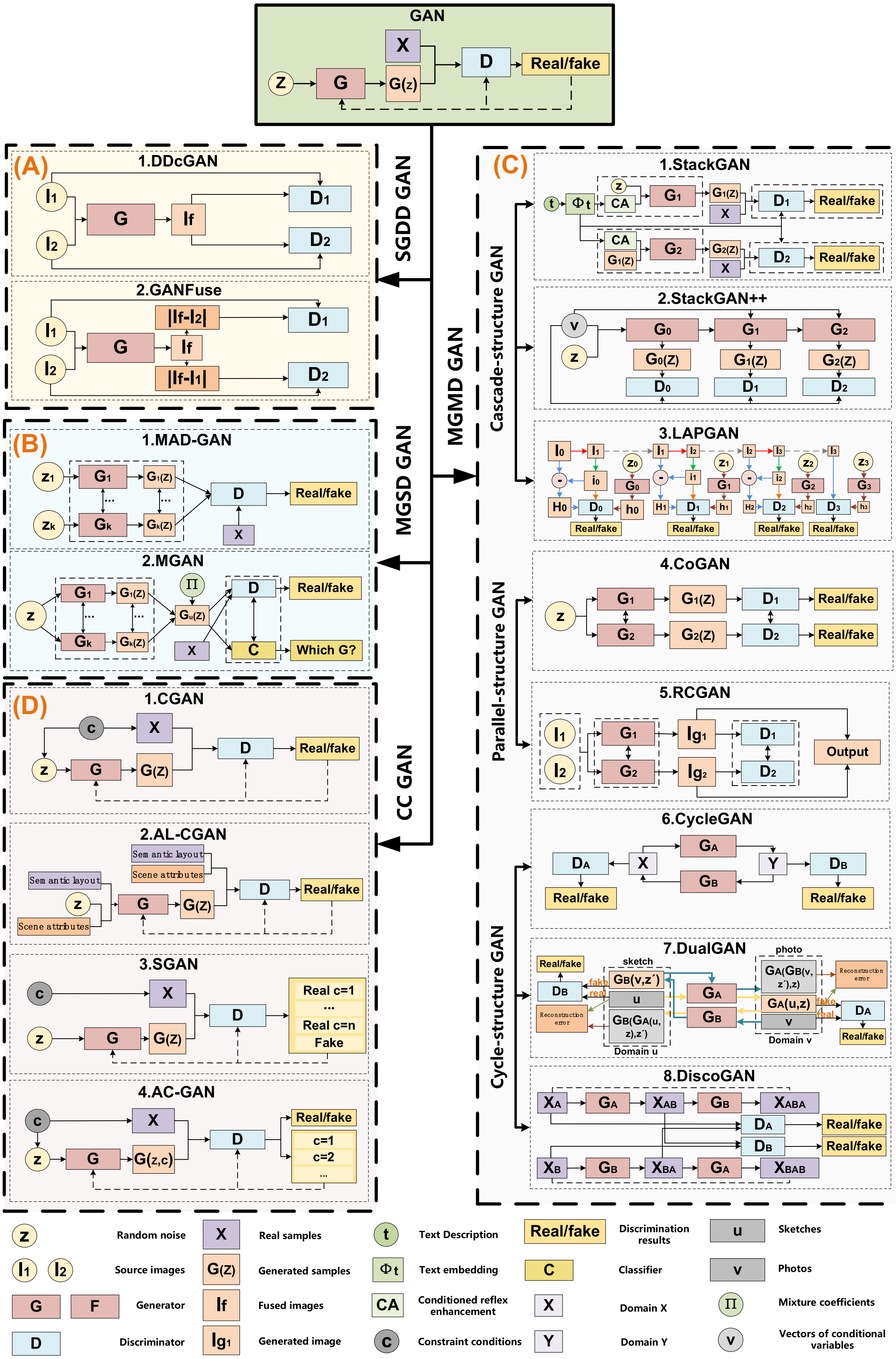

Fig. 5. Overall network architecture. SGDD GAN expresses Single-Generator and Dual-Discriminators GAN; MGSD GAN expresses Multi-Generators and Single-Discriminator GAN; MGMD GAN expresses Multi-Generators and Multi-Discriminators GAN; CC GAN expresses Conditional Constraint GAN.

3.2.1. SGDD GAN

SGDD GAN refers to a model that uses one Generator and two Discriminators for the field of image fusion. DDcGAN [9] is a typical model of this architecture, as shown in Fig. 5(A)-1. Compared to the standard GAN model that inputs the generated sample $G(z)$ directly into a Discriminator for discrimination, DDcGAN inputs the obtained fused image I_f into D_1 and D_2 for discrimination, respectively. It establishes an adversarial game between a Generator and two Discriminators. By using two Discriminators, the Generator can obtain more fully training and better fused image can be obtained. GANFuse [33] further processes the fused image I_f on the basis of the network architecture of DDcGAN, as shown in Fig. 5(A)-2. The model uses $|I_f\text{-}I_2|$ and $|I_f\text{-}I_1|$ as inputs to D_1 and D_2, respectively. It optimizes the Generator by training two Discriminators, which can improve the ability of Generator and Discriminators of the model at the same time.

3.2.2. MGSD GAN

MGSD GAN refers to a model that uses multiple Generators and a Discriminator for alleviating the problem of pattern collapse. Broadly speaking, there are two ways to alleviate this problem: (1) improving learning methods and (2) using multiple Generators on a network architecture. MAD-GAN [34] is a typical MGSD GAN model, as shown in Fig. 5(B)-1. First, k random noises are input into k Generators to get k generated samples, where k Generators share the initial layer parameters. Then k generated samples are used as input to the Discriminator. Similar to the standard GAN model, the purpose of each Generator is to maximize the errors of the general Discriminator. By using multiple Generators that share initial layer parameters not only can reduce redundant computations, but also can effectively avoid the problem of pattern collapse. Compared with the model of MAD-GAN, MGAN [35] randomly selects a $G_u(z)$ from k generated samples as the input to the Discriminator, as shown in Fig. 5(B)-2. By training k Generators with shared parameters at the same time, the problem of pattern collapse can be effectively alleviated.

3.2.3. MGMD GAN

MGMD GAN refers to some models that use multiple Generators and multiple Discriminators for the tasks of image generation, image fusion and image translation. The adversarial learning between multiple Generators and multiple Discriminators can improve the performance of Generators and Discriminators. According to the logical structure between the Generators and Discriminators, the models of MGMD GAN is divided into Cascade-structure GAN, Parallel-structure GAN and Cycle-structure GAN, as shown in Fig. 5(C).

(1) *Cascade-structure GAN*

Cascade-structure GAN model are some models that can generate images in stages and the model can gradually generate high-resolution images from low-resolution images. StackGAN [36] is a typical model with two Generators and two Discriminators, as shown in Fig. 5(C)-1. In this model, the processing of generating images is a two-stage cascade process. The first stage draws the basic shape and color of the image according to the given text description t and draws the background layout of the image according to random noise z. In the first stage can generate low-resolution image $G_1(z)$. The second stage corrects the defects in $G_1(z)$ and reads the text description again to complete the detail drawing. Through the two-stage cascade process that ultimately generate a high-resolution image $G_2(z)$. This model improves the quality of the generated images. However, its training process is unstable, so the StackGAN++ [37] improves it with three Generators and three Discriminators, changing the two-stage stacking structure of StackGAN into a tree structure, as shown in Fig. 5(C)-2. LAPGAN [38], on the other hand, uses four Generators and four Discriminators to generate high-resolution images step by step based on the idea of the Laplacian pyramid [47], as shown in Fig. 5(C)-3. This model trains a pair of Generators and Discriminators on each layer of the pyramid, where the Generators on each layer only learn differential images from the previous layer to that layer. This reduces what the Generator needs to learn each time and it is easy for Generators to learn from the module of residuals.

