Generative Adversarial Networks: A Literature Review

Jieren Cheng[1,2], Yue Yang[1,2*], Xiangyan Tang[1,2,3], Naixue Xiong[3,4] Yuan Zhang[1,2], and Feifei Lei[1]
[1] School of Compute Science and Cyberspace Security
Hainan University, Haikou, 570228, China
[e-mail: yangyuehnu@163.com , 357734432@qq.com, 2467967614@qq.com]
[2] Hainan Blockchain Technology Engineering Research Center
Haikou, 570228, China
[e-mail: cjr22@163.com]
[3] College of Intelligence and Computing
Tianjin University, Tianjin, 300350, China
[e-mail: tangxy36@163.com]
[4] Department of Mathematics and Computer Science
Northeastern State University, Tahlequah, OK, USA
[e-mail: xiongnaixue@gmail.com]
*Corresponding author: Yue Yang

Received August 21, 2020; revised September 14, 2020; accepted November 23, 2020; published December 31, 2020

Abstract

The Generative Adversarial Networks, as one of the most creative deep learning models in recent years, has achieved great success in computer vision and natural language processing. It uses the game theory to generate the best sample in generator and discriminator. Recently, many deep learning models have been applied to the security field. Along with the idea of "generative" and "adversarial", researchers are trying to apply Generative Adversarial Networks to the security field. This paper presents the development of Generative Adversarial Networks. We review traditional generation models and typical Generative Adversarial Networks models, analyze the application of their models in natural language processing and computer vision. To emphasize that Generative Adversarial Networks models are feasible to be used in security, we separately review the contributions that their defenses in information security, cyber security and artificial intelligence security. Finally, drawing on the reviewed literature, we provide a broader outlook of this research direction.

Keywords: Generative Adversarial Networks, Artificial Intelligence, Generative Model, Deep Learning, Security

1. Introduction

With the emergence of autonomous driving, machine translation and other technologies [1-3], Artificial intelligence (AI) gradually entered public life [4-6]. All the time researchers were trying to improve the learning and optimization ability of computers, machine learning (ML) has officially entered the stage of AI. Machine learning is a method of learning with data or previous experience to optimize the objective model's performance [7]. According to the way of learning, it can be divided into supervised learning, semi-supervised learning, unsupervised learning, transfer learning and reinforcement learning (RL) [8]. Transfer learning is a learning method that transfers the trained model parameters to a new model for training [9-11]. Reinforcement learning is used to describe and solve the problem that agents use learning strategies to maximize returns or achieve specific goals in their interaction with the environment [12, 13]. It contains the value-based algorithm and the policy-based algorithm. Supervised learning is a method to use the existing data and their relationship to train the optimal model [14]. The classical algorithms are Naive Bayesian Classifier [15], Decision Tree [16], k-Nearest Neighbor (KNN) [17], Support Vector Machine (SVM) [18] and so on. However, once there is a lack of prior knowledge in real life, it is difficult for machines to serve humans due to the increased difficulty in manual labeling of categories and the high cost of categories.

Fortunately, unsupervised learning represented by clustering algorithm provides ideas for solving such problems [19, 20]. In 2014, Ian Goodfellow et al. [21] proposed a generative model named Generative Adversarial Networks (GAN). It is composed of a generator and a discriminator. The generator is responsible for producing samples, and the discriminator is responsible for distinguishing the authenticity of samples. Since the goal of each side is to defeat the other, the model that optimizes itself is continuously modified, and the generator after the final training can generate a nearly real sample through any input.

In the past few years, GAN has superior output samples compared with other generative models, and it has been widely applied in the fields of image generation and natural language processing [22]. In this paper, we review the progress of GAN in recent years. We firstly introduce the existing generative models in section 2, the latest derived models are listed and their pros and cons with GAN are contrasted. Afterwards, we describe two factors (generator, discriminator) in GAN including the training process and the evaluation index in section 3. In part 4, we introduce three typical models of GAN. By modifying the network structure or improving the model algorithm, these models solve some GAN training problems. After that, there is an introduction to the GAN`s classical applications in section 5, some recent applications of GAN in computer vision and natural language processing are shown in this part. In chapter 6, we discuss the applications of security on GAN. It proves that GAN has an excellent application prospect in the field of security. Next, we summarize the current development of GAN and several questions need to be solved in section 7. The last part is the conclusion.

2. Previous Work

2.1 Game Theory

The inspiration of GAN comes from the zero-sum game in game theory [23]. The zero-sum game, a non-cooperative game, is defined as a game in which two parties are strictly opposed to each other, where the gains of one party are bound to bring losses of the other party, and the

gains and losses of both parties add up to zero [24]. In GAN, the discriminator judges the samples generated by the generator. The more real the images generated by the generator, the more difficult it is for the discriminator to judge the authenticity and falsity. Similarly, in the early training period, the samples with poor quality generated can be easily identified as false samples. Eventually they will reach an equilibrium point in the game named the Nash equilibrium. Nash equilibrium refers to the strategy adopted by both sides of the game to maximize their own interests [25]. In continuous training, the generator seeks to generate enough samples to fool the discriminator, which tries to recognize the authenticity of all the samples. The process of mutual game is also a characteristic and highlight of generative adversarial networks.

2.2 The Classic Generation Models

2.2.1 Auto-regressive Network

Auto-regressive models are often applied to the prediction of economics, informatics and natural phenomena [26]. The auto-regressive model is a directed probability model without potential random variables and belongs to the category of supervised learning. Auto-regressive models are often used to deal with problems of time series [27].

Pixel recursive neural network (Pixel RNN) [28] is a relatively new generation method in recent years, the basic idea of this model is to generate images from one pixel to another, and the former also can be a reference for the latter. Finally, the prediction of the joint distribution of each pixel on the image is converted into the conditional distribution. The specific prediction is as follows.

$$p(x)=\prod_{i=1}^{n^2} p\left(x_i \middle| x_1,\cdots,x_{i-1}\right) \tag{1}$$

In addition to RNN [29, 30], the author also adopted the method of convolutional neural network (CNN) to carry out convolution around the generated pixel points and later researchers also improved and optimized this kind of approach [31, 32]. For DeepMind's latest research, an Autoregressive Model of 3D Meshes was proposed [33]. This method improves the quality of grid vertex prediction from 2D to 3D. At the same time, the development of autoregressive networks proves that this method is still applicable to the latest generation requirements.

Compared with GAN, the advantage of the Auto-regressive network is to explicitly calculate likelihood $P(x)$ and put forward a good evaluation measure explicitly. The disadvantage is that the generation speed is slow, and the resolution of the picture is not high.

2.2.2 Variational Autoencoders (VAE)

Autoencoder is a data compression algorithm, in which the data compression and decompression are realized by neural network self-learning [34]. The encoder maps the input data to the low-latitude features we need, and then reconstructs the original input data through the decoder. Variational Autoencoder [35] is a method that adds "Gaussian noise" to the result of the encoder in Autoencoder to make the result of decoder robust to noise.

$$p(X)=\sum_{Z} p(X|Z)\,p(Z) \tag{2}$$

In the formula above, X is training data, and Z is the hidden feature that cannot be observed in X data. The characteristic of VAE is that every one-dimensional distribution of Z conforms to a normal distribution, and the learning of characteristics is introduced to make the decoding effect better [36, 37]. However, VAE adopted the Variational Inference [38] for approximation. Compared with GAN, the fitting of the real data is not as good as GAN. From the result of the generated picture, the picture clarity of GAN is also better than that of VAE.

3. Principles and criteria of GAN

3.1 Basic Concept

It is mentioned in 2.1 that GAN originates from a zero-sum game in game theory, in which the two players are generator and discriminator. The random noise z is input into the generator and generates a sample $G(z)$, and then the test sample is put into the discriminator. The more realistic the generated sample is, the closer the result is to 1; otherwise, the result is close to 0 [21].

The loss of generator lies in the discriminator's misclassification of the generated non-real samples, generator should continuously train to produce new generated samples to cheat the discriminator. The loss of discriminator lies in the wrong judgment of real samples and generated samples. The optimization of the model is described in 3.2. The general process is shown in **Fig. 1**.

