This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2905015, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Recent Progress on Generative Adversarial Networks (GANs): A Survey

Zhaoqing Pan[1], (Member, IEEE), Weijie Yu[1], Xiaokai Yi[1], Asifullah Khan[2], Feng Yuan[1], AND Yuhui Zheng[1], (Member, IEEE)

[1]School of Computer and Software, Jiangsu Engineering Center of Network monitoring, and Jiangsu Collaborative Innovation Center on Atmospheric, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: {zhaoqingpan, weijieyu, xyi, yuanfeng, zheng_yuhui}@nuist.edu.cn)
[2]Department of Computer and Information Sciences, Pakistan Institute of Engineering and Applied Sciences (PIEAS), Nilore, Pakistan (e-mail: asif@pieas.edu.pk)

Corresponding author: Yuhui Zheng (e-mail: zheng_yuhui@nuist.edu.cn).

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61501246, in part by the Natural Science Foundation of Jiangsu Province of China under Grant No. BK20150930, in part by the Six Talent Peaks Project of Jiangsu Province under Grant No. XYDXXJS-041, in part by the Project through the Priority Academic Program Development of Jiangsu Higher Education Institutions, in part by the Collaborative Innovation Center of Atmospheric Environment and Equipment Technology.

ABSTRACT Generative Adversarial Network (GANs) is one of the most important research avenues in the field of artificial intelligence, and its outstanding data generation capacity has received wide attention. In this paper, we present the recent progress on GANs. Firstly, the basic theory of GANs, and the differences among different generative models in recent years were analyzed and summarized. Then, the derived models of GANs are classified, and introduced one by one. Thirdly, the training tricks and evaluation metrics were given. Fourthly, the applications of GANs were introduced. Finally, the problem we need to address, and future directions were discussed.

INDEX TERMS Deep learning, Machine learning, Unsupervised learning, Generative adversarial networks

I. INTRODUCTION

THE past several years have witnessed a burgeoning development of computer science and data accumulation. Artificial intelligence (AI) is becoming a thriving field with a great number of meaningful applications and valuable research topics. In the AI community, machine learning [1] exerts a huge impact on different aspects of our daily life. All of these machine learning algorithms need the representation of the data they are given. But when one want to use this skill in other fields or tasks, it is very difficult to extract useful features. Hence, researchers proposed a new approach called representation learning [2] to automatically extract useful information when doing classification and detection. Deep learning [3] is a type of representation learning methods which can easily extract high-level, more abstract features than other methods by composing some simple representations.

Generally, according to whether the data set is labeled or not, machine learning algorithms can be divided into two categories, supervised and unsupervised learning. For the supervised learning, a dataset with diverse features is required, and each example in the dataset must be labeled. The representatives of supervised learning are classification, regression and structured output problems. However, the unsupervised learning requires a dataset with no more than the same label. The goal is to explore the special structure in this dataset. Customarily, density estimation, clustering, synthesis, and denoising are always regarded as unsupervised learning.

In case of supervised learning, it is difficult to collect or annotate labels automatically. Hence, researchers pay more attention to unsupervised learning. In the task of unsupervised learning, generative model is one of the most promising technologies. The typical generative models are usually based on Markov chains, maximum likelihood, and approximate inference. Restricted Boltzmann Machines [4] and its extension models (e.g., Deep Belief Networks (DBNs) [5], Deep Boltzmann Machinescite [6]) are always based on maximum likelihood estimation. The models generated by these methods represent distributions, and these distributions have a number of parameters that are intended to fit with the empirical distribution of the training data.

However, these early models [4]–[6] had serious limitations, they may not have good generalization. In 2014, Goodfellow et al. [7] proposed a novel generative model, named Generative Adversarial Networks (GANs). Based on

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2905015, IEEE Access

the game theory, there are two networks in GANs, one is generator, and the other is discriminator. The role of the generator is to create as realistic data as possible to deceive the discriminator. The role of discriminator tries to distinguish fake samples from real ones. In this case, we can train both models using backpropagation [8], and dropout algorithms. In addition, the approximate inference or Markov chains are unnecessary for the GANs.