(2) *Parallel-structure GAN*

Parallel-structure GAN models typically contain two Generators and two Discriminators. CoGAN [39] is a typical model of this architecture, as shown in Fig. 5(C)-4. The random noise z is input to the Generators G_1 and G_2 that share the interlayer weights, respectively. And the generated images $G_1(z)$ and $G_2(z)$ are input to the Discriminators D_1 and D_2 that share the interlayer weights, respectively. By training two GAN models with shared weights at the same time, CoGAN solves the problem of joint distribution learning of multi-domain images without paired training data. RCGAN [40] implemented the image fusion task using a parallel-structure GAN model, as shown in Fig. 5(C)-5. Through the adversarial training of the coupled Generators and the coupled Discriminators, a high-quality fusion image can be obtained. Compared with other models used for image fusion tasks, this model averages the two images obtained as the final fused image, avoiding the problem that traditional fusion methods require complex fusion rules to be designed manually and accelerating the convergence speed of the network.

(3) *Cycle-structure GAN*

Cycle-structure GAN models use two Generators and two Discriminators for the task of image-to-image translation. CycleGAN [41] is a typical cycle-structure model, as shown in Fig. 5(C)-6. By learning the bidirectional mappings $G_A{:}A\text{->}B$ and $G_B{:}B\text{->}A$, this model achieves the translation task of images between domain A and domain B, which can be expressed as $G_B(G_A(x))=x$ and $G_A(G_B(y))=y$. DiscoGAN [43] directly implements bidirectional mapping via $G_A(G_B(x_A))=x_A$ and $G_B(G_A(x_B))=x_B$ as shown in Fig. 5(C)-8. DualGAN [42], inspired by the ideas of Dual learning [48], considers the problems of translation between two images from a closed-loop perspective, thus realizing the mutual translation of sketch image u and photo v. As shown in Fig. 5(C)-7, Generator G_A is used to translate u to v and Generator G_B is used to translate v to u. First, the random noise z and u are used as inputs to G_A to get $G_A(u,z)$ and then it is used with z' as inputs to G_B to get $G_B(G_A(u,z),z')$ to achieve the translation from u to v and the same for the translation from v to u. Although the network architecture of these three models is similar and all implement image translation tasks, from the perception of experimental tasks, CycleGAN and DualGAN focus on image translation tasks between two domains, of which DualGAN contains more types of images, while DiscoGAN focuses on how to avoid the problem of pattern collapse and improve the quality of the generated samples.

3.2.4. CC GAN

CC GAN are some models that add multiple forms of conditional constraints such as object names, bounding boxes or key point locations on the basis of the standard GAN model. By adding constraints to the standard GAN model, the problem of uncontrollable the generated image patterns can be alleviated and it can improve the stability of model training process and the quality of the generated images. CGAN [8] is a typical model of CC GAN, as shown in Fig. 5(D)-1. This model adds conditional constraint c to both the Generator and the Discriminator based on the standard GAN model, where c can be any type of auxiliary information, such as category labels or data from other patterns. It can adjust the model through the conditional constraint c, which guides the processing of generating the data. AL-CGAN [44] uses the layout of semantic vectors and scene attributes, such as sunlight and fog, as conditional vectors to generate realistic outdoor scenes (ocean, mountains, urban scenes) under various conditions such as sunny or cloudy, as shown in Fig. 5(D)-2. The model uses specified scene attributes as conditional constraints to provide coding directions for the global appearance characteristics of the scene. And it can draw the boundaries of the scene under the guidance of the semantic layouts, resulting in a diverse scene with rich details, vivid colors and clear boundaries. SGAN [45] only adds conditional constraints to the real sample *x*, as shown in Fig. 5(D)-3. By using the category label *c*, the Discriminator is forced to output the category label of the real samples while discriminating the real-fake samples. Based on the implementation of the binary classification pattern of the traditional models, SGAN becomes the pattern of multi-classification, which improves the quality of the generated images and reduces the training time of the Generator. AC-GAN [46], on the other hand, combines the ideas of CGAN and SGAN by adding category labels *c* to both *z* and *x*. As shown in Fig. 5(D)-4, the Discriminator discriminates the real-fake images and the category it belongs to base on the auxiliary classifier. By using the loss function value of the corresponding category, the Generator can generate a high-resolution image toward the target category. The model effectively alleviates the problem of pattern collapse and implements the function of image classification.

3.3. Neural network structure

With the rapid development of models based on deep learning, how to design the internal neural network structure of generators and discriminators also has a very important impact on the performance of GAN models. At present, the common neural network structure models mainly include Convolutional Neural Network structure and Auto-Encoder Neural Network structure. Thus, this section will introduce the Neural Network Structure of GAN models from the following two aspects: Convolutional Neural Network structure GAN and Auto-Encoder Neural Network structure GAN. As shown in Fig. 6, the left branch is the Convolutional Neural Network structure GAN and the right branch is the Auto-Encoder Neural Network structure GAN, which are divided into four categories according to the position of encoder and decoder: the first is Generator encoding-decoding GAN, the second is Discriminator encoding-decoding GAN, the third is Variational Auto-Encoder GAN (VAE GAN) and the fourth is Inverse mapping encoding GAN.

3.3.1. Convolutional neural network structure GAN

Convolutional Neural Network (CNN) is a deep feed-forward neural network that is mainly cross-stacked by convolutional layers, pooling layers and fully connected layers. It has three structural characteristics: local connection, weight sharing and pooling, which give CNN a certain degree of translation invariance, scale invariance and rotation invariance. The network structure is widely used in the field of image processing. By using CNN, better quality images can be generated and the problems in the standard GAN model can be solved. DCGAN [49] is a typical Convolutional Neural Network structure GAN model, as shown in Fig. 6(A)-1. The model first uses CNN to replace the structure of Multilayer Perceptron used in the standard GAN model, eliminates the pooling layer and fully connected layer used in CNN and uses Batch Normalization (BN) in Generators and Discriminators. It can improve the stability of the standard GAN model and the quality of the generated samples. The typical models also contain ESRGAN [50], RaGAN-EF [51], FusionGAN [52], FLGC-FusionGAN [53], SAGAN [54] and MEF-GAN [55]. Due to the larger capacity and easier training characteristics of the residual dense blocks, ESRGAN [50] uses *B* residual dense blocks that remove all BN layers as the basic block of the Generator network, as shown in Fig. 6(A)-2. This model can reduce the computational complexity and memory usage while improving the generalization capabilities of the models. Based on RaGAN [56], it uses a relative average Discriminator, which no longer directly discriminate whether an image is realistic, but learns to discriminate whether one image is more realistic than another. By predicting relative authenticity, the model can guide the Generator to generate more realistic texture details. RaGAN-EF [51] uses a multi-exposure image sequence as input to a Generator consisting of convolutional layers whose weights are shared between them. This model solves the problem of uncertain number of input image sequences in multi-exposure image fusion algorithm, as shown in Fig. 6(A)-3. The Generator simply stacks multiple convolutional layers on the basis of the method of VDSR [57] and adds the input image and the residual image at the end of the network to obtain the final output image $G(z)$. The Discriminator uses a 9-layer convolution to discriminate $G(z)$ and the real sample *x*. The model can directly fuse multiple source images with different exposure levels and it also can directly generate high-quality images that can be displayed on a common device without further processing.

FusionGAN [51] is the first model using GAN to achieve the task of infrared and visible images fusion, as shown in Fig. 6(A)-4. Both the Generator and Discriminator of this model use a CNN with a 5-layer structure. By extracting the feature mapping of the input image by convolutional layer, this model can improve the diversity of the generated images and follow the rules of DCGAN for batch normalization, which can effectively avoid the problem of gradient disappearance and make the training process of the model more stable. It is further improved by FLGC-Fusion GAN [53], as shown in Fig. 6(A)-5. The Generator of this model uses a residual dense block as the basic network unit, which is similar to the structure used in the model of ESRGAN removing the BN layer, which can prevent delayed computation of the deeper network training process. The Discriminator is designed on the basis of the VGG network [58], which adds a BN layer after each convolutional layer to improve the training speed of the model and performance of the network. In addition, the model adds another convolutional layer at the end of the network to reduce the size of the network. This model can enhance the significance of the target information in the fused images and can complete a variety of different image fusion tasks.