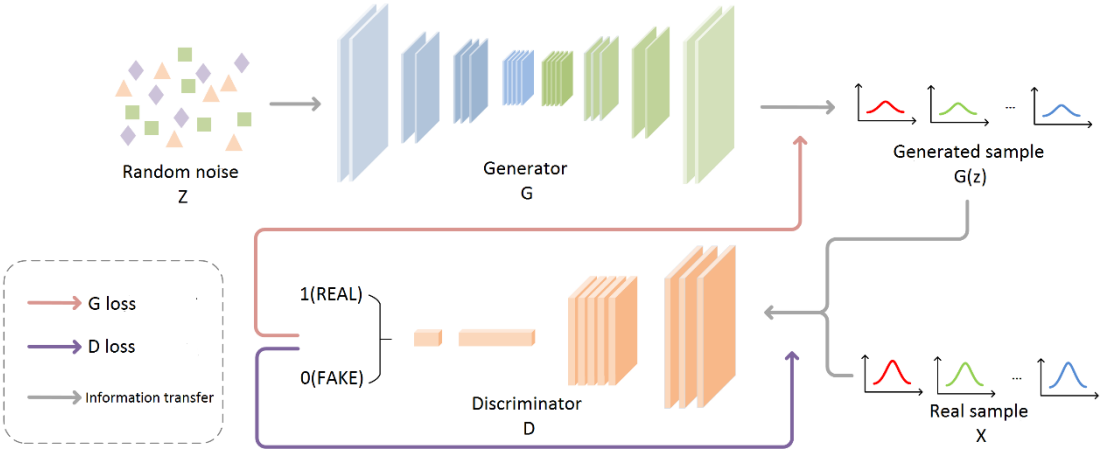


Fig. 1. GAN model frame diagram

3.2 Training Process

Since it is difficult for the first generated sample to reach the level of real data, we need to continuously train and optimize the GAN. However, the training of GAN is different from the previous single neural network training, and we adopt separate alternating iterative training [39-41]. In GAN, we use fixed generator to optimize the discriminator, or fixed discriminator to optimize the generator [21]. The formula for the entire GAN is as follows.

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}\left[\log(D(x))\right] + E_{z\sim P_z(z)}\left[\log(1-D(G(z)))\right] \tag{3}$$

E is the expected value of the distribution function, $P_{data}(x)$ is the distribution of the real sample, $P(z)$ is the distribution that generates the sample.

3.2.1 Training Generator

After the discriminator is fixed, the generator generates samples for the discriminator to judge, and the final result is fed back to the generator. Since there is a certain gap between the generated samples and the real samples at the beginning of training, the discriminator will feed back loss to the generator. The final result should be that the generator can generate the generated samples that cheat the discriminator, and the discriminator's judgment on the sample is near to 1.

$$\min_{G} V(D,G) = E_{z\sim p_z(z)}\left[\log(1-D(G(z)))\right] \tag{4}$$

Since the generator is optimized to make the generation of samples as real as possible and the discriminator identifies the result as close to 1, $G(z)$ is the sample generation, the value of $D(G(z))$ is as large as possible.

3.2.2 Training Discriminator

In the training process of the frequency discriminator, the generator is fixed. The frequency discriminator improves the frequency discrimination ability by continuously judging the real samples and the generated samples, and finally achieves higher frequency discrimination abil ity. Therefore, the frequency discrimination of the non-real samples is close to 0.

$$\max_{D} V(D,G) = E_{x\sim P_{data}(x)}\left[\log(D(x))\right] + E_{z\sim P_z(z)}\left[\log(1-D(G(z)))\right] \tag{5}$$

The x of the first half of the formula is the real sample of the input, and the discriminator should make the result of the real sample as big as possible. The latter part is the sample generated by the generator, and the discriminator should make the result of sample generation as small as possible, while on the contrary, the result of $1\text{-}D(G(z))$ should be as large as possible, and the maximum value of the sum of the two should be calculated at last.

3.3 The Evaluation Index

After repeating iterative training, the generator and discriminator should cheat and identify each other, but they cannot measure the quality and diversity of sample generation. In chapter 4, we will enumerate some typical GAN models, and the merits of these models also need some general indicators to measure. Next, we will introduce two general evaluation indicators.

3.3.1 Inception Score (IS)

During the ILSVRC competition in 2014, Google proposed a network called Inception Net [42]. It not only controlled the number of parameters, but also achieved good classification performance. In 2015, Inception net-v3 was proposed and it was exactly the model required for IS calculation in this part [43, 44].

As analyzed above, IS focuses on the sharpness and diversity of images, which IS is not high resolution but clear classification [45]. The formula for IS is as follows.

$$IS\ （(G) = \exp\left(E_{x\sim p_g} D_{KL}\left(p(y|x)\|p(y)\right)\right) \tag{6}$$

Generate the 1000-dimensional vector y obtained from the input model of picture x, and the probability distribution of x belonging to each category is $P(y|x)$. $P(y)$ is the marginal distribution of all the images obtained after the input model of a large number of images is generated. $P(y|x)$ represents sharpness, while $P(y)$ is related to variety.

However, the evaluation index of IS was limited to the use of data set, because the Inception V3 model adopted by IS was trained on ImageNet [46], so the image generation of GAN also needs to rely on ImageNet to serve as a real sample. It no longer makes sense to use the classification model and the generation model in different data sets. At the same time, in the case of insufficient samples, the estimation of sample distribution will become very difficult, so we also need other indicators to evaluate the capability of GAN.

3.3.2 Fréchet Inception Distance (FID)

FID [47] also uses the Inception V3 model. Due to the limitations of ImageNet dataset, images that do not exist in the data set are uniformly judged to be not real images. FID chooses the method of extracting image features to calculate the distance between the generated samples and the real samples in the feature space. The closer the distance is, the better the effect and diversity of the naturally generated images will be.

$$FID(x,g) = \left\|\mu_x - \mu_g\right\|^2 + Tr\left(\sum{}_x + \sum{}_g - 2\left(\sum{}_x \sum{}_g\right)^{\frac{1}{2}}\right) \tag{7}$$

According to the formula, FID is negatively correlated with picture quality. Compared with IS, it is more robust to noise. Even in the case of insufficient samples (only one type of picture is produced), it will have a high FID value [48].

The following table lists FID values of the more classical generation models in recent years under different datasets [49]. MM GAN and NS GAN are expressed as mini-max [50] loss function and non-saturating [51] loss function respectively.

Table 1. FID values of different GAN models on corresponding datasets [48]

Datasets / Models	**MNIST** [52]	**FASHION** [53]	**CIFAR** [54]	**CELEBA** [55]
MM GAN	9.8±0.9	29.6±1.6	72.7±3.6	65.6±4.2
NS GAN	6.8±0.5	26.5±0.6	58.5±1.9	55.0±3.3
LSGAN [56]	7.8±0.6	30.7±2.2	87.1±47.5	53.9±2.8
WGAN [57]	6.7±0.4	21.5±1.6	55.2±2.3	41.3±2.0
WGAN-GP [58]	20.3±5.0	24.5±2.1	55.8±0.9	30.0±1.0
DRAGAN [59]	7.6±0.4	27.7±1.2	69.8±2.0	42.3±3.0
BEGAN [60]	13.1±1.0	22.9±0.9	71.4±1.6	38.9±0.9
VAE [35]	23.8±0.6	58.7±1.2	155.7±11.6	85.7±3.8

It can be seen that FID values of different GAN models are quite different under different data sets. In recent years, more and more GAN models take FID value as the index of GAN

model optimization. In addition, Mode Score (MS) [61], Kernel Maximum Mean Discrepancy (Kernel MMD) [62] and 1-nearest Neighbor (1-NN) [63] classifier are also the evaluation indicators of GAN model. With the unceasing development of GAN, more accurate evaluation criteria will be continuously proposed to evaluate the model.

4. Typical GAN models

Since GAN was proposed by Goodfellow in 2014, it has become one of the hottest research fields in artificial intelligence and machine learning [64]. Different kinds of GAN models have been blossoming in recent years, the latest International Conference on Learning Representations (ICLR) on GAN publications still contribute in the front row. In this section, we will list some typical models.

4.1 Conditional Generative Adversarial Nets (CGAN)

In the second part, we introduced two typical generation models PixelRNN and VAE. The advantage of GAN is that the input does not rely on an expected data distribution; the generated samples can be more real by using direct sampling to input random noise Z. At the same time, such a disadvantage is that the generated samples are too free to control and cannot focus on the specified samples, and we need to add some constraints to GAN. It is from this idea that CGAN came onto the stage of GAN.