This survey analyzes and summarizes the recent state-of-the-art GANs, including definition, motivations, and applications of these networks. The structure of this survey is as follows. Section II introduces several generative models and highlights the basic theory of GANs. In addition, a brief comparison of these models is provided. A series of derived GANs models are presented in Section III. Section IV provides several training tricks of GANs. Section V discusses the pros and cons of different evaluation metrics. The applications of GANs in various fields are reviewed in Section VI. Section VII discusses the limitations of GANs, and provides future suggestions. Finally, Section VIII draws conclusions.

II. GENERATIVE MODELS AND GANS

GANs are one of the deep generative models, they can well process the generative problems. In this section, we will firstly present several kinds of deep generative models, which are more popular to use, and compare the differences between these models. Then, we will introduce the theory and architecture of the basic GANs.

A. DEEP GENERATIVE MODELS

The artificial intelligence aims to achieve that the machines can understand complex world just like humans. Based on this idea, researchers of machine intelligence have proposed generative models, which are dedicated to describing the world around them in terms of probability and statistics. Now the generative models can be divided into three categories: Generative Adversarial Networks (GANs) [7], Variational Autoencoder (VAE) [9], and AutoRegressive Networks [10]. The VAE is a probabilistic graphical model, which attempts to model the probability distribution of data. However, its final probabilistic simulation has a certain bias. So it mostly generates more blurred samples than GANs. PixelRNN [11] is one of the autoregressive networks, which translates the problem of image generation into the problem of pixel prediction and generation. Therefore, each pixel needs to be processed one by one, while GANs directly process the sample in one shot, and this causes GANs to produce a sample faster than PixelRNN.

As a probabilistic generative model, when the density of probability is not provided, some of the traditional generative models that rely on the natural interpretation of data cannot be trained and applied. But GANs can still be used in this situation, because GANs introduce very clever internal adversarial training mechanism.

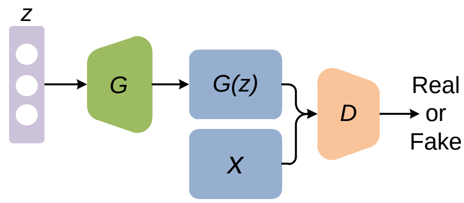


FIGURE 1: The architecture of generative adversarial networks

B. THE PRINCIPLE OF GANS

GANs were inspired by the game theory, the generator and discriminator will complete with each other to achieve the Nash equilibrium in the training processing. The architecture of GANs is illustrated in Fig. 1. The principle of generator G is to generate fake data as much as possible to fit the potential distribution of real data, while the principle of discriminator D is to correctly distinguish real data from fake data. The input of the generator is a random noise vector z (usually a uniform or normal distribution). The noise is mapped to a new data space via generator G to obtain a fake sample, $G(z)$, which is a multi-dimensional vector. And, the discriminator D is a binary classifier, it takes both the real sample from dataset, and the fake sample generated by generator G as the input, and the output of discriminator D represents the probability that the sample is a real rather than a fake. When the discriminator D cannot determine whether the data comes from the real dataset or the generator, the optimal state is reached. At this point, we obtain a generator model G, which has learned the distribution of real data.

C. LEARNING MODEL OF GANS

As two players in game theory, both the generator and the discriminator have their own loss functions. In this case, we call them $J^{(G)}$ and $J^{(D)}$, respectively. In [7], the discriminator D is defined as a binary classifier, and the loss function is represented by the cross entropy,

$$\begin{aligned} J^{(D)} = -\frac{1}{2}\mathbb{E}_{x\sim pdata}\log D\left(x\right) - \\ \frac{1}{2}\mathbb{E}_{z}\log\left(1 - D\left(G\left(z\right)\right)\right) \end{aligned} \tag{1}$$

where x represents the real sample, z represents the random noise vector, $G(z)$ is the data generated by the generator, and $\mathbb{E}$ represents the expectation. $D(x)$ indicates the probability that D discriminates x as real data, and $D(G(z))$ indicates the probability that D determines the data generated by G. The goal of D is to correctly determine the source of the data, so it wants $D(G(z))$ approach 0, while the goal of G is to bring it closer to 1. Based on this idea, there exists a conflict between these two models (i.e., zero-sum game). Therefore, the loss of the generator can be derived by the discriminator:

$$J^{(G)} = -J^{(D)} \tag{2}$$

TABLE 1: Classification of GANs models

Architecture Optimization Based GANs	Convolution based GANs	DCGAN [12]
	Condition based GANs	CGANs [13]; InfoGAN [14]; ACGAN [15]
	Autoencoder based GANs	AAE [16]; BiGAN [17]; ALI [18]; AGE [19]; VAE-GAN [20]
Objective Function Optimization Based GANs	unrolled GAN [21]; f-GAN [22]; Mode-Regularized GAN [23]; Least-Square GAN [24]; Loss-Senstive GAN [25]; EBGAN [26]; WGAN [27]; WGAN-GP [28]; WGAN-LP [29]	

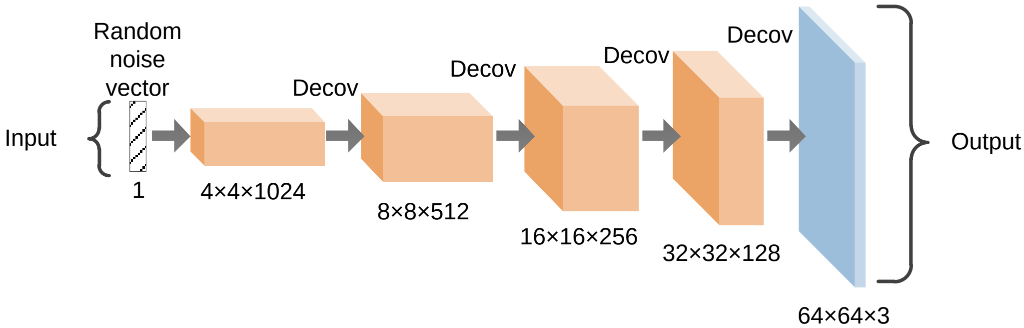


FIGURE 2: The generator of deep convolutional GANs.

Consequently, the optimization problem of GANs is transformed into the minimax game as shown below,

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim pdata(x)}\left[\log D(x)\right] + \mathbb{E}_{z\sim p(z)}\left[\log\left(1 - D\left(G\left(z\right)\right)\right)\right] \tag{3}$$

In the training process, the parameters in G are updated along with the parameters updating process in D. When $D(G(z))$=0.5, the discriminator cannot determine the differences between these two distributions, and in this status, the model will achieve the global optimal solution.

III. THE DERIVED GANS MODELS

Due to the deficiencies of original GANs, various derived GANs models were proposed, and these derived GANs models can be classified into two groups, architecture optimization based GANs, and objective function optimization based GANs, as shown in Table 1. In this section, we will introduce a series of selected derived GANs models in details.

A. ARCHITECTURE OPTIMIZATION BASED GANS

1) Convolution based GANs

Convolutional Neural Network (CNN) [30] is regarded as a very effective model of supervised learning, and is one of the most common network structures in image processing. In terms of the network structure of generator and discriminator, the original GANs adopts the Multi-Layer Perceptron (MLP) to make it work. Due to the fact that CNN is better than MLP in extracting image features, Radford et al. [12] proposed a Deep Convolutional Generative Adversarial Networks (D-CGAN). As shown in Fig. 2, this approach innovatively replaces the fully connected layer with the deconvolution layer in the generator, which achieved great performance in image generation tasks.

2) Condition based GANs

Since the input of the generator is the random noise vector z, these unrestricted inputs can lead to the collapse of the training mode. Therefore, Mirza et al. [13] proposed a Conditional Generative Adversarial Networks (CGANs), which introduced the conditional variable c (variable c can be labels, text or other data) in both generator and discriminator to add conditions to the model using additional information to affect the data generation process. In Fig. 3(a), the input of the generator is the conditional variable c and the noise vector z, the input of the discriminator is $G(z|c)$ which is from generator, and the real sample under the control of the same conditional variable c. Therefore, the objective function can be described as,

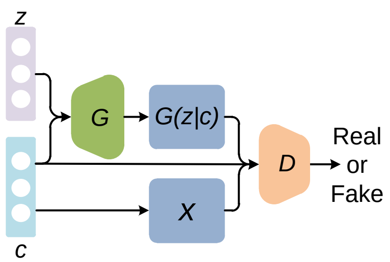
$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim pdata(x)}\left[\log D(x|c)\right] + \mathbb{E}_{z\sim p(z)}\left[\log\left(1 - D\left(G\left(z|c\right)\right)\right)\right] \tag{4}$$