Although CNN-based GAN models have been widely used in various fields, the learning of long-term dependence on the local receptive field in convolutional operators is limited. The presentation capacity of the network can be improved by increasing the size of the convolutional kernel, but the computational and statistical efficiency of the models obtained by using the local convolutional structure is reduced. The Self-Attention Mechanism [59] shows a better balance between the establishment of long-term dependence capacity and the efficiency of computational and statistical. The self-attention module calculates the response of a location as a weighted sum of all location features and the calculation consumption of the weights is small. Therefore, SAGAN [54] introduces the Self-Attention Mechanism into the Convolutional Neural Network structure GAN, as shown in Fig. 6A-(6), where the Generator and the Discriminator add self-attention modules that help to establish long-distance and multi-level dependencies between image regions. With the module of self-attention, the Generator can generate a detailed image by using cues from all feature locations and the Discriminator can discriminate whether the features of the detailed part of the image are consistent with each other. This model not only takes into account global information at each layer of the network structure, but also does not introduce too many parameters. It can effectively achieve a balance between increasing the receptive field and reducing the number of parameters. In addition, in the field of multi-exposure image fusion, the detail information captured in each image will change greatly with the different spatial positions because the image has different exposure degrees. Therefore, MEF-GAN [55] introduces the Self-Attention Mechanism into the Generator, as shown in Fig. 6A-(7), which can solve the problem of local distortions or inappropriate representations in the fusion images. As a result, this model can improve the performance of the fused images.

3.3.2. Auto-encoder neural network structure GAN

Auto-Encoder Neural Network [60], also known as Auto-Encoder. Its basic network structure includes an Encoder (Enc), a Hidden Layer and a

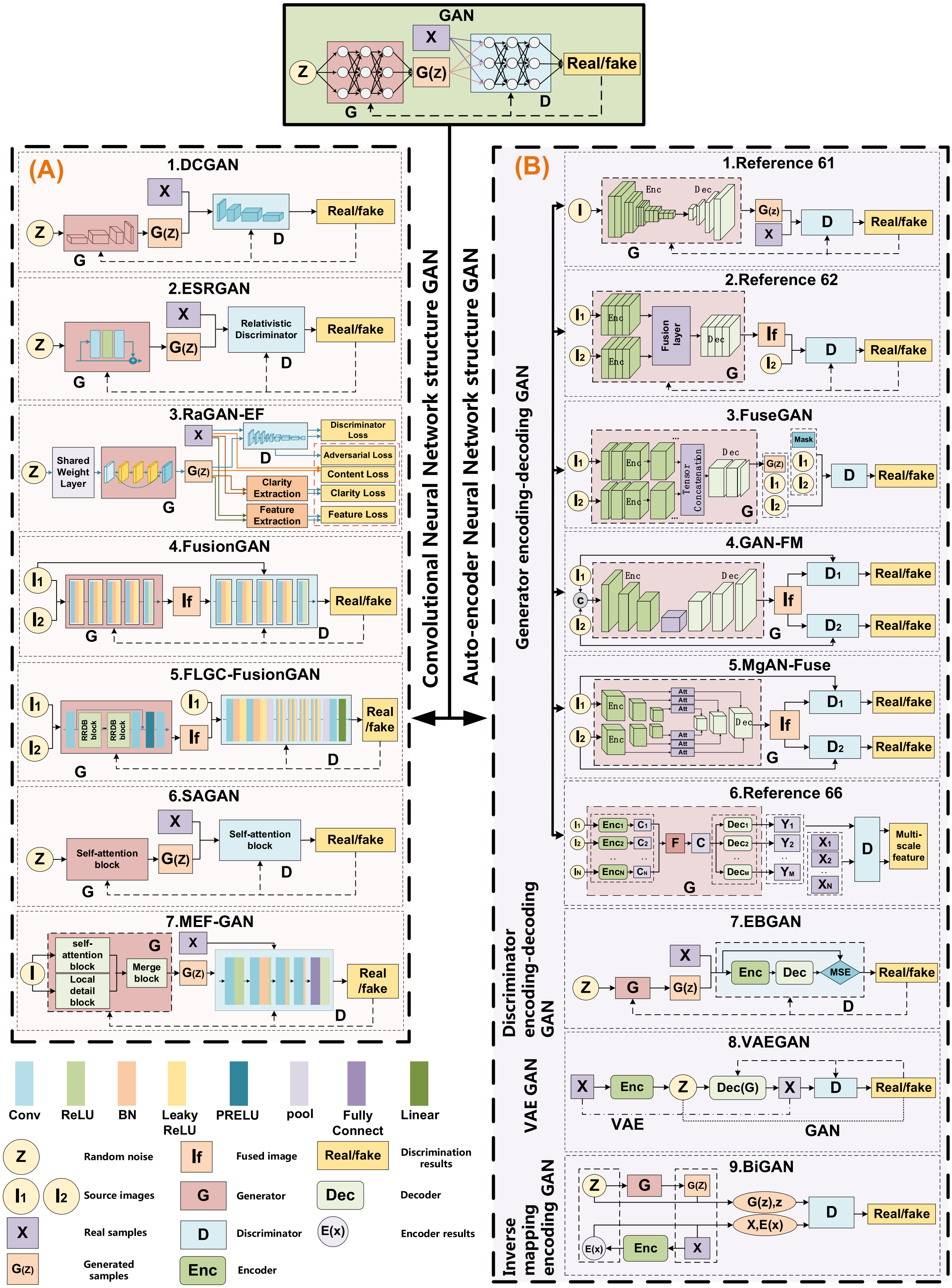

Fig. 6. Neural network structure.

Decoder (Dec), which is a neural network that uses BP algorithm to make the output values equal to the input values. It can automatically learn feature information from unlabeled data and can give better feature descriptions than the original data, with the strong ability of feature learning. Combining the Auto-Encoder structure with the GAN models, we can extract the images features using Encoder and reconstruct the images using Decoder, which can not only learn from the sample images automatically, but also can extract the feature information of the sample images.

As shown in Fig. 6(B), four aspects of the Auto-Encoder Neural Network structure GAN models have been studied in depth by researchers from Generator encoding-decoding GAN, Discriminator encoding-decoding GAN, VAE GAN and Inverse mapping encoding GAN.

(1) *Generator encoding-decoding GAN*

Generator encoding-decoding GAN refers to some models that add encoder and decoder structures to the Generator. The encoder can automatically learn the characteristics of data without labeled. Combining it with Generator can overcome the problem of complex computational quantities required by traditional methods. The model is able to retain more detail information about the real image and can generate an image with better quality. Typical models include Reference 61 [61], Reference 62 [62], FuseGAN [63], GAN-FM [64], MgAN-Fuse [65] and Reference [66].

Reference 61 [61] adds encoder and decoder structures to the Generator for the task of defective image recovery, as shown in Fig. 6 (B)-1. The encoder extracts the feature information of the original defect images and then enters it into the decoder to obtain the recovered image. Compared to the standard GAN model, which directly takes random noise as input, the hidden representation obtained by the encoder can capture more variations and relationships between unknown regions and known regions. The model improves the accuracy of the recovered images. Compared with Reference 61 [61], which directly inputs the feature information extracted by the encoder into the decoder for decoding, Reference 62 [62] adds a L_{max} norm based on the fusion layer between the structure of encoder and decoder to process the extracted feature information, as shown in Fig. 6(B)-2. The processed features are used as the input of the decoder and the encoder uses a two-branch structure to extract the feature information of the two source images separately. The model solves the problems of multiple inputs and single outputs faced in the multimodal medical image fusion task and it also can simultaneously preserve the bone structure information of the original image and the original image information of the soft tissue structure. FuseGAN [63] builds a dual-branch structure encoder based on the Siamese network [67], as shown in Fig. 6(B)-3. The encoders share the weights between the same branches and then the output tensors obtained by dual-branch input into the layer of tensor joins and finally the output result is used as the input to the decoder. This method can alleviate the problem that the focus area cannot be detected perfectly in the multi-focus image fusion task and it also can provide an accurate decision map for the focus area in the field of multi-focus image fusion. Instead of directly using an encoder with a dual-branch structure, GAN-FM [64]concatenates the two source images in the channel dimension and then inputs the image into the encoder and decoder structure, which is built on the basis of a full-scale jump connection structure, as shown in Fig. 6(B)-4. The encoder is used to extract multi-scale depth features of the image. The full-scale jump connection structure processes the obtained feature maps to the same channel and then connect them into the decoder. This method can improve the effect of the fusion result of infrared and visible images. Compared to the above models using an encoder and a decoder in the network architecture, MgAN-Fuse [65] uses two encoders to extract the feature information of the two source images separately, as shown in Fig. 6(B)-5. In this network architecture, the multi-grained attention module is integrated into the multi-scale layer of the encoder, which can distinguish the difference between the infrared and visible images. The obtained multi-grained attention mapping is then concatenated with the corresponding multi-scale feature information of the decoder to obtain the final fused image. This method can preserve both the foreground target information of the infrared image and the background information of the visible image. Reference 66 [66], on the other hand, uses N encoders and M decoders for the field of medical image synthesis, as shown in Fig. 6(B)-6. N encoders with independent parameters encode N source images of different modules respectively, which can effectively use more spatial information while preserving the local detail information of the images. Adding a feature fusion operation F between the structure of the encoder and decoder, which can convert N high-quality feature representations into a unified feature representation C in the space of the common representation, which is decoded by M decoders with independent parameters. This model can synthesize high-quality medical images on both global and detail textures.