Due to GAN's double randomness (random noise and random samples), Mehdi Mirza et al proposed CGAN [65], and *y* tag was added as input in GAN generator and discriminator. With the addition of *y*, the input of generator becomes noise and label, while the input of discriminator becomes real sample and generates sample and label.

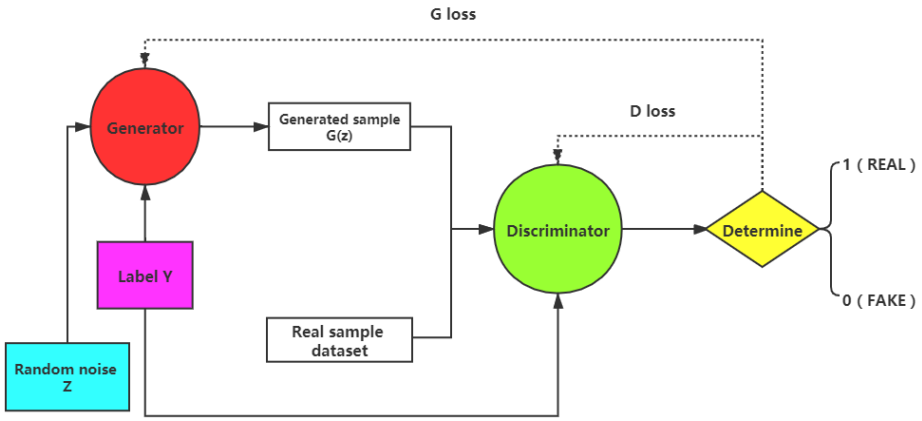
As for the training process of CGAN, *y* is increased compared with GAN's formula, the details are as follows.

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}\left[\log(D(x|y))\right] + E_{z\sim P_z(z)}\left[\log(1-D(G(z|y)))\right] \tag{8}$$

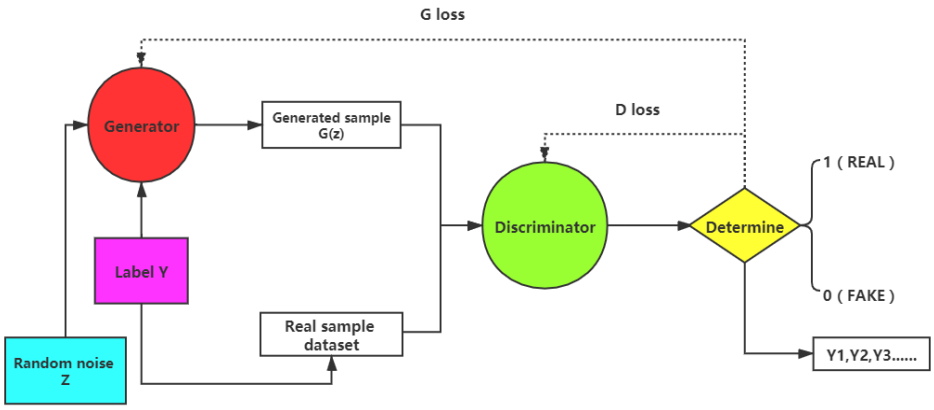
With the addition of y, the formula becomes a binary minimization maximum problem with conditional probability. The generation of CGAN is controlled by y condition, which realizes the supervision of generator and can generate different samples according to different y values. Since the discriminator also has y input, we can also select the type we need when comparing the sample generated by the generator with the real sample.

CGAN can not only produce samples of specific labels, but also improve the quality of generated samples. Then, semi-supervised GAN (SGAN) [66] proposed by Augustus Odena turned GAN into a semi-supervised model and made it possible to judge its output category labels. It was the thought of the two models that was received, Auxiliary Classifier GAN (ACGAN) [67], which further improved the quality of CGAN's generated samples.

ACGAN's improvement on CGAN is that it not only uses tags in the input of generator, but also adds tags into the real sample, so that the generator can better understand the required sample structure. Two full connection layers are used in the output part of discriminator. Borrowing the idea of SGAN, in addition to generating discriminant results (Real or Fake), the classification results of samples are added.



(a)



(b)

Fig. 2. CGAN (b), an improvement on CGAN (a)

4.2 Deep Convolution Generative Adversarial Networks (DCGAN)

The traditional neural network is simply composed of three sections: input layer, hidden layer and output layer [68]. Within each layer, neurons with weights are mapped to the next layer by activation function. This way of receiving input from the previous layer's output to the next layer is also called full connection. The disadvantage of this way is that it is affected by a large number of parameters such as weight, the convergence of training is too slow, and the generalization effect is not good.

Convolutional neural network (CNN) is a feed forward neural network with convolution calculation [69]. It trains the weights of the CNN through the back-propagation algorithm, and finally obtains the classification results. CNN is similar to the traditional neural network structure, which is composed of input layer, hidden layer and output layer. The hidden layer in the traditional neural network structure includes convolutional layer, pooling layer and full connection layer, among which weight sharing of convolution layer solves the problem of low training efficiency caused by excessive parameters of traditional neural network. The birth of CNN provides a new idea for the development of deep learning.

CNN has a good effect in supervised learning [70], but it is seldom used in unsupervised direction. At the beginning, some researchers tried to combine CNN and GAN, but no good

effect was achieved [71]. Until 2015, Alec Radford et al. proposed DCGAN [72], which improved the network structure of GAN and greatly improved the quality of GAN generated images and the stability of training.

DCGAN chose to remove the hidden layer of full connection because the full connection mode of GAN made the training inefficient. Batch Normalization [73] is used in the convergence of the model and avoids the collapse of the generator, allowing for deeper gradient propagation.

Another feature of the model is that in the generator, except for the tangent h function used for output, the activation function selects ReLU function, while in the discriminator, Leaky ReLU function is selected to prevent gradient sparse selection [74]. The reason for the use of the tangent h function is briefly explained here. Since the pixel range is 0 to 255, the ReLU function's result may exceed this range. It is beneficial to fix the final output value with a function whose range is -1 to 1.

The birth of DCGAN realized unsupervised feature extraction, and the image realized the addition and subtraction function similar to the word vector and this idea is widely used in image synthesis.

4.3 Wasserstein GAN (WGAN)

In the previous part, we constructed a relatively stable network structure by introducing DCGAN. While Martin Arjovsky, the author of WGAN [57], did an experiment on DCGAN, in which the generator was fixed, the discriminator was trained iteratively, and the relation graph of the gradient of the generator's objective function and the number of iterations was established. From this experiment, with the iterative training of discriminator, the gradient of generator decays rapidly. It can be known that the poorly trained discriminator will make the generator gradient unstable. Sufficient training of discriminator will cause generator gradient to disappear. Therefore, the training degree of discriminator is one of the important reasons for GAN instability.

According to Goodfellow's paper, under the condition of optimal discriminator, the gradient of generator can be composed of the Kullback-Leibler (KL) divergence and the Jensen-Shannon (JS) divergence [21]. In this case, mode collapse is easily caused, which means that the generated sample focuses on part of mode and lacks diversity. For this reason, the author proposed the Earth-Mover (EM) distance and compared it with KL divergence and JS divergence, and found that the change of the EM distance was more sensitive and more useful gradient could be proposed.

The Kullback-Leibler (KL) divergence

$$KL(P_r \| P_g) = \int \log\left(\frac{P_r(x)}{P_g(x)}\right) P_r(x) d\mu(x) \tag{9}$$

The Jensen-Shannon (JS) divergence

$$JS(P_r, P_g) = KL(P_r \| P_m) + KL(P_g \| P_m) \tag{10}$$

The Earth-Mover (EM) distance or Wasserstein-1

$$W(P_r, P_g) = \inf_{\gamma \in \prod(P_r, P_g)} E_{(x,y)\sim\gamma}\left[\|x - y\|\right] \tag{11}$$

Due to the advantages of EM distance, Arjovsky tried to apply it on the GAN, borrow the

Kantorovich - Rubinstein duality theory [75] will get EM distance formula for the deformation, and then adapt the method of neural network to solve, the writer named the neural network Critic. To satisfy the following equation, the Critic truncates the parameter to a range called weight clipping after each update.

$$W(P_r, P_\theta) = \sup_{\|f\|_L \le 1} E_{x \sim P_r}[f(x)] - E_{x \sim P_\theta}[f(x)] \tag{12}$$

$$\max_{w \in W} E_{x \sim P_r}[f_w(x)] - E_{z \sim P_{(z)}}[f_w(g_\theta(z))] \tag{13}$$

There are three main differences between the Critic and the traditional discriminator of GAN. First, since the discriminator is responsible for dichotomies in GAN, and the Critic's function is to fit EM distance, sigmoid function is removed, the probability is no longer output, but the general score. Then, the target function of the Critic no longer contains log functions. Last but not least, there is no need to worry about the effect of excessive training of the discriminator on the generation of samples. The more training of the Critic, the better samples will be generated.