In addition, Chen et al. [14] proposed another CGANs named InfoGAN. By introducing mutual information, the InfoGAN makes the generation process more controllable, and the results can be more interpreted. Where, mutual information represents the correction between the latent code c and the generated data x. In order to enhance the relationship between x and c, the value of mutual information needs to be maximized. Its generator is similar to CGANs, but the difference is that the latent code c is not known, and thus needs to be discovered through training process. In addition to the original GANs'discriminator, InfoGAN has an additional network Q to output the conditional variables $Q(c|x)$. The objective function is shown as follows,

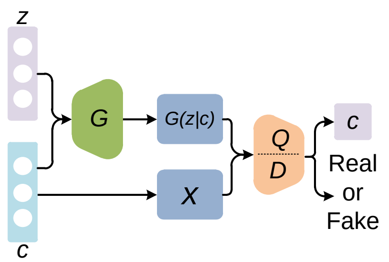
$$\min_{G} \max_{D} V(D,G) - \lambda I(c, G(z,c)) \tag{5}$$

where λ is a hyper-parameter of the constraint function $I(c, G(z,c))$, this mutual information makes the latent code c more and more reasonable for the generated data. The architecture of InfoGAN is shown in Fig. 3(b).

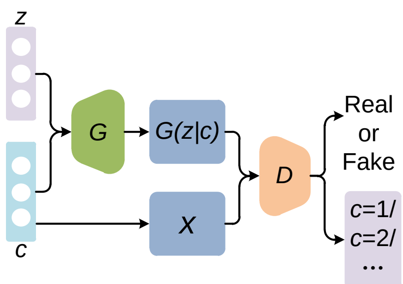
Based on CGANs, Odena et al. [15] proposed an Auxiliary Classifier GAN(ACGAN). In Fig. 3(c), for the discriminator, the condition variable c will not be added, and another classifier will be used to show the probability over the class labels. The loss function is then modified to increase the probability of correct class prediction.



(a)



(b)



(c)

FIGURE 3: The architecture of derived GANs models. (a) CGANs; (b) InfoGAN; (c) ACGAN

3) Autoencoder based GANs

Autoencoder is one type of neural networks that is trained to reconstruct the input into an output. It includes two parts, encoder $z = f(x)$ and decoder $\hat{x} = g(z)$, where the encoder is used to convert the input x (usually image, video, audio or text data) into the hidden layer (latent code z), which is a process of decreasing the dimension; the decoder is used to receive code from the hidden layer h as input. After training, the decoder attempts to reconstruct the input x as its output $\hat{x}$. It is also an unsupervised model because the labels are not required during training process. In recent years, it has been used in conjunction with latent variable model theory to apply autoencoder to generative models.

Autoencoder is also imperfect that the hidden layer obtained by the encoder is not evenly distributed in the specified space, which results in that there are a large number of gaps in the distribution. Hence, Makhzani et al. [16] proposed an Adversarial Autoencoder (AAE), combining the idea of adversarial networks with autoencoder. In this approach, the arbitrary prior distribution is imposed to the distribution of hidden layer obtained by the encoder. This is to ensure that there are no gaps in the prior distribution, so that the decoder can reconstruct meaningful samples from any part of it. The architecture of AAE is shown in Fig. 4, where the latent code z (hidden layer) represents fake data and z' represents prior with the specified distribution $p(z)$. They are the input of discriminator. After training process, the encoder can learn the distribution we expect, and the decoder can finally output the samples which are reconstructed by the required distribution.