(2) *Discriminator encoding-decoding GAN*

Discriminator encoding-decoding GAN refers to the models that add the structure of the encoder and decoder to the Discriminator. Combining the Discriminator with the encoder and decoder allows pre-training operations on the real images without a Generator and can judge the reconstruction capability of the input images. EBGAN [25] is a typical Discriminator encoding-decoding GAN model, as shown in Fig. 6 (B)-7. It treats the Discriminator as a function of energy, providing higher energy for the generated samples and lower energy for the real samples and its output is the Mean Squared Error (MSE) of the encoder and decoder. With this architecture, the Generator can get a relatively large amount of energy at the beginning, so the training process of the network progresses very quickly.

(3) *VAE GAN*

VAE GAN refers to the GAN models based on VAE [2]. Since the images decoded by the Decoder of VAE is easily blurred, in addition, the training process of the model is more difficult by randomly sampling a certain noise and using it as input to the Generator of the GAN model. So VAEGAN [68] combines the Decoder with the Generator for the task of image generation. As shown in Fig. 6(B)-8, sharing parameters between Decoders and Generators and training them together. This method can not only make the image generated by VAE clearer through the discrimination of the Discriminator, but also make the input image of the Generator have a rough spatial distribution through the decoding operation of VAE, so that an image with higher quality can be generated.

(4) *Inverse mapping encoding GAN*

In the standard GAN model, the Generator can map samples from arbitrary latent space to data space, but it cannot map samples from data space back to latent space, that is, it cannot learn the inverse mapping. Therefore, BiGAN [69] adds an encoder to the standard GAN model that can map samples from the data space back to the latent space, as shown in Fig. 6(B)-9. Compared to the standard GAN model, BiGAN takes $(x,E(x))$ and $(z,G(z))$ as the input to the Discriminator. The model successfully maps samples from the data space back to the latent space and is able to extract data features without supervision using an Encoder to generate clear images.

4. Advantages and application

There are common mainly anatomical images and functional images in medical images. The former mainly includes Computed Tomography (CT) images and Magnetic Resonance Imaging (MRI) images; the latter mainly includes Single-Photon Emission Computed Tomography (SPECT) images and Positron Emission Tomography (PET) images.

Among them, CT images can clearly reflect the structural information of human bones and have high spatial resolution, but it cannot show the information of soft tissues well; PET images have high soft tissue resolution, which can clearly show organs, soft tissues and other information, but it has limitations in displaying bone information. It can be seen that a single mode medical image is not enough to provide sufficient auxiliary information for clinicians to refer to. Image fusion technology can realize the organic combination of medical image information, reduce the randomness and redundancy of information, thus it can improve the diagnostic efficiency of some complex diseases. This section will introduce the advantages and application of GAN in medical image fusion field.

4.1. Advantages

The widely used defines of medical image fusion as the process that combines two or more geometrically registered images. The main aim is to generate a fused image with improved quality and salient features. So far, a large number of image fusion methods have been proposed by researchers, including the methods based on spatial domain, the methods based on transform domain and the methods based on deep learning. Among them, the methods based on spatial domain usually takes image pixels, image blocks or image regions as units. Pixels or regions are directly chosen according to some salience measures, then merged with either linear or non-linear operations. The advantages of these methods are: (1) the fusion process is relatively simple; (2) it has high calculation efficiency. The disadvantages are: (1) it is based on a single pixel in the process of fusion, ignoring the similarity of information; (2) it is easy to cause information loss. In the image fusion methods based on transform domain, this paper takes the medical image fusion methods based on multi-scale decomposition as an example to analyze. These methods can be summarized as three steps: source images are first decomposed into multi-scale coefficients using a decomposition transform; then coefficients are merged according to some fusion rules; and finally, the inverse of the transform is applied to reconstruct the fused image. It has the following advantages: (1) it is sensitive to the texture features of the source images; (2) the smooth part and detail information in the source images can be distinguished, and then fused separately to improve the visual effect of the fused image. The disadvantages are: (1) all the effective information contained in different modal medical images may not be obtained by using the same feature extraction algorithm; (2) the end-to-end image fusion task cannot be realized; (3) the quality of the fused image depends on the number of decomposition levels to a large extent.

In recent years, with the rise of deep learning [70], researchers have proposed various models based on deep learning and achieved good results in the corresponding fields [71–74]. Among them, neural networks such as Convolution Neural Network (CNN) and Generative Adversarial Network (GAN) [4] have made remarkable achievements in the field of image fusion through their powerful feature extraction and feature reconstruction capabilities. The image fusion methods based on CNN has the following advantages: (1) the fusion process is end-to-end; (2) strong feature extraction ability. However, these methods have the following disadvantages: (1) strong dependence on data. Because of the particularity of medical images, manual annotation of training data requires professional background knowledge, and the currently published medical image datasets that have been registered are small, so it has certain limitations to use the image fusion methods based on CNN in the field of medical image fusion; (2) Poor model interpretability. On the basis of game theory, GAN has become a research hotspot in the field of medical image fusion through the confrontation game between the generator and the discriminator. This section discusses the main advantages of GAN in the application of medical image fusion filed by comparing the methods based on GAN and the methods based on multi-scale transformation. Fig. 7 shows the fusion process of CT and PET using the above two methods respectively.

Firstly, GAN has a flexible network architecture and can adapt to more complex fusion tasks. For example, (1) multiple Generators are used to extract the deeper level features of the source images, which can not only fully obtain the detail information in the source images of different modes, improve the quality of the fused images, but also can alleviate the problem of mode collapse in the training process of GAN to a certain extent; (2) multiple Discriminators are used to distinguish the generated fusion image from the source images. Fig. 7(b) shows the fusion process of a GAN model with one Generator and two Discriminators. Taking CT source image and fused image, PET source image and fused image as the inputs of CT Discriminator and PET Discriminator respectively, as shown by the blue arrow in the figure. In this way, the fused image can retain more effective information contained in the two source images at the same time. However, the methods based on multi-scale transformation need to strictly follow the steps shown in Fig. 7(a). In addition, GAN can also improve the quality of fused images by designing Generators and Discriminators with different internal network structures.