However, WGAN also has problems such as difficult convergence and poor sample quality, so the researchers finally focused on the weight clipping method in WGAN. In other words, this method will lead to the weakening of model generation ability or gradient explosion (disappeared). Gradient penalty was considered as one of the ways to speed up a solution to the problem, and WGAN-GP [58] was proposed.

The EM distance is proposed under the Lipschitz constraint, and the Lipschitz constraint requires that the discriminator gradient not exceed K. The authors of WGAN-GP believe that Lipschitz restriction does not need to be added to the entire sample space. Instead, they focus on the generated samples, real samples and their intermediate regions.

$$L = \underset{\hat{x} \sim P_g}{E}[D(\tilde{x})] - \underset{x \sim P_r}{E}[D(x)] + \lambda \underset{\hat{x} \sim P_{\hat{x}}}{E}\left[\left(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1\right)^2\right] \tag{14}$$

The first two parts on the right of the equal sign of the above equation are the loss of the Critic, and the third part is the added gradient penalty item proposed by the author. By adding gradient penalty term, WGAN-GP improves the slow convergence of WGAN model and the training speed.

The experiment from this paper shows that in terms of Generator iterations and Wall clock time (in seconds), the performance of WGAN-GP is close to that of DCGAN and far better than that of WGAN. At the same time, due to the balance problem between discriminator and generator in DCGAN, WGAN-GP is a better choice.

5. Classical applications of GAN

Multimedia technology refers to the comprehensive processing and management of text, data, image, animation, sound and other media information through the computer, enabling the user to interact with the computer through a variety of senses for real-time information [76]. As GAN's application direction is mainly in natural language processing and computer vision [77], and it is closely combined with many aspects of multimedia technology, GAN's birth provides infinite prospects for the development of multimedia technology. In this section, we introduce

the application of GAN in real life by using multimedia technology such as text, audio, image and video as the carrier.

5.1 Natural Language Processing (NLP)

Natural language processing is the study of normal communication between humans and computers using natural language [78]. At present, NLP is mostly based on statistical machine learning and applied in emotional processing, machine translation, text extraction and other directions [79-81]. However, NLP did not make great progress in the early days of GAN. The reason is that GAN is mainly applied to continuous data, while text is mainly discrete data. According to the discriminant result, the discriminator will give feedback to the generator after the sequence generated by the generator is inputted. With the efforts of researchers, GAN has made some achievements on NLP in recent years.

5.1.1 Text Processing

Reinforcement learning is often used in NLP. Yu et al. raised a model named SeqGAN, which innovatively combined RL with GAN and made a breakthrough in the development of GAN on NLP [82]. SeqGAN treats error as a reward for RL and the generation of a sequence as a sequential decision makes process. The strategy gradient algorithm is used in reinforcement learning. The result was a Chinese poem and a collection of Obama's speeches.

Li et al. followed Yu's steps and proposed Dialog Generation [83]. The model adopted Seq2Seq instead of the generator. By entering the historical Dialogue, RNN generated each word to answer the question one by one, and finally got the reply message. This paper shows the generated statement effect, further proving GAN's infinite potential on NLP.

In the field of information retrieval (IR), Wang et al. proposed IRGAN [84]. The model transforms the generator and discriminator into two IR systems according to the characteristics of IR, generates the model to predict the related documents, and discriminant model to judge the correlation between the given documents. Under GAN framework, RL method based on strategy gradient is still adopted, which greatly improves the performance of information retrieval.

5.1.2 Audio Generated

With the development of advanced learning, the ability of computer natural language processing is gradually improved, but it is seldom applied in audio processing. Previous audio generation methods are generally based on text. This approach requires humans to record a large number of voice databases, which is inefficient and produces unnatural audio [85]. In 2016, Google DeepMind proposed a deep generation model of raw audio waveforms called WaveNet [86], which chose to directly model the original waveform of audio signals, expanding the variety of audio and increasing the authenticity of audio generation. However, since the audio has a sequence, WaveNet, which belongs to the autoregressive model, needs a long time to conduct continuous sampling, and researchers have been trying more generation methods, such as MelodyRNN, DeepBach and so on [87].

Due to GAN's high efficiency and quality in image generation, researchers have been trying to use GAN to generate music. SeqGAN proposed by Yu et al. in the last session can also generate audio, but it does not show the generated samples. Inspired by WaveNet, Yang et al. proposed a MidiNet model that combines CNN with GAN and generates a music that reaches the level of MelodyRNN in realism and pleasant [88]. Jesse et al. proposed a method to quickly generate high fidelity audio, which could be 50,000 times faster than current most commonly

used WaveNet methods named GANSynth which adopts the architecture of Progressive GAN [89]. Different from WaveNet, GANSynth adopts the method of parallel sequence generation, and makes use of convolution to generate audio fragments on a single potential vector, so as to separate global features such as pitch and timbre.

5.2 Computer Vision (CV)

The goal of CV is to enable computers and robots to achieve the ability of human vision to track and measure objects [22]. It also uses image processing to allow computers to produce images that are easier for humans and machines to recognize. In the field of computer vision, the generated samples have undergone countless times of self-comparison and optimization before being compared with other samples as a result of GAN's own counter mechanism, and the performance of GAN in image processing is amazing [90].

5.2.1 Image Synthesis

Image synthesis literally means the process of synthesizing some existing images into new ones. At present, it is mainly applied to human faces and natural scenes. In 4.2, we mentioned that the idea of vector addition and subtraction of DCGAN is applied in image synthesis, the model obtained a group of smiling male images through multiple images of smiling female and natural male and female facial expressions [71].

It has always been a difficult problem for image synthesis to generate a face from a person's side face. Two-Pathway Generative Adversarial Network (TP-GAN) was proposed by Huang et al. solved this problem [91]. The feature of the model is that the generator contains two paths, one infers global structure, the other infers local texture, and the final features of the two paths are fused.

5.2.2 Image Conversion

Image conversion is different from the traditional GAN image generation to pursue high quality. It aims to generate images of another style and pursues the diversity of generated images. The best known are Pixel to Pixel (Pix2pix) and CycleGAN.

Phillip et al. [92] proposed that Pix2pix adopted CGAN as the basic framework, and obtained the desired sample style by adding the requirements of tag generation to the generator. Different from CGAN, Pix2pix input is not noise but image. Meanwhile, the discriminator adopts PatchGAN model, instead of distinguishing the authenticity of the whole image, it divides the image into N×N patches for distinguishing, which improves the operation speed. CycleGAN [93] creatively mirrors the GAN structure, generating a network sharing two generators with a discriminator on each side. Compared with Pix2pix, which requires certain correlation between images and labels, this model realizes simultaneous training of two uncorrelated data sets, so it can meet the requirements of image generation of various styles.

5.2.3 Super- Resolution (SR)

Super-resolution usually is the reconstruction of one or more low-resolution images to generate high-resolution images [94]. In deep learning, super-resolution adopts SRCNN, DRCN and other methods [95]. In terms of super-resolution reconstruction, Wang et al. proposed Pix2pixHD [96], which further optimized Pix2pix by using a coarse-thin generator, a multi-scale discriminator structure and a robust antagonism learning objective function to achieve the purpose of high-resolution reconstruction and generate high-resolution images. In addition, SRGAN [97] based on GAN model training uses generator to generate detailed parts

of images, and adopts perceptive loss function and counter loss function to increase the sense of reality of images. This method can not only be applied in the sharpening of old photos, but also to the interface optimization of some early games. Subsequently, based on the idea of SRGAN, ESRGAN [98] and other optimization models with clearer edge images were gradually proposed, which promoted the development of SR.