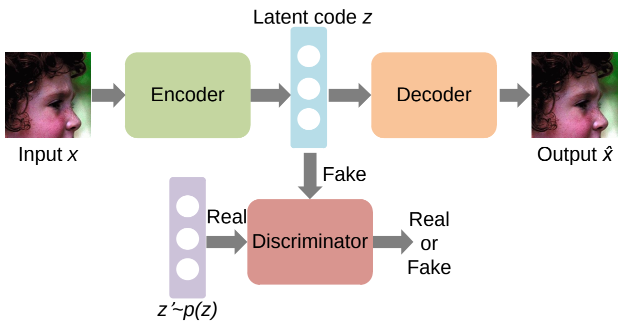


FIGURE 4: The architecture of Adversarial Autoencoder (AAE)

Some models [17]–[19] only add encoder to GANs. The generator of these models can learn the features in the latent space, and capture the semantic changes in the data distribution. However, it cannot learn the mapping from data sample distribution to latent space. To address this problem, Donahue et al. [17] proposed a Bidirectional Generative Adversarial Networks (BiGAN), which can not only make valid inferences, but also guarantee the quality of the generated samples. The architecture of BiGAN and ALI is shown in Fig. 5(a). In the architecture of BiGAN, in addition to the discriminator and generator, an encoder is added to the model, which is used to inversely map the data which was generated by GANs in the data distribution back to the latent feature space. The input of discriminator becomes a tuple consisting of data, and its corresponding latent code. For the data generated by the generator, this tuple is the generated data $G(z)$ and the noise vector z that was used to generate the data. For real samples x from dataset, the tuple are sample x

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2905015, IEEE Access

and $E(x)$ obtained from x through the inverse mapping of the encoder. In this approach, the encoder can be used as a feature capture tool for discriminator. Similarly, Dumoulin et al. [18] proposed an Adversarially Learned Inference (ALI), which uses the encoder to learn the latent feature distribution. These two approaches can both learn the generator and encoder in parallel.

In addition to the approach of combining the autoencoder and adversarial networks mentioned above, Ulyanov et al. [19] proposed an approach named Adversarial Generator-Encoder Network (AGE), in which the adversarial network acts between the generator and the encoder, and the network does not require the participation of the discriminator. Fig. 5 shows the architecture of AGE in which R represents the reconstruction loss function. In the structure of their model, the generator aims to minimize the divergence between the latent distribution z and the generated data distribution, while the encoder aims to maximize the divergence between z and $E(G(z))$, and minimize the divergence of real data x. Furthermore, they constructed the reconstruction loss function to avoid the model falling into the mode collapse.

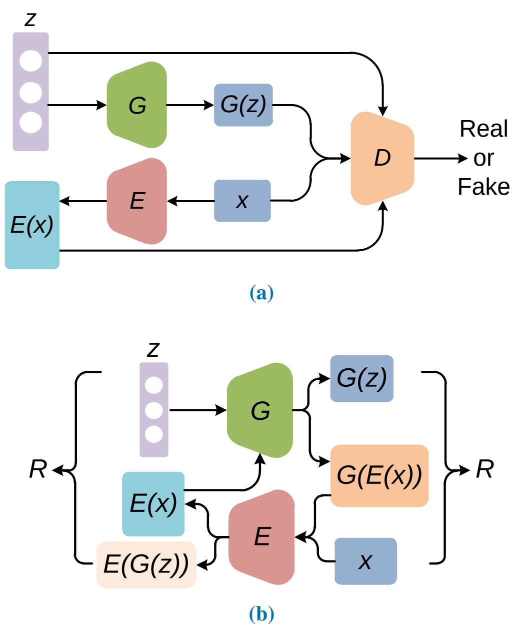


FIGURE 5: The architecture of (a) BiGAN and ALI; (b) Adversarial Generator-Encoder Network(AGE)

In Section 2, we have briefly compared VAEs with GANs. The advantage of VAEs is that they are less affected by the mode collapse, but the generated samples are blurred. The generated model based on GANs generates higher quality samples than VAEs, but it also has the problem of mode collapse. Larsen et al. [20] combined the advantages of GANs and VAEs to replace the decoder of VAEs with generator of GANs. In their approach, they combine the adversarial loss of GANs with the objective function of VAEs, which can reduce the problem of the VAEs generating blurry images, while keeping the VAEs able to learn the distribution of the latent code.