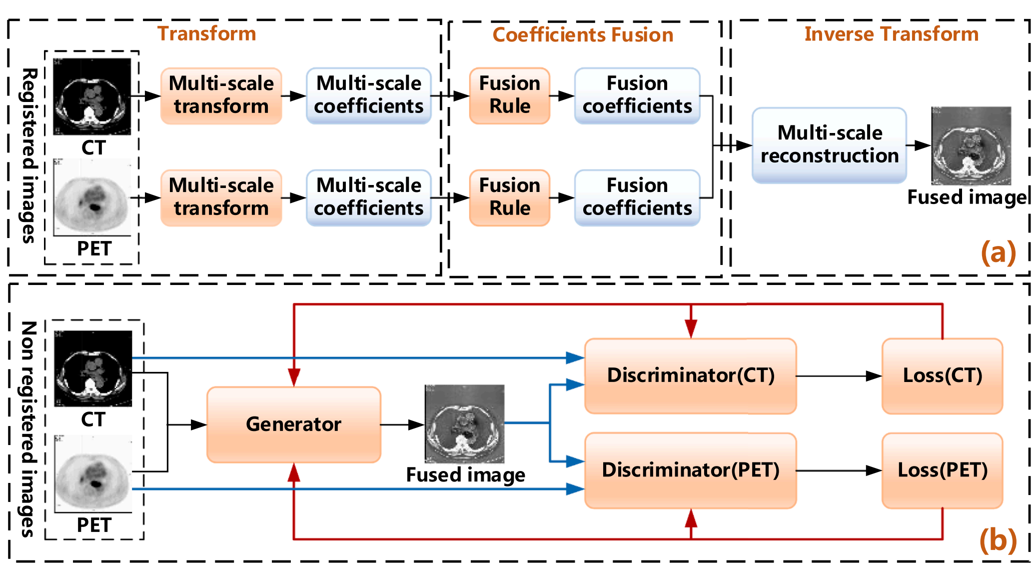

Fig. 7. The process of medical image fusion. (a) The process of medical image fusion based on multi-scale decomposition. (b) The process of medical image fusion based on GAN.

Secondly, the requirements of external conditions are reduced. As an end-to-end model, GAN can adaptively generate fused images under the condition of given label images, reducing the requirement for source image registration because it has strong feature extraction capability. Avoiding the complexity of manually designing fusion rules in Fig. 7(a), and improving the efficiency of the fusion process.

Thirdly, the quality of fused images has been improved. GAN can continuously optimize the Generator and Discriminator through the improved loss function according to the characteristics of different modal medical images, adjust the most appropriate fusion weight according to the experimental results, and constantly improve the quality of the fused images through the Backpropagation algorithm, as shown by the red arrow in Fig. 7(b). However, in the methods based on multi-scale transformation shown in Fig. 7(a), once the fused image is generated, it cannot be modified. This is also one of the advantages of GAN models in medical image fusion field.

4.2. Application

Based on the above advantages, GAN models have been successfully applied in the tasks of medical image fusion, as shown in Fig. 8. This section will discuss the current application status of the GAN models in the field of medical image fusion from three aspects based on the number of Generators and Discriminators: Single-Generator and Single-Discriminator GAN (SGSD GAN), Single-Generator and Dual-Discriminators GAN (SGDD GAN) and Multi-Generators and Multi-Discriminators GAN (MGMD GAN).

4.2.1. SGSD GAN

SGSD GAN models establish an adversarial game between a Generator and a Discriminator. Typical models include DSAGAN [5], MGM-GAN [6], TA-cGAN [12] and Reference 62 [62].

Fu et al [5] proposed the model of DSAGAN, which used a Generator and a Discriminator to achieve the task of fusing anatomical and functional medical images. It extracts the deep features of the image using a dual-stream architecture and multi-scale convolution in the Generator and further enhances the feature information of fused images using Attention Mechanism. The fused images and the multimodal input images are then used as input to the Discriminator. This model can directly output the fused images with sharp edges and sufficient detail textures information without additional manipulation.

MGM-GAN [6] uses a Generator and a Discriminator to achieve the task of fusing multimodal MRI. The Generator of the model treats each unique modality as a separate input channel and uses the corresponding encoder to learn the characteristic distribution of each modality, while introducing GM modules instead of the overlay strategies used in the traditional fusion methods. This model can effectively suppress the irrelevant noise during the fusion processing, but it can only obtain a single output mode at present.

TA-cGAN [12] uses CGAN [8] on the basis of the network architecture of a Generator and a Discriminator to achieve the fusion of brain PET and MRI, in which both the Generator and Discriminator use the tissue label image generated by the MRI as a condition. In this model, the Generator attempts to minimize the objective function by generating a fused image and the Discriminator attempts to maximize the objective function by causing the fused image to contain more information about the MRI structure. This method enhances the anatomical detailed information of the fused image while effectively preserving the color information of the PET image.

Reference 62 [62] implements the task of multimodal medical image fusion through a Generator and a Discriminator, in which the Generator combines the dense blocks with the structure of Encoder and Decoder. Its Generator uses the Encoder to extract the feature information of the images and processes it using the fusion rules based on the L_{max} norm and then inputs it into the Decoder to obtain the final fused image. This method can process the information of the middle layer according to the dense block structure, so that the loss of information during the image fusion processing can be avoided.

4.2.2. SGDD GAN

SGDD GAN models establish an adversarial game between a Generator and two Discriminators. Typical models include DDcGAN [9] and MMFGAN [13].

DDcGAN [9] is used for the task of multimodal medical image fusion of different resolutions. The model enables the Generator to be more fully trained through the adversarial game between one Generator and two Discriminators. In this model, the purpose of the Generator is to generate a real fused image based on a specially designed content loss function and used it to trick the two Discriminators, while the purpose of the two Discriminators is to distinguish the structural differences between the fused image and the two source images, respectively. Experimental results show that the model shows better performance in the task of multi-resolution medical image fusion.

MMFGAN [13] proposed the residual attention mechanism for the first time and applied it to the field of multimodal medical image fusion. The mechanism can use attention-driven and long-term dependency mechanisms to draw the relationship of global dependencies. The addition of dual Discriminators and detail texture blocks to the model effectively alleviates the problem of losing key feature information of the model during the fusion process and it can obtain higher quality fused images.

4.2.3. MGMD GAN

MGMD GAN models establish an adversarial game between multiple Generators and multiple Discriminators. Typical models include Fp^{roi}-GAN [75] and MGMDcGAN [7].

Fp^{roi}-GAN [75] achieves a high-quality image synthesis task of paired medical images and their corresponding regions of interest (ROI) on the basis of CycleGAN through two Generators and two Discriminators. By supplementing the characteristic information of the prior region, it is possible for the Generators to learn the tissue characteristic information of ROI and non-ROI. Experiment shows that this method can effectively improve the structural consistency of the generated images and the real images.

MGMDcGAN [7] implements the medical image fusion tasks of different resolutions on the basis of CGAN with two Generators and two Discriminators. In the first CGAN, the purpose of the Generator is to generate a real fused image based on a specially designed content loss function and trick the two Discriminators with the fused image, which is designed to distinguish the structural differences between the fused images and the source images. On this basis, the model uses a second CGAN with a mask to enhance the dense structural information in the final fused images while preventing the functional information from being weakened. Thus, the final fused images can preserve structural information and functional information

It can be seen that GAN models have been successfully applied in the field of medical image fusion and have achieved good performance in this field. This paper summarizes the above GAN models from five aspects: Model Architecture, Date, Datasets, Evaluation Metrics and Value, as shown in Table 1.

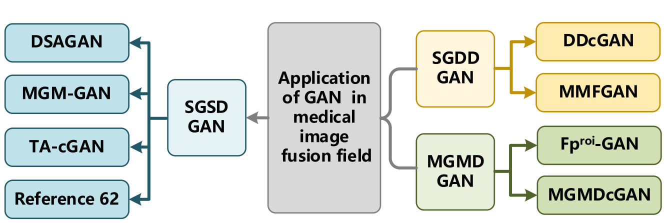

Fig. 8. Application in the field of medical image fusion.

Table 1
GAN models used in the medical image fusion field.