5.3 Domain Transformation

Text to image is an input sentence to generate an image. Due to GAN's excellent performance in graphic image processing, many new models have been created on the result of this kind of domain transformation. Scott et al. proposed GAN-INT-CLS [99], which generates a network that inputs text features to obtain images. Then, GAWWN [100] was proposed by this team to improve the accuracy of the generated images. In terms of image to text, Liu et al. proposed the method of generating poetry through images, which for the first-time incorporated image processing and poetry generation into a framework, making the cognition of machines have the ability to approach human beings [101].

6. Security applications of GAN

AI is a double-edged sword, not only should we protect the security of AI technology implementation, but we adopt AI technology to create a safe environment. This part will introduce the applications of GAN in security.

6.1 Information Security

Information is universal, shareable, value-added, manageable and multifunctional, which makes it especially important for human beings [102]. The meaning of information security is to protect all kinds of information resources from all kinds of threats, interference and destruction [103, 104]. The game mechanism of attackers and defenders is similar to GAN, which has made some contributions to the improvement of information security, especially in cryptography.

In 2016, Google proposed an encryption technology based on GAN [105], which could effectively solve the data protection problem in the process of data sharing. Cryptography can be both defensive and offensive, and GAN is also applied to decryption technology. Briland et al. raised a way of cryptography generation based upon machine learning theory to replace artificially generated cryptography rules, which was named PassGAN [106]. By using the leaked password list as the real sample to train discriminator, the sample generated by the generator will be closer and closer to the real user's password and complete the password guessing process. In 2018, Aidan N came up with an unsupervised method of deciphering the code, named cipherGAN [107]. After the training of unmatched plaintext ciphertext, this method can decode the Caesar shift code or Virginia code with high fidelity. Inspired by CycleGAN, this model adopts unmatched plaintext and ciphertext, and completes the key decoding of long word level without parallel text.

6.2 Cyber Security

Nowadays, Cyber Security has attracted more and more attention from researchers because of the progress of big data, Internet of things (IoT), blockchain and other hot spots [108]. At the same time, the gradually increasing numbers of network anomalies threaten the normal operation of the network, such as Challenge Collapsar (CC) attack, distributed denial of

service (DDoS) attack, malware, worm [109]. However, network abnormal behavior is not a simple image or text, and it is difficult to process with GAN. Meanwhile, as the detection technology of abnormal network behaviors, the researchers turn to the source of the attack and plan to use the method of generating abnormal behavior samples to simulate the attack, so as to improve the detection ability of the existing detection technology.

Hu et al. proposed MalGAN [110], a generation model of malware. This model used a neural network-based alternative detector to match the black box detection to generate samples that could fool the detector so as to by passing the black box detection. The DeepDGA [111] algorithm can generate a large number of pseudo-random domain names through the training of GAN. Kim et al. proposed tDCGAN [112], which turned the malicious software into pictures, adopted self-encoder as the generator of GAN, and finally trained discriminator that adopted transfer learning method to detect zero-day malicious software.

6.3 AI Security

In the development of artificial intelligence, deep learning is one of the most critical technologies in the field of artificial intelligence. However, researchers have found that deep learning algorithms are vulnerable to attacks against samples, so it is urgent to improve the robustness and security of the algorithms. Akhtar et al. [113] summarized three main directions of deep learning algorithm against adversarial attacks:

(a) Using modified training or modified input.

(b) Modifying networks.

(c) Using external models as network

Currently, most of the defenses against such attacks are gradient regularization/masking, but researchers have shown that such methods can be circumvented. Bao et al. [114] propose FBGAN to capture the semantic features of the input and filter the non-semantic perturbation, after pre-training and bidirectional mapping, the adversarial data was de-noised and classified to effectively weaken the adversarial attack. Experiments have shown the effectiveness of defense.

7. Summary and perspectives

In recent years, GAN model has flourished in natural language processing, computer vision and other aspects, and has been deeply involved in medicine, physics, security and other fields. The arrival of AI era has raised people's spiritual demands, which has promoted the development of GAN. However, GAN also has many shortcomings, which leads to many optimization models of GAN put forward in recent years. The optimization of GAN is mainly modified and improved from the loss function and structure. We list three optimization models (CGAN, DCGAN, WGAN) that have a profound impact on the development of GAN, and introduce in detail the principles, processes, and variations of these models. According to previous studies, GAN has unlimited development potential. However, there are still some problems yet to be solved in the development of GAN.

The first problem is the convergence problem of GAN. In equation (3), we can theoretically reach Nash equilibrium through iterative training according to the idea of game theory. In reality, the loss function of generator and discriminator, we need to add many parameters to reach the final balance. The two main cases of non-convergence are as follows. On the one hand, the mode collapse means that the generation model cannot solve the problem of sample diversity in GAN function. When the generator is optimized, the gradient descent method is adopted to generate some similar samples that are easy to fool the discriminator. The

generation of some similar or identical samples will affect the diversity of the final generated model. On the other hand, gradient disappearance is caused by too many layers or loss function. Since there is a big gap between the initial generation of samples and the real samples, the discriminator can easily achieve a high recognition ability, which leads to the generator being prone to the gradient cannot be updated or even disappear. How to find a better method with a lower cost in the future is the problem that GAN needs to solve.

The second is the quality problem of GAN generating samples. The method of characterizing VQ-VAE [115, 116] model proposed by the Google team has surpassed the image generation ability of GAN`s models. The Glow model generated along the flow-based model has gained a foothold in the development of deep learning due to its reversible advantages over GAN. On the one hand, the specific solution is to combine the advantages of other generated models to produce hybrid models. On the other hand, better evaluation criteria are proposed. In section 3.3.2, we introduced some GAN evaluation indicators, but they all have corresponding shortcomings, such as IS cannot solve the over-fitting situation, FID cannot solve the spatial relationship of sample characteristics, etc. How to put forward a better and more identical evaluation index is also an urgent problem to be solved.

Last but not least, GAN model has limitations in application. In the fifth part, we mainly introduce the scene and object of classical GAN application. Due to its limitations in discrete data processing, GAN's achievements in natural language processing are often not as outstanding as those in computer vision. Therefore, GAN is mainly applied in image repair, resolution improvement, medical detection, pedestrian recognition and other directions in daily life. In the sixth part, we mentioned that GAN has great potential in the field of security. At present, few people adopt GAN's idea to achieve a good network anomaly detection method. In the face of DDoS attack and other network attacks with high speed and large flow, timely defense will play a key role in the development of the entire Internet [117]. The application of GAN to the detection of network abnormal behavior such as DDOS attack has a great prospect.

In general, GAN is still one of the hot spots in the field of artificial intelligence. The development of GAN in the future is still expected to optimize the model structure and expand the application field [118]. On the one hand, researchers will propose a more stable internal structure to solve the problem of training and mode collapse. On the other hand, with the increasing application of GAN in the security field, more and more security problems will be solved. The combination of GAN and other deep learning models will also improve its applicability in the security field.

8. Conclusion

GAN has excellent performance in natural language processing and computer vision in recent years. With the continuous improvement of artificial intelligence, cyber security and even artificial intelligence security are paid more and more attention by researchers. While deep learning models such as convolutional neural network and recurrent neural network have been applied in the security field, the existing literature demonstrates that GAN can play a better role than other models in information encryption and decryption, simulation of attack data and improvement of model robustness.

This paper explains the related theories and the evolution of GAN. It introduces multiple models based on GAN and the latest development of classical applications. At the meanwhile, the paper presents the great potential of GAN in the field of security. By analyzing the latest development of GAN, this paper summarizes the three existing problems of GAN and hopes that GAN's powerful generating ability will be able to solve the problem of missing or

unbalanced attack samples in the future.

Acknowledgement

This work was supported by the Hainan Provincial Natural Science Foundation of China (Grant No. 2019RC041 and 2019RC098), National Natural Science Foundation of China (Grant No. 61762033), Opening Project of Shanghai Trusted Industrial Control Platform (Grant No. TICPSH202003005-ZC), Education and Teaching Reform Research Project of Hainan University (Grant No. hdjy1970) and Innovative research project for Graduate students in Hainan Province (Grant No. Hys2020-85).