B. OBJECTIVE FUNCTION OPTIMIZATION BASED GANS

To enhance the GANs stability, many efforts [21]–[29] have been proposed by optimizing the objective function. To adjust the GANs'training processes, Metz et al. [21] proposed an unrolled GANs, which uses a gradient-based loss function to enhance the generator. And, the original GANs were achieved by minimizing the Jensen-Shannon (JS) divergence to minimize the loss function of generator. [22] pointed out that this is just a special case, and any f-divergence can be used in the architecture of GANs. [22], [24], [26] used different divergences to construct the objective function to enhance the GANs' stability.

The other methods for improving the stability of GANs are using different regularization. Che et al. [23] proposed two regularizers to make the learning more stable. If there is no overlap between the distribution of generated data and real data, or the overlap is negligible, the divergence will set to a constant. At this time, the gradient is zero, which will cause the vanishing gradient problem. In order to address the problem, Arjovsky et al. [27] proposed a Wasserstein Generative Adversarial Networks (WGAN). They firstly theoretically showed that the Earth-Mover (EM) distance produces better gradient behaviors in distribution learning compared to other distance metrics. This approach provided a weight clipping method to enforce the Lipschitz constraint, and found a novel loss metric to address the problem of unstable training process. Gulrajani et al. [28] found that the WGAN might still have unsatisfactory results or could not converge due to the use of weight clipping in discriminator. Hence, they proposed a gradient penalty named WGAN-GP to enforce the Lipschitz constraint. The method also has a better performance than the original WGAN, and enables training of various GANs architectures more stable than before with almost no hyper-parameter tuning. Furthermore, Petzka et al. [29] proposed a new penalty term to enforce the Lipschitz constraint, which is known as WGAN-LP. This method further improves the stability of network training.

In addition, the original GANs assume that the discriminator has the ability to model infinitely without any restrictions on the real sample distribution. This easily leads to overfitting and poor generalization capability. In order to limit the infinite modeling ability of GANs, Qi et al. [25] limited the loss function which is obtained by minimizing the objective function to a space that satisfies Lipschitz continuous functions. Both [25] and [27] used the Lipschitz regularity to address the problem of mode collapse and vanishing gradient. The difference is that the Lipschitz constraint of [27] comes from the Kantorovich-Rubinstein duality.

IV. TRICKS FOR TRAINING GANS

The goal of GANs is to achieve Nash equilibrium, but this is very difficult in the implementation process. This section will provide some suggestions for achieving excellent training performance.

In [31], Salimans et al. proposed several tricks which can improve the performance, and stabilize the training process. Firstly, feature matching can make GANs training more stable by giving the generator a new objective function. In this way, the generated data will be more consistent, and generator can generate more sample information. Secondly, using the minibatch layer allows the discriminator to reflect the diversity of the sample to avoid the problem of mode collapse. Thirdly, historical averaging can help the model to converge. When there is a large difference between the current value and the average value of the parameter, a term is added to the generator and discriminator to make a penalty for the current parameter. Fourthly, one-sided label smoothing is proposed to set the estimation of the discriminator for the real samples to a value adjacent to 1, which can smooth the classifier boundary.

By using separate learning rates, [32] proposed a two time-scale update rule(TTUR) for both generator and discriminator to guarantee that the model can converge to a stable local Nash equilibrium. In [33], Miyato et al. proposed a spectral normalization, which is a weight normalization technique to stabilize the training of the GAN's discriminator. Their approach is to add the Lipschitz constant as a constraint in the discriminator. Different from the weight clipping and gradient penalty in [27] and [28], they restrict the spectral norm of each layer to stabilize the training. The computational cost of this approach is small and there is no need to tune other hyper-parameters. In [34], Zhang et al. demonstrated that it is also useful to use spectral normalization in the generator.

V. EVALUATION METRICS

Recently, the GANs model has been applied to different tasks, and each task has its own evaluation metric. However, there is still no universal quantitative evaluation metrics, which will cause significant confusion about how researchers can determine evaluation metrics for different tasks. Therefore, we show several evaluation metrics that are widely used at present, and discuss their strengths and weaknesses.

(a) Inception Scores (IS) This metric has been widely used in GANs, and it was proposed by Salimans et al. [31] A higher IS indicates that the generated model can generate high quality samples, while the samples are also diverse. However, the IS also has limitations, if the generative model falls into mode collapse, the IS might be still pretty, but the real situation is very bad.