Model	Model Architecture	Date	Datasets	Evaluation Metrics	Value
DSAGAN [5]	SGSD GAN	2021	The Whole Brain Atlas database of Harvard Medical School	AG	7.6204
MGM-GAN [6]	SGSD GAN	2022	BraTS 2015 Dataset	NRMSE	0.232 ±0.112
TA-cGAN [12]	SGSD GAN	2021	AANLIB database	SD	5.3057
Reference 62 [62]	SGSD GAN	2021	The Whole Brain Atlas database of Harvard Medical School	MI	1.0627
DDcGAN [9]	SGDD GAN	2020	AANLIB database	PSNR	65.6339
MMFGAN [13]	SGDD GAN	2022	BrainWeb database, AANLIB database	SSIM	0.8946
Fp^{roi}-GAN [75]	MGMD GAN	2021	INbreast Dataset, BraTS 2017 Dataset	SSIM	0.882 ±0.16
MGMDcGAN [7]	MGMD GAN	2020	AANLIB database	PSNR	67.6702

*Notes: SGSD GAN expresses Single-Generator and Single-Discriminator GAN; SGDD GAN expresses Single-Generator and Dual-Discriminators GAN; MGMD GAN expresses Multi-Generators and Multi-Discriminators GAN; AG expresses Average Gradient; NRMSE expresses Normalized Root Mean Square Error; SD expresses Standard Deviation; MI expresses Mutual Information; PSNR expresses Peak Signal-to-Noise Ratio; SSIM expresses Structure Similarity Index Measure.

(1) Model Architecture: it mainly includes Single-Generator and Single-Discriminator GAN (SGSD GAN), Single-Generator and Dual-Discriminators GAN (SGDD GAN) and Multi-Generators and Multi-Discriminators GAN (MGMD GAN).
(2) Datasets: The Whole Brain Altes is a medical image dataset released by Harvard Medical School in 1999. The dataset is a collection of medical images, UCI data and biomedical literature, containing more than 13,000 pairs of CT-MRI, PET-MRI, and SPECT-MRI images, the dataset is available for download. The BraTS (Brain Tumor Segmentation Challenge) 2015 dataset is a dataset for brain tumor image segmentation. It consists of 220 high grade gliomas (HGG) and 54 low grade gliomas (LGG) MRIs. The four MRI modalities are T1, T1c, T2 and T2FLAIR. AANLIB database is provided by Harvard Medical School. It includes normal MRI along with MRI for diseases cerebrovascular, neoplastic, degenerative and infectious. BrainWeb: Simulated Brain Database (SBD) contains a set of realistic MRI data volumes produced by an MRI simulator. These data can be used by the neuroimaging community to evaluate the performance of various image analysis methods in a setting where the truth is known. Currently, the SBD contains simulated brain MRI data based on two anatomical models: normal and multiple sclerosis (MS). These data are available for viewing in three orthogonal views (transversal, sagittal, and coronal), and for downloading. The INbreast database is a mammographic database, with images acquired at a Breast Centre, located in a University Hospital. The BraTS (Brain Tumor Segmentation Challenge) 2017 dataset contains 285 brain tumor MRI scans, with four MRI modalities as T1, T1ce, T2, and Flair for each scan. The dataset also provides full masks for brain tumors, with labels for ED, ET, NET/NCR. The segmentation evaluation is based on three tasks: WT, TC and ET segmentation. Because the datasets used by various methods are different, the datasets listed in this article only provide a dataset name index.
(3) Table 1 summarizes the convincing evaluation metrics and values used by each network. The evaluation metrics mainly include AG (Average Gradient), SD (Standard Deviation), MI (Mutual Information), PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structure Similarity Index Measure), which can provide reference for researchers in the in-depth study of medical image fusion methods based on GAN.

5. Challenges and prospects

Although researchers have proposed various GAN models and have been widely used in the medical image fusion field, GAN and its application in the field of medical image fusion still face many challenges, as discussed by Arora et al [76], Grnarova et al [77] and Mescheder et al [78]. Here, these challenges are discussed in this paper and the future development directions are pointed out from the following four aspects:

(1) The study of the theoretical problems. The theoretical problem faced by GAN is also the main problem that faced by GAN in the field of medical image fusion. The basic theory of the GAN model is based on a two-person zero-sum game and when the Generator and Discriminator reach the Nash equilibrium, it can be assumed that Generator learns the probability distribution of the real samples. However, the existence of the equilibrium point has not been proved in theory. In addition, there are some problems in Generator and Discriminator are not resolved, such as difficulty converging of the training processing, pattern collapses and unstable training processing. Although WGAN [21] has made positive progresses in the study of gradient disappearance and training process instability in GAN models, these problems will exist for a long time. Therefore, how to make a breakthrough in theory and find the root cause is the urgent need to improve the task of image fusion based on GAN.
(2) The study of the impact of the datasets. The GAN models are characterized by learning the probability distribution of the datasets in real samples. Therefore, a large number of high-quality medical image datasets are needed to improve the training efficiency of GAN models and the quality of fused images. But it is difficult to obtain large-scale and high-quality datasets. At present, the public datasets available for medical image fusion task is mainly the Harvard Datasets, and the methods of transfer learning or the methods of supplementing datasets with natural images may lead to unsatisfactory fusion results. Therefore, how to obtain large-scale and high-quality datasets, improve the training efficiency of GAN models on small-scale datasets to a certain extent, and avoid the negative impact of low-quality samples will be an important research direction of GAN model in the field of medical image fusion.
(3) The study of the interpretability of GAN models. Although the medical image fusion methods based on GAN can effectively extract the deep level features of the source images, the interpretability of GAN is still one of the research focuses at this stage. The standard GAN model is trained by inputting random noise. Its training process is "hidden" and the correspondence relation between the random noise and real samples is unclear. As a result, GAN models are difficult to generate the required samples and the interpretability of models are difficult. InfoGAN [79] has improved the problem in some degree, however, systematic research in this field is not enough, some key questions are not discussed sufficiently. Hence, it is very important to find a way to improve the interpretability of GAN.
(4) The study of the evaluation metrics. At present, there are many objective evaluation metrics that can be used in the field of

medical image fusion, such as Average Gradient (AG), Entropy (EN), Standard Deviation (SD), Spatial Frequency (SF), Structure Similarity Index Measurement (SSIM) and Mutual Information (MI). But at present, there is no unified methods that can be used to evaluate the quality and diversity of various models, researchers can only choose reasonable evaluation metrics according to the actual problems to be solved. Therefore, how to establish a standardized and universal scientific evaluation system to better improve the application of GAN models in the field of medical image fusion and obtain better fused images is also an urgent task in the future.

Therefore, in-depth research from the above four aspects is of great significance for improving the performance of GAN models and the quality of fused images, and assisting in the diagnosis of clinical diseases.

6. Summary

This paper firstly describes the basic principles of GAN in terms of the basic model and the training process. Secondly, the GAN models are summarized from three aspects: Probability Distribution Distance, Overall Network Architecture and Neural Network Structure. Thirdly, the advantages and application of GAN models in the field of medical image fusion are discussed from three aspects. Finally, the challenges faced by GAN and the challenges faced by GAN models in medical image fusion field are discussed, and the future development directions are prospected.

To sum up, it can be predicted from the development speed of deep learning that medical image fusion methods based on GAN will be more widely used in the future, which will greatly promote the progress and development of medical image fusion and make greater contributions to the improvement of medical diagnosis level.

CRediT authorship contribution statement

Tao Zhou: Writing – review & editing, Validation, Supervision. **Qi Li:** Conceptualization, Methodology, Investigation, Writing – review & editing, Visualization. **Huiling Lu:** Data curation, Writing – review & editing, Investigation, Supervision. **Qianru Cheng:** Writing – review & editing, Resources. **Xiangxiang Zhang:** Formal analysis, Writing – review & editing.

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.

Data Availability

No data was used for the research described in the article.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China under Grant No. 62062003, Natural Science Foundation of Ningxia under Grant No. 2022AAC03149 and North Minzu University Research Project of Talent Introduction under Grant No. 2020KYQD08.