References

[1] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature*, vol. 529, no. 7587, pp. 484-489, 2016. Article (CrossRef Link)

[2] W. Li, C. W. Pan, R. Zhang, J. P. Ren, Y. X. Ma, J. Fang, F. L. Yan, Q. C. Geng, X. Y. Huang, H. J. Gong, W. Xu, G. Wang, D. Manocha, and R. G. Yang, "AADS: Augmented autonomous driving simulation using data-driven algorithms," *Science Robotics*, vol. 4, no. 28, 2019. Article (CrossRef Link)

[3] Y. Cheng, L. Jiang, and W. Macherey, "Robust Neural Machine Translation with Doubly Adversarial Inputs," *arXiv preprint arXiv*:*1906.02443*, pp. 4324-4333, 2019. Article (CrossRef Link)

[4] J. He, S. L. Baxter, J. Xu, X. Zhou, and K. Zhang, "The practical implementation of artificial intelligence technologies in medicine," *Nature Medicine*, vol. 25, no. 1, pp. 30-36, 2019. Article (CrossRef Link)

[5] B. Y. Liu, L. J. Wang, and M. Liu, "Lifelong federated reinforcement learning: a learning architecture for navigation in cloud robotic systems," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4555-4562. Article (CrossRef Link)

[6] T. Baltrušaitis, C. Ahuja, and L. P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423-443, 2018. Article (CrossRef Link)

[7] K. T. Butler, D. W. Davies, C. Hugh, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," *Nature*, vol. 559, pp. 547-555, 2018. Article (CrossRef Link)

[8] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85-117, 2015. Article (CrossRef Link)

[9] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Q. Zhang, "Transfer learning using computational intelligence: A survey," *Knowledge-Based Systems*, vol. 80, pp. 14-23, 2015. Article (CrossRef Link)

[10] L. Shao, F. Zhu, and X. Li, "Transfer Learning for Visual Categorization: A Survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 5, pp. 1019-1034, 2015. Article (CrossRef Link)

[11] X. Zhang, Y. Zhuang, W. Wang, and W. Pedrycz, "Transfer Boosting with Synthetic Instances for Class Imbalanced Object Recognition," *IEEE Transactions on Cybernetics*, vol. 48, no. 1, pp. 357-370, 2016. Article (CrossRef Link)

[12] S. J. Gershman and N. D. Daw, "Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework," *Annual Review of Psychology*, vol. 68, no. 1, pp. 101-128, 2017. Article (CrossRef Link)

[13] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep Reinforcement Learning: A Brief Survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26-38, 2017. Article (CrossRef Link)
[14] A. I. Kadhim, "Survey on supervised machine learning techniques for automatic text classification," *Artificial Intelligence Review*, vol. 52, pp. 273-292, 2019. Article (CrossRef Link)
[15] L. Dong, J. Wesseloo, Y. Potvin, and X. Li, "Discrimination of Mine Seismic Events and Blasts Using the Fisher Classifier, Naive Bayesian Classifier and Logistic Regression," *Rock Mechanics and Rock Engineering*, vol. 49, pp. 183-211, 2016. Article (CrossRef Link)
[16] K. Khosravi, B. T. Pham, K. Chapi, A. Shirzadi, H. Shahabi, I. Revhaug, I. Prakash, and D. T. Bui, "A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran," *The Science of the total environment*, vol. 627, pp. 744-755, 2018. Article (CrossRef Link)
[17] J. Tian, C. Morillo, M. H. Azarian, and M. Percht, "Motor Bearing Fault Detection Using Spectral Kurtosis-Based Feature Extraction Coupled With K-Nearest Neighbor Distance Analysis," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 3, pp. 1793-1803, 2016. Article (CrossRef Link)
[18] M. Balachandran, T. H. Shin, and L. Gwang, "PVP-SVM: Sequence-Based Prediction of Phage Virion Proteins Using a Support Vector Machine," *Frontiers in Microbiology*, vol. 9, pp. 476, 2018. Article (CrossRef Link)
[19] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised Deep Feature Extraction for Remote Sensing Image Classification," *IEEE Transactions on Geoscience & Remote Sensing*, vol. 54, no. 3, pp. 1349-1362, 2015. Article (CrossRef Link)
[20] S. J. Wetzel, "Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders," *Physical Review*, vol. 96, no. 2, 2017. Article (CrossRef Link)
[21] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Proc. of International Conference on Neural Information Processing Systems*, 2014. Article (CrossRef Link)
[22] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53-65, 2017. Article (CrossRef Link)
[23] M. Gong, X. Niu, P. Zhang, and Z. Li, "Generative Adversarial Networks for Change Detection in Multispectral Imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2310-2314, 2017. Article (CrossRef Link)
[24] Q. Wei, R. Song, and P. Yan, "Data-Driven Zero-Sum Neuro-Optimal Control for a Class of Continuous-Time Unknown Nonlinear Systems with Disturbance Using ADP," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 2, pp. 444-458, 2016. Article (CrossRef Link)
[25] J. Zheng, Y. Cai, Y. Wu, and X. Shen, "Dynamic computation offloading for mobile cloud computing: A stochastic game-theoretic approach," *IEEE Transactions on Mobile Computing*, vol. 18, no. 4, pp. 771-786, 2019. Article (CrossRef Link)
[26] K. Zhu, "Bootstrapping the portmanteau tests in weak auto-regressive moving average models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 78, no. 2, pp. 463-485, 2016. Article (CrossRef Link)
[27] H. Akaike, "Fitting autoregressive models for prediction," *Annals of the Institute of Statistical Mathematics*, vol. 21, no. 1, pp. 243-247, 1969. Article (CrossRef Link)
[28] A. Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," in *Proc. of the 33rd International Conference on International Conference on Machine Learning*, vol. 48, pp. 1747-1756, 2016. Article (CrossRef Link)
[29] L. Mou, P. Ghamisi, and X. Zhu, "Deep Recurrent Neural Networks for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3639-3655, July 2017. Article (CrossRef Link)
[30] K. Gregor, I. Danihelka, A. Graves, D. J. Rezende, and D. Wierstra, "DRAW: A Recurrent Neural Network for Image Generation," *Computer Science*, vol. 37, pp. 1462-1471, 2015. Article (CrossRef Link)

[31] A. Oord, N. Kalchbrenner, O. Vinyals, L. Espeholt, A. Graves, and K. Kavukcuoglu, "Conditional image generation with pixelcnn decoders," *Advances in neural information processing systems*, pp. 4797-4805, 2016. Article (CrossRef Link)
[32] T. Salimans, A. Karpathy, X. Chen, and D. P. Kingma, "Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications," *arXiv preprint arXiv:* 1701.05517, 2017. Article (CrossRef Link)
[33] C. Nash, Y. Ganin, S. M. Ali Eslami, and P. W. Battaglia, "PolyGen: An autoregressive generative model of 3D meshes," *arXiv preprint arXiv:2002.10880*, Feb. 2020. Article (CrossRef Link)
[34] D. E. Rumelhart, "Learning Representations by Back-Propagating Errors," *Nature*, vol. 323, pp. 533-536, 1986. Article (CrossRef Link)
[35] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013. Article (CrossRef Link)
[36] H. Li and S. Misra, "Prediction of Subsurface NMR T2 Distributions in a Shale Petroleum System Using Variational Autoencoder-Based Neural Networks," *IEEE Geoscience & Remote Sensing Letters*, vol. 14, no. 12, pp. 23995-2397, 2017. Article (CrossRef Link)
[37] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lioret, "Conditional variational autoencoder for prediction and feature recovery applied to intrusion detection in iot," *Sensors*, vol. 17, no. 9, 2017. Article (CrossRef Link)
[38] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 859-877, 2017. Article (CrossRef Link)
[39] W. Cao, X. Wang, Z. Ming, and J. Gao, "A Review on Neural Networks with Random Weights," *Neurocomputing*, vol. 275, pp. 278-287, 2017. Article (CrossRef Link)
[40] L. Zhang and P. N. Suganthan, "A Survey of Randomized Algorithms for Training Neural Networks," *Information Sciences*, vol. 364, pp. 146-155, 2016. Article (CrossRef Link)
[41] S. Wang, T. Z. Huang, J. Liu, and X. G. Lv, "An alternating iterative algorithm for image deblurring and denoising problems," *Communications in Nonlinear Science and Numerical Simulation*, vol. 19, no. 3, pp. 617-626, 2014. Article (CrossRef Link)
[42] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. of the IEEE conference on computer vision and pattern recognition*, pp. 1-9, 2015. Article (CrossRef Link)
[43] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proc. of International Conference on Machine Learning*, pp. 448-456, 2015. Article (CrossRef Link)
[44] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. of Thirty-First AAAI Conference on Artificial Intelligence*, 2017. Article (CrossRef Link)
[45] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," *Advances in neural information processing systems*, 2016. Article (CrossRef Link)
[46] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, pp. 211-252, 2015. Article (CrossRef Link)
[47] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in Neural Information Processing Systems*, 2018. Article (CrossRef Link)
[48] Q. Xu, G. Huang, Y. Yuan, C. Guo, Y. Sun, F. Wu, and K. Weinberger, "An empirical study on evaluation metrics of generative adversarial networks," *arXiv preprint arXiv:1806.07755*, 2018. Article (CrossRef Link)
[49] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet, "Are GANs created equal? a large-scale study," *Advances in neural information processing systems*, vol. 1, 2017. Article (CrossRef Link)