(b) Mode Score (MS) Based on the IS, Che et al. [22] proposed another evaluation metrics named MS, which can reflect the variety and visual quality of the generated samples at the same time. This evaluation metric addresses the problem of IS which is not sensitive to prior distributions over the ground truth labels.

(c) Fréchet Inception Distance (FID) The FID was proposed by Heusel et al. [32], which is used to detect the intra-class mode dropping. In this approach, the generated samples are embedded into the feature space provided by specific layer of Inception network. Based on the assumption that the generated samples follow a multi-dimensional Gaussian, the mean and covariance are computed between the generated samples and real data. Then, the Fréchet distance between these two Gaussians is measured to evaluate the quality of the generated samples. However, the IS and FID cannot well process the overfitting problem. To address this problem, the Kernel Inception Distance (KID) was proposed by Bińkowski et al. [35].

(d) Multi-scale Structural Similarity for Image Quality (MS-SSIM) Different from single scale SSIM [36] metric, which is used to measure the similarity between two images, Wang et al. [37] firstly proposed the MS-SSIM for multiple scales image quality assessment. It quantitatively evaluates the similarity of images by predicting human perception similarity judgment. Odena et al. [15] and Fedus et al. [38] used this evaluation metric to determine the diversity of generated data. [39] suggested that FID and IS can be used as the auxiliary evaluation metrics with MS-SSIM for testing sample diversity.

In addition, the Classifier Two-sample Tests (C2ST) [40] is another metric based on training a binary classifier. It assess whether different samples came from the same distribution. 1-Nearest Neighbor classifier (1-NN) [41] is the variant of C2ST which does not require special training and much hyper-parameter tuning.

Objective functions can also be used as the metrics to judge whether the mode is suitable for their problems. The Wasserstein Critic [27] and Maximum Mean Discrepancy (MMD) [42] are proposed to measure the distance between the real sample distributions and the generated sample distributions. Both of them will have a low value if the distribution between target and output are similar.

How to select an appropriate evaluation metric is still a difficult problem, [43] presented several measures as meta-measures to guide researchers to choose quantitative evaluation metrics. A good evaluation metric should distinguish generated samples from real one, verify mode collapse, mode drop and detect overfitting. We hope that there are more suitable methods for evaluating the quality of the GANs model.

VI. APPLICATIONS OF GANS

As a kind of generative model, the most direct application of GANs is data generation. That is to learn from the distribution of real samples, and generate samples consistent with the distribution. This section will introduce selected applications of GANs, including the applications in the computer vision, natural language processing, and other fields.

A. COMPUTER VISION

At present, the most successful applications of GANs are used in the computer vision areas, including image translation, image super-resolution, image synthesis and video generation, etc. The details of these applications are introduced as follows.

1) Image super-resolution

To improve the resolution of image, a Super-Resolution Generative Adversarial Networks (SRGAN) was proposed by Ledig et al. [44], which takes a low-resolution image as input, and generates a high-resolution image with 4x up-scaling. To address the problem that the texture information generated by SRGAN is not real enough, and is often accompanied by some noise, Wang et al. [45] proposed an Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN). In ESRGAN, the architecture of the network, the adversarial loss, and the perceptual loss are improved. Furthermore, a new network unit named Residual-in-Residual Dense Block (RRDB) based on the relativistic GAN [46] was introduced. A generated performance comparison is shown in Fig. 6. We can see that the ESRGAN achieves better performance than SRGAN.

2) Image Translation

To convert the image content from one domain to another, an image-to-image translation approach was proposed by Isola et al. [47] using CGANs, which is named pix2pix. Experiments have shown that pix2pix can be effective not only in graphics tasks but also in vision tasks. As a follow-up to the work, pix2pixHD [48] further improves the quality, and definition of the generated samples. This approach used a novel adversarial loss term to generate images with a resolution of 2048×1024. Pix2pix can be used for image translation problems, but it requires the training space to be strictly paired in the X and Y Spaces. However, in our daily life, such paired data is hard to find. Based on this situation, CycleGAN [49], DiscoGAN [50] and DualGAN [51] all adopt the idea of cyclic consistency, which can use unpaired data to train the mapping from X space to Y space. All these three generators are encoder-decoder framework. The difference is that different feature representations are used in the encoders and decoders. Fig. 7 presents the performance of CycleGAN.

These previous works were about image-to-image translation in two domains. Choi et al. [52] proposed a StarGAN, which can solve the problem of image translation among multi-domains by learning one model. In particular, the use of StarGAN in the tasks of facial expression synthesis and facial attribute transfer has surprising effects.

3) Texture synthesis

Texture synthesis is a very classical problem in image domain. Based on GANs, Li et al. [53] proposed a texture synthesis approach, which is named Markovian Generative Adversarial Networks (MGAN). By capturing the texture data of Markovian patches, MGAN can generate stylized images and videos in a very short time, so as to realize real-time texture synthesis. Jetchev et al. [54] proposed a Spatial GAN(SGAN), which was the first application of GANs with fully unsupervised learning in texture synthesis. As the follow-up work of SGAN, Bergmann et al. [55] proposed a Periodic Spatial GAN (PSGAN), which can learn periodic textures from a single image or complex large dataset. Besides, it can handle texture information in noise space flexibly and synthesize high-resolution textures.

4) Face synthesis

Face synthesis is also an important direction. How to generate realistic face samples has always been the problem that needs people to address. Huang et al. [56] proposed a Two-Pathway Generative Adversarial Network (TP-GAN) that can consider both global and local information like humans, and can be used to synthesize high-resolution frontal face images from a single side photo. The face image synthesized by this method can retain the identity feature well, and can process a large number of photos of different postures and illuminations.

In addition, for image synthesis, Zhang et al. [34] combined the self-attention block with GAN to handle long-range dependency (SAGAN), thus ensuring that the discriminator can determine the dependency between two distant features. This approach further improves the quality of image synthesis. Based on the SAGAN, Brock et al. [57] proposed a BigGAN to increase the diversity and fidelity of the generated samples by increasing the batch size and using "truncation trick". For the latent distribution z, the traditional approach is to embed z as input into the initial layer of the generator G. But in BigGAN, z is embedded in multiple layers of the generator G to affect features of different resolutions and levels. On ImageNet, the Inception Score (IS) reached to 166.3, while the Frechet Inception Distance (FID) dropped to 9.6. As shown in Fig. 8, the generated samples are realistic, and the approach achieves state-of-the-art effect. For video generation, Tulyakov et al. [58] proposed a MoCoGAN to generate video in an unsupervised manner. For text-to-image translation, [59] and [60] used the textual description to generate images.

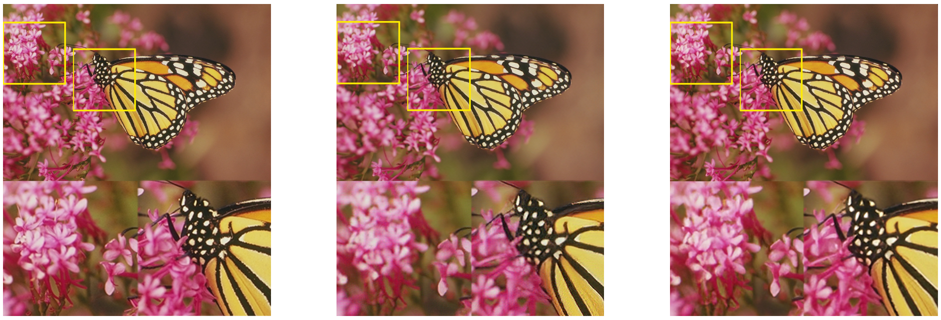
B. NATURAL LANGUAGE PROCESSING

At present, GANs also has some achievements in the field of language and speech processing. Yu et al. [61] proposed a SeqGAN based on the policy gradient to train the generator. Experiments show that the SeqGAN can outperform traditional methods in terms of speech, poetry and music generation. Lin et al. [62] proposed a RankGAN to generate sentences. They used a ranker instead of the discriminator, and achieved excellent performances. Li et al. [63] generated the open-domain dialogue by using the adversarial training method. This task is used as a reinforcement learning [64] problem, joint training generator and discriminator. The result of using the discriminator is used as a reward part of the reinforcement learning to reward the generator, and the dialogue generated by the push generator is similar to the human conversation.

C. OTHER DOMAINS