References

[1] D. Rumelhart, J. Mcclelland, Information processing in dynamical systems: foundations of harmony theory, in: Proceedings of the 1986 Parallel Distributed Processing: Explorations in the Microstructure of Cognition, 1986, pp. 194–281.
[2] D. Kingma, M. Welling, Auto-encoding variational bayes, 2014, arXiv preprint arXiv: 1312.6114.
[3] H. Larochelle, I. Murray, The neural autoregressive distribution estimator, in: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, 2011, pp. 29–37.
[4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Adv Neural Inf Process Syst 63 (11) (2014) 139–144.
[5] J. Fu, W. Li, J. Du, L Xu, DSAGAN: a generative adversarial network based on dual-stream attention mechanism for anatomical and functional image fusion, Inf. Sci. 576 (9) (2021) 484–506, https://doi.org/10.1016/j.ins.2021.06.083.
[6] B. Zhan, D. Li, X. Wu, J. Zhou, Y. Wang, Multi-modal MRI image synthesis via GAN with multi-scale gate mergence, IEEE J. Biomed. Health Inf. 26 (1) (2022) 17–26, https://doi.org/10.1109/JBHI.2021.3088866.
[7] J. Huang, Z. Le, Y. Ma, F. Fan, L. Yang, MGMDcGAN: medical image fusion using multi-generator multi-discriminator conditional generative adversarial network, IEEE Access 8 (2020) 55145–55157, https://doi.org/10.1109/ACCESS.2020.2982016.
[8] M. Mirza, S. Osindero, Conditional generative adversarial nets, PeerJ Comput Sci (2014) 2672–2680.
[9] J. Ma, H. Xu, J. Jiang, X. Mei, XP. Zhang, DDcGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion, IEEE Trans. Image Process. 29 (2020) 4980–4995, https://doi.org/10.1109/TIP.2020.2977573.
[10] M. Jiang, M. Zhi, L. Wei, X. Yang, J. Zhang, Y. Li, P. Wang, J. Huang, G. Yang, FA-GAN: fused attentive generative adversarial networks for MRI image super-resolution, Comput. Med. Imaging Graph. 92 (2021), 101969, https://doi.org/10.1016/j.compmedimag.2021.101969.
[11] L. Wang, C. Chang, B. Hao, C. Liu, Multi-modal medical image fusion based on GAN and the shift-invariant shearlet transform, in: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020, pp. 2538–2543, https://doi.org/10.1109/BIBM49941.2020.9313288.
[12] J. Kang, W. Lu, W. Zhang, Fusion of brain PET and MRI images using tissue-aware conditional generative adversarial network with joint loss, IEEE Access 8 (2020) 6368–6378, https://doi.org/10.1109/ACCESS.2019.2963741.
[13] K. Guo, X. Hu, X. Li, MMFGAN: a novel multimodal brain medical image fusion based on the improvement of generative adversarial network, Multimedia Tools and Applications. 81 (4) (2022) 5889-5927, https://doi.org/10.1007/s11042-021-11822-y.
[14] Z. Wang, Q. She, T.E. Ward, Generative adversarial networks in computer vision: a survey and taxonomy, ACM Comput. Surv. 54 (2) (2022) 1–38. https://dl.acm.org/doi/10.1145/3439723.
[15] J. Jeong, A. Tariq, T. Adejumo, H. Trivedi, J. Gichoya, I. Banerjee, Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation, J. Digit. Imaging 35 (2022) 137–152, https://doi.org/10.1007/s10278-021-00556-w.
[16] A. Wali, Z. Alamgir, S. Karim, A. Fawaz, M. Barkat Ali, M. Adan, M. Mujtaba, Generative adversarial networks for speech processing: a review, Comput Speech Lang 72 (2022), https://doi.org/10.1016/j.csl.2021.101308.
[17] S. Nowozin, B. Cseke, R. Tomioka, f-GAN: training generative neural samplers using variational divergence minimization, in: Proceedings of the Advances in Neural Information Processing Systems 29: NIPS'16, 2016, pp. 271–279. https://di.acm.org/doi/10.5555/3157096.3157127.
[18] A. Müller, Integral probability metrics and their generating classes of functions, Adv. Appl. Probab. 29 (2) (1997) 429–443, https://doi.org/10.2307/1428011.
[19] E. David, G. Hinton, R. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536. https://dl.acm.org/doi/10.5555/65669.104451.
[20] L. Ratliff, S. Burden, S. Sastry, Characterization and computation of local Nash equilibria in continuous games, in: 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2013, pp. 917–924, https://doi.org/10.1109/Allerton.2013.6736623.
[21] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein GAN, 2017, arXiv preprint arXiv: 1701.07875.
[22] Q. Xu, G. Huang, Y. Yuan, C. Guo, Y. Sun, F. Wu, K. Weinberger, An empirical study on evaluation metrics of generative adversarial networks, 2018.
[23] Y. Mroueh, T. Sercu, Fisher GAN, in: Proceedings of the Advances in Neural Information Processing Systems 30: NIPS'17, 2017, pp. 2510-2520, https://dl.acm.org/doi/10.5555/3294996.3295012.
[24] X. Mao, Q. Li, H. Xie, R. Lau, Z. Wang, SP Smolley, On the effectiveness of least squares generative adversarial networks, IEEE Trans. Pattern Anal. Mach. Intell. 41 (12) (2019) 2947–2960, https://doi.org/10.1109/TPAMI.2018.2872043.
[25] J. Zhao, M. Mathieu, Y. Lecun, Energy-based generative adversarial network, in: The 5th International Conference on Learning Representations, ICLR, 2017.
[26] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. Courville, Improved training of Wasserstein GANs, NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 5769-5779, https://dl.acm.org/doi/10.5555/3295222/3295327.
[27] H. Petzka, A. Fischer, D. Lukovnicov, On the regularization of Wasserstein GANs, in: International Conference on Learning Representations (ICLR), 2017.
[28] J. Wu, Z. Huang, J. Thoma, Wasserstein divergence for GANs, in: European Conference on Computer Vision (ECCV) 11209, 2018, pp. 673–688, https://doi.org/10.1007/978-3-030-01228-1_40.
[29] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, Spectral normalization for generative adversarial networks, 2018.