[50] M. Sion, "On General Minimax Theorems," *Pacific Journal of Mathematics*, vol. 8, no. 1, pp.171-176, 1958. Article (CrossRef Link)
[51] T. Chavdarova and F. Fleuret, "SGAN: An alternative training of generative adversarial networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9407-9415, 2018. Article (CrossRef Link)
[52] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proc. of the IEEE international Conference*, vol. 86, no. 11, pp. 2278-2324, 1998. Article (CrossRef Link)
[53] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017. Article (CrossRef Link)
[54] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Neural information processing systems*, vol. 60, no. 6, 2012. Article (CrossRef Link)
[55] Z. Liu, L. Ping, X. Wang, and X. Tang, "Deep Learning Face Attributes in the Wild," in *Proc. of IEEE International Conference on Computer Vision*, pp. 3730-3738, 2015. Article (CrossRef Link)
[56] X. Mao, Q. Li, H. Xie, R. Lau, Z. Wang, and S. P. Smolly, "Least squares generative adversarial networks," in *Proc. of the IEEE International Conference on Computer Vision*, pp. 2813-2821, 2017. Article (CrossRef Link)
[57] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *arXiv preprint arXiv: 1701.07875*, 2017. Article (CrossRef Link)
[58] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein GANs," *Advances in neural information processing systems*, 2017. Article (CrossRef Link)
[59] N. Kodali, J. Abernethy, J. Hays, and Z. Kira, "On convergence and stability of GANs," *arXiv preprint arXiv:1705.07215*, 2017. Article (CrossRef Link)
[60] D. Berthelot, T. Schumm, and L. Metz, "BEGAN: Boundary equilibrium generative adversarial networks," *arXiv preprint arXiv:1703.10717*, 2017. Article (CrossRef Link)
[61] T. Che, Y. Li, A. P. Jacob, Y. Bengio, and W. Li, "Mode regularized generative adversarial networks," *arXiv preprint arXiv:1612.02136*, 2017. Article (CrossRef Link)
[62] A. Slimene and E. Zagrouba, "Kernel maximum mean discrepancy for region merging approach," in *Proc. of International Conference on Computer Analysis of Images and Patterns*, vol. 8048, pp. 475-482, 2013. Article (CrossRef Link)
[63] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21-27, 1967. Article (CrossRef Link)
[64] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, "Recent Progress on Generative Adversarial Networks (GANs): A Survey," *IEEE Access*, vol. 7, pp. 36322-36333, 2019. Article (CrossRef Link)
[65] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," *arXiv preprint arXiv:1411.1784*, 2014. Article (CrossRef Link)
[66] A. Odena, "Semi-Supervised learning with generative adversarial networks," *arXiv preprint arXiv:1606.01583*, 2016. Article (CrossRef Link)
[67] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. of the 34th International Conference on Machine Learning*, vol. 1, pp. 2642-2651, 2017. Article (CrossRef Link)
[68] H. Nielsen, "Theory of the backpropagation neural network," *Neural Networks*, vol. 1, no. 1, p. 445, 1988. Article (CrossRef Link)
[69] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in neural information processing systems*, pp. 1097-1105, 2012. Article (CrossRef Link)
[70] A. Yuan, B. Gang, L. Jiao, and Y. Liu, "Offline Handwritten English Character Recognition Based on Convolutional Neural Network," in *Proc. of 2012 10th IAPR International Workshop on Document Analysis Systems*, pp. 125-129, 2012. Article (CrossRef Link)

[71] E. L. Denton, S. Chintala, and R. Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks," *Advances in neural information processing systems*, pp. 1486-1494, 2015. Article (CrossRef Link)
[72] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," *Computer Science*, 2016. Article (CrossRef Link)
[73] Y. Li, N. Y. Wang, J. P. Shi, X. Hou, and J. Liu, "Adaptive Batch Normalization for practical domain adaptation," *Pattern Recognition*, vol. 80, pp. 109-117, 2018. Article (CrossRef Link)
[74] B. Xu, N. Y.Wang, T. Q. Chen, and M. Li, "Empirical Evaluation of Rectified Activations in Convolutional Network," *Computer Science*, 2015. Article (CrossRef Link)
[75] J. Lellmann, D. A. Lorenz, C. Schönlieb, and T. Valkonen, "Imaging with Kantorovich-Rubinstein discrepancy," *Siam Journal on Imaging Sciences*, vol. 7, no. 4, pp. 2833-2859, 2014. Article (CrossRef Link)
[76] X. Chang, Z. Ma, Y. Yang, Y. Zeng, Z. Hauptmann, and G. Alexander, "Bi-Level Semantic Representation Analysis for Multimedia Event Detection," *IEEE Transactions on Cybernetics*, pp. 1-18, 2016. Article (CrossRef Link)
[77] S. R. Zhou, M. L. Ke, and P. Luo, "Multi-Camera Transfer GAN for Person Re-Identification," *Journal of Visual Communication and Image Representation*, vol. 59, no. 1, pp. 393-400, 2019. Article (CrossRef Link)
[78] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Computational intelligence magazine*, vol. 13, no. 3, pp. 55-75, 2018. Article (CrossRef Link)
[79] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. Alsaadi, "A Survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11-26, 2017. Article (CrossRef Link)
[80] R. Xia and Z. Ding, "Emotion-Cause Pair Extraction: A New Task to Emotion Analysis in Texts," in *Proc. of the 57th Annual Meeting of the Association for Computational Linguistics*, pp1003-1012, 2019. Article (CrossRef Link)
[81] Y. Li, Q. Pan, S. Wang, T. Yang, and E. Cambria, "A Generative Model for Category Text Generation," *Information Sciences*, vol. 450, pp. 301-315, 2018. Article (CrossRef Link)
[82] L. Yu, W. Zhang, J. Wang, and Y. Yu, "SeqGan: Sequence generative adversarial nets with policy gradient," in *Proc. of Thirty-First AAAI Conference on Artificial Intelligence*, pp. 2852-2858, 2017. Article (CrossRef Link)
[83] J. Li, W. Monroe, T. Shi, S. Jean, A. Ritter, and D. Jurafsky, "Adversarial Learning for Neural Dialogue Generation," in *Proc. of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 2157-2169, 2017. Article (CrossRef Link)
[84] J. Wang, L.Yu, W. Zhang, Y. Gong, Y. Xu, B. Wang, P. Zhang, and D. Zhang, "Irgan: A minimax game for unifying generative and discriminative information retrieval models," in *Proc. of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 515-524, 2017. Article (CrossRef Link)
[85] Y. Xu, J. Du, L. R. Dai, and C. H. Lee, "A Regression Approach to Speech Enhancement Based on Deep Neural Networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 1, pp. 7-19, 2015. Article (CrossRef Link)
[86] A. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016. Article (CrossRef Link)
[87] G. Hadjeres, F. Pachet, and F. Nielsen, "Deepbach: a steerable model for bach chorales generation," in *Proc. of the 34th International Conference on Machine Learning*, vol. 50, pp. 1362-1371, 2017. Article (CrossRef Link)
[88] L. C. Yang, S. Y. Chou, and Y. H. Yang, "MidiNet: A convolutional generative adversarial network for symbolic-domain music generation," *arXiv preprint arXiv:1703.10847*, 2017. Article (CrossRef Link)