[30] Y. Mroueh, T. Sercu, V. Goel, McGan: mean and covariance feature matching GAN, in: The 34th International Conference on Machine Learning 70, 2017, pp. 2527–2535, https://doi.org/10.48550/arXiv.1702.08398.
[31] Y. Li, K. Swersky, R. Zemel, Generative moment matching networks, in: International Conference on Machine Learning, 2015, pp. 1718–1727.
[32] C. Li, W. Chang, Y. Cheng, Y. Yang, B. Póczos, GAN MMD, Towards deeper understanding of moment matching network, in: The 31st International Conference on Neural Information Processing Systems, 2017, pp. 2200–2210. https://dl.acm.org/doi/10.5555/3294771.3294981.
[33] Z. Yang, Y. Chen, Z. Le, Y. Ma, GANFuse: a novel multi-exposure image fusion method based on generative adversarial networks, Neural Comput Appl 33 (2021) 6133–6145, https://doi.org/10.1007/s00521-020-05387-4.
[34] A. Ghosh, V. Kulharia, V. Namboodiri, P. Torr, P. Dokania, Multi-agent diverse generative adversarial networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8513–8521, https://doi.org/10.1109/CVPR.2018.00888.
[35] H. Quan, D. Tu, T. Le, D. Phung, MGAN: training generative adversarial nets with multiple generators, 6th Int. Conf. Learn. Represent. (2018).
[36] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, D. Metaxas, StackGAN: text to photo-realistic image synthesis with stacked generative adversarial networks, in: 2017 IEEE International Conference on Computer Vision (ICCV, 2017, pp. 5908–5916.
[37] Z. Han, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, D. Metaxas, StackGAN++: realistic image synthesis with stacked generative adversarial networks, IEEE Trans. Pattern Anal. Mach. Intell. 41 (8) (2019) 1947–1962, https://doi.org/10.1109/TPAMI.2018.2856256.
[38] E. Denton, S. Chintala, A. Szlam, R. Fergus, Deep generative image models using a Laplacian pyramid of adversarial networks, in: Proceedings of the 28th International Conference on Neural Information Processing Systems, 2015, pp. 1486–1494.
[39] M. Liu, O. Tuzel, Coupled generative adversarial networks, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016, pp. 469–477. https://dl.acm.org/doi/abs/10.5555/3157096.3157149.
[40] Q. Li, L. Lu, Z. Li, W. Wu, X. Yang, Coupled GAN with relativistic discriminators for infrared and visible images fusion, IEEE Sensors J. 21 (6) (2021) 7458–7467, https://doi.org/10.1109/JSEN.2019.2921803.
[41] J. Zhu, T. Park, P. Isola, A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: 2017 IEEE International Conference on Computer Vision (ICCV, 2017, pp. 2242–2251, https://doi.org/10.1109/ICCV.2017.244.
[42] Z. Yi, H. Zhang, P. Tan, M. Gong, DualGAN: unsupervised dual learning for image-to-image translation, in: 2017 IEEE International Conference on Computer Vision (ICCV, 2017, pp. 2868–2876, https://doi.org/10.1109/ICCV.2017.310.
[43] T. Kim, M. Cha, H. Kim, Learning to discover cross-domain relations with generative adversarial networks, in: Proceedings of the 34th International Conference on Machine Learning, 2017, pp. 1857-1865.
[44] L. Karacan, Z. Akata, A. Erdem, E. Erdem, Learning to generate images of outdoor scenes from attributes and semantic layouts, 2016.
[45] A. Odena, Semi-supervised learning with Generative Adversarial Networks, 2016.
[46] A. Odena, C. Olah, J. Shlens, Conditional image synthesis with auxiliary classifier GANs, in: Proceedings of the 34th International Conference on Machine Learning, 2016, pp. 2642–2651. https://dl.acm.org/doi/10.5555/3305890.3305954.
[47] P. Burt, E. Adelson, The Laplacian pyramid as a compact image code, IEEE Trans. Commun. 31 (4) (1983) 532–540, https://doi.org/10.1016/B978-0-08-051581-6.50065-9.
[48] Y. Xia, D. He, T. Qin, L. Wang, N. Yu, T. Liu, W. Ma, Dual learning for machine translation, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016, pp. 820–828.
[49] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks,2015.
[50] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C. Loy, ESRGAN: enhanced super-resolution generative adversarial networks, Eur. Conf. Comput. Vision (ECCV), 11133 (2018) 63–79, https://doi.org/10.1007/978-3-030-11021-5_5.
[51] J. Wang, X. LI, H. Liu, Exposure fusion using a relative generative adversarial network, IEICE Trans. Inf. Syst. (2021) 1017–1027, https://doi.org/10.1587/transinf.2021EDP7028.
[52] J. Ma, Y. Wei, P. Liang, L. Chang, J. Jiang, FusionGAN: A generative adversarial network for infrared and visible image fusion, Inf Fusion 48 (2019) 11–26, https://doi.org/10.1016/j.inffus.2018.09.004.
[53] C. Yuan, C. Sun, X. Tang, R. Liu, GAN FLGC-Fusion, An enhanced fusion GAN model by importing fully learnable group convolution, Math. Probl. Eng. (2020) 1–13, https://doi.org/10.1155/2020/6384831.
[54] H. Zhang, I. Goodfellow, D. Metaxas, A. Odena, Self-attention generative adversarial networks, 2018.
[55] H. Xu, J. Ma, X. Zhang, MEF-GAN: multi-exposure image fusion via generative adversarial networks, IEEE Trans. Image Process. 29 (2020) 7203–7216, https://doi.org/10.1109/TIP.2020.2999855.
[56] A. Jolicoeur-Martineau, The relativistic discriminator: a key element missing from standard GAN, 2018.
[57] J. Kim, J. Lee, K. Lee, Accurate image super-resolution using very deep convolutional networks, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654, https://doi.org/10.1109/CVPR.2016.182.
[58] K. SIMONYAN, A. ZISSERMAN, Very deep convolutional networks for large-scale image recognition, in: The 3rd International Conference on Learning Representations, 2015, pp. 1–14.
[59] A. Vaswani, N. Shazeer, N. Parmar, Attention Is All You Need, arXiv preprint arXiv: 1706.03762. 2021, https://dl.acm.org/doi/10.5555/3295222.3295349.
[60] E. David, L. James, Learning internal representations by error propagation, Parallel Distrib. Process. (1987) 318–362. URL, http://ieeexplore.ieee.org/document/6302929.
[61] X. Jin, Y. Hu, C. Zhang, Image restoration method based on GAN and multi-scale feature fusion, in: 2020 Chinese Control and Decision Conference (CCDC), 2020, pp. 2305–2310, https://doi.org/10.1109/CCDC49329.2020.9164498.
[62] C. Zhao, T. Wang, B. Lei, Medical image fusion method based on dense block and deep convolutional generative adversarial network, Neural Comput Appl 33 (2021) 6595–6610, https://doi.org/10.1007/s00521-020-05421-5.
[63] X. Guo, R. Nie, J. Cao, D. Zhou, L. Mei, K. He, FuseGAN: learning to fuse multi-focus image via conditional generative adversarial network, IEEE Trans. Multimedia 21 (8) (2019) 1982–1996, https://doi.org/10.1109/TMM.2019.2895292.
[64] H. Zhang, J. Yuan, T. Xin, J. Ma, GAN-FM: infrared and visible image fusion using GAN with full-scale skip connection and dual Markovian discriminators, IEEE Trans Comput Imaging 7 (2021) 1134–1147, https://doi.org/10.1109/TCI.2021.3119954.
[65] J. Li, H. Huo, C. Li, R. Wang, C. Sui, Z. Liu, Multigrained attention network for infrared and visible image fusion, IEEE Trans. Instrum. Meas. 70 (2021) 1–12, https://doi.org/10.1109/TIM.2020.3029360.
[66] X. Liu, A. Yu, X. Wei, Z. Pan, J. Tang, Multimodal MR image synthesis using gradient prior and adversarial learning, IEEE J Sel Top Signal Process 14 (6) (2020) 1176–1188, https://doi.org/10.1109/JSTSP.2020.3013418.
[67] S. Chopra, R. Hadsell, Y. LeCun, Learning a similarity metric discriminatively, with application to face verification, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 539–546, https://doi.org/10.1109/CVPR.2005.202.
[68] A. Larsen, S.Kaae Snderby, H. Larochelle, O. Winther, Autoencoding beyond pixels using a learned similarity metric, in: The 33rd International Conference on International Conference on Machine Learning, 2016, pp. 1558–1566.
[69] J. Donahue, P. Krähenbühl, T. Darrell, Adversarial Feature Learning, 2016.
[70] T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu, Y. Liu, Dense convolutional network and its application in medical image analysis, Biomed. Res. Int. 2022 (2022) 1–22, https://doi.org/10.1155/2022/2384830.
[71] N. Zeng, H. Li, Y. Peng, A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease, Neural Comput Appl. (2021) 1–12, https://doi.org/10.1007/s00521-021-06149-6.
[72] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, X. Liu, A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection, IEEE Trans. Instrum. Meas. 71 (2022) 1–14, https://doi.org/10.1109/TIM.2022.3153997.
[73] H. Li, N. Zeng, P. Wu, K. Clawson, Cov-Net: a computer-aided diagnosis method for recognizing COVID-19 from chest X-ray images via machine vision, Expert Syst. Appl. 207 (2022) 1–12, https://doi.org/10.1016/j.eswa.2022.118029.
[74] T. Zhou, H. Lu, Z. Yang, S. Qiu, B. Huo, Y. Dong, The ensemble deep learning model for novel COVID-19 on CT images, Appl. Soft Comput. 98 (2021), https://doi.org/10.1016/j.asoc.2020.106885.
[75] J. Dong, C. Liu, P. Man, G. Zhao, Y. Wu, Y. Lin, Fproi-GAN with fused regional features for the synthesis of high-quality paired medical images, J. Healthc. Eng. 2021 (2021), https://doi.org/10.1155/2021/6678031.
[76] S. Arora, G. Rong, Y. Liang, T. Ma, Z. Yi, Generalization and equilibrium in generative adversarial nets (GANs), Proceedings of the 34th International Conference on Machine Learning, 70 (2017) pp. 224-232.
[77] P. Grnarova, K. Levy, A. Lucchi, T. Hofmann, A. Krause, An online learning approach to generative adversarial networks, 2017.
[78] L. Mescheder, S. Nowozin, A. Geiger, The Numerics of GANs, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 1823–1833. https://dl.acm.org/doi/10.5555/3294771.3294945.
[79] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, InfoGAN: interpretable representation learning by information maximizing generative adversarial Nets, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016, pp. 2172–2180. https://dl.acm.org/doi/10.5555/3157096.3157340.