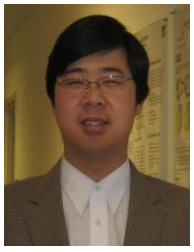
[89] J. Engel, K. Agrawal, S. Chen, I. Gulrajani, C. Donahue, and A. Roberts, "GANsynth: Adversarial neural audio synthesis," *arXiv preprint arXiv:1902.08710*, 2019. Article (CrossRef Link)
[90] J. Liu, C. K. Gu, and J. Wang, "Geumran Youn, Jeong-Uk Kim, "Multi-Scale Multi-Class Conditional Generative Adversarial Network for Handwritten Character Generation," *The Journal of Supercomputing*, vol.75, no.4, pp.1922-1940, 2019. Article (CrossRef Link)
[91] R. Huang, S. Zhang, T. Li, and R. He, "Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis," in *Proc. of the IEEE International Conference on Computer Vision*, pp. 2439-2448, 2017. Article (CrossRef Link).
[92] P. Isola, J. Y. Zhu, T. Zhou, and A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. of IEEE conference on computer vision and pattern recognition*, pp. 5967-5976, 2017. Article (CrossRef Link).
[93] J. Y. Zhu, T. Park, P. Isola, and A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. of the IEEE international conference on computer vision*, pp. 2242-2251, 2017. Article (CrossRef Link)
[94] K. Nguyen, C. Fookes, S. Sridharan, M. Tistarelli, and M. Nixon, "Super-Resolution for Biometrics: A Comprehensive Survey," *Pattern Recognition*, vol. 78, pp. 23-42, 2018. Article (CrossRef Link)
[95] X. Zhang, C. Li, Q. Meng, S. Liu, Y. Zhang, and J. Wang, "Infrared image super resolution by combining compressive sensing and deep learning," *Sensors*, vol. 18, no. 8, 2018. Article (CrossRef Link)
[96] T. C. Wang, M. Y. Liu, J. Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. of the IEEE conference on computer vision and pattern recognition*, pp. 8798-8807, 2018. Article (CrossRef Link)
[97] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. of the IEEE conference on computer vision and pattern recognition*, pp. 105-114, 2017. Article (CrossRef Link)
[98] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proc. of the European Conference on Computer Vision (ECCV)*, pp. 63-79, 2018. Article (CrossRef Link)
[99] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *Proc. of the 33rd International Conference on International Conference on Machine Learning*, vol. 48, pp. 1060-1069, 2016. Article (CrossRef Link)
[100] S. Reed, Z. Akata, S. Mohan, S. Tenka, B. Schiele, and H. Lee, "Learning what and where to draw," in *Proc. of the 30th International Conference on Neural Information Processing Systems*, pp. 217-225, 2016. Article (CrossRef Link)
[101] B. Liu, J. Fu, M. P. Kato, and M. Yoshikawa, "Beyond narrative description: Generating poetry from images by multi-adversarial training," in *Proc. of 2018 ACM Multimedia Conference on Multimedia Conference*, pp. 783-791, 2018. Article (CrossRef Link)
[102] K. Zhang, J. Ni, K. Yang, X. Liang, J. Ren, and X. S. Shen, "Security and Privacy in Smart City Applications: Challenges and Solutions," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 122-129, 2017. Article (CrossRef Link)
[103] Y. Tu, Y. Lin, J. Wang, and J. U. Kim, "Semi-supervised Learning with Generative Adversarial Networks on Digital Signal Modulation Classification," *Computers Materials & Continua*, vol. 55, no. 2, pp. 243-254, May, 2018. Article (CrossRef Link)
[104] C. Zhang, J. R. Cheng, X. Y. Tang, S. Victor, Z. D. Sheng, and J. Q. Li, "Novel DDoS Feature Representation Model Combining Deep Belief Network and Canonical Correlation Analysis," *Computers, Materials & Continua*, vol. 61, no. 2, pp. 657-675, 2019. Article (CrossRef Link)
[105] M. Abadi and D. G. Andersen, "Learning to protect communications with adversarial neural cryptography," *arXiv preprint arXiv:1610.06918*, 2016. Article (CrossRef Link)
[106] B. Hitaj, P. Gasti, G. Ateniese, and F. Perez-Cruz, "Passgan: A deep learning approach for password guessing," in *Proc. of International Conference on Applied Cryptography and Network Security*, vol. 11464, pp. 217-237, 2019. Article (CrossRef Link)

[107] A. N. Gomez, S. Huang, I. Zhang, I. Zhang, B. M. Li, M. Osama, and L.Kaiser, "Unsupervised cipher cracking using discrete GANs," i*arXiv preprint arXiv:1801.04883*, 2018. Article (CrossRef Link)
[108] R. Zhang, R. Xue, and L. Liu, "Security and Privacy on Blockchain," *ACM Computing Surveys*, vol. 51, no. 3, 2019. Article (CrossRef Link)
[109] J. Cheng, C. Zhang, X. Tang, V.S. Sheng, Z. Dong, and J. Li, "Adaptive DDoS attack detection method based on multiple-kernel learning," *Security and Communication Networks*, vol. 2018, 2019. Article (CrossRef Link)
[110] W. Hu and Y. Tan, "Generating adversarial malware examples for black-box attacks based on GAN," *arXiv preprint arXiv:1702.05983*, 2017. Article (CrossRef Link)
[111] H. S. Anderson, J. Woodbridge, and B. Filar, "DeepDGA: Adversarially-tuned domain generation and detection," in *Proc. of the 2016 ACM Workshop on Artificial Intelligence and Security*, pp. 13-21, 2016. Article (CrossRef Link)
[112] J. Y. Kim, S. J. Bu, and S. B. Cho, "Zero-day Malware Detection using Transferred Generative Adversarial Networks based on Deep Autoencoders," *Information Sciences*, vol. 460, pp. 83-102, 2018. Article (CrossRef Link)
[113] N. Akhtar and A. Mian, "Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey," *IEEE Access*, vol. 6, pp. 14410-14430, 2018. Article (CrossRef Link)
[114] R. Bao, S. Liang, and Q. Wang, "Featurized bidirectional GAN: Adversarial defense via adversarially learned semantic inference," *arXiv preprint arXiv:1805.07862*, 2018. Article (CrossRef Link)
[115] A. Oord and O. Vinyals, "Neural discrete representation learning," *Advances in Neural Information Processing Systems*, 2017. Article (CrossRef Link)
[116] A. Razavi, A. Oord, and O. Vinyals, "Generating Diverse High-Fidelity Images with VQ-VAE-2," *arXiv preprint arXiv:1906.00446*, 2019. Article (CrossRef Link)
[117] R. Cheng, R. Xu, X. Tang, V. S. Sheung, and C. Cai, "An abnormal network flow feature sequence prediction approach for DDoS attacks detection in big data environment," *Computers, Materials & Continua*, vol. 55, no. 1, pp. 95-119, 2018. Article (CrossRef Link)
[118] D. J. Zeng, Y. Dai, F. Li, R. S. Sherratt, and J. Wang, "Adversarial Learning for Distant Supervised Relation Extraction," *Computers Materials & Continua*, vol. 55, no. 1, pp.121-136, 2018. Article (CrossRef Link)

Jieren Cheng received the Ph.D. degree from School of Computer, National University of Defense Technology in 2010. Now he is a professor and Ph.D supervisor at Hainan University, and the member of CCF. His research interests include cloud computing, artificial intelligence, network security and intelligent transportation, etc.

Yue Yang received the B.S. degree in information security from Hainan University, Haikou, China, in 2018. He is currently working toward the M.S. degree in software engineering at Hainan University, Hainan, China. His research interests focus on cyberspace security and machine learning.

Xiangyan Tang received the M.S. degree from School of Computer, Hunan Agricultural University in 2011. Now she is an associate professor at Hainan University, and currently working toward the Ph.D. degree in computer science and technology at Tianjin University, Tianjin, China. Her research interests include artificial intelligence, network security and intelligent transportation, etc.



Naixue Xiong received his Ph.D. degrees in Japan Advanced Institute of Science and Technology, respectively. He is currently an Associate Professor at the Department of Mathematics and Computer Science, Northeastern State University. His research interests include Cloud Computing, Security and Dependability, Parallel and Distributed Computing.

Yuan Zhang received the B.S. degree in information engineering from Southeast University, Nanjing, China, in 2019. He is currently working toward the M.S. degree in cyber space security at Hainan University, Hainan, China. His research interests focus on blockchain and machine learning.

Feifei Lei is currently working toward the B.S. degree in computer science and technology at the Hainan University, Hainan, China. Her research interests focus on computer version, and machine learning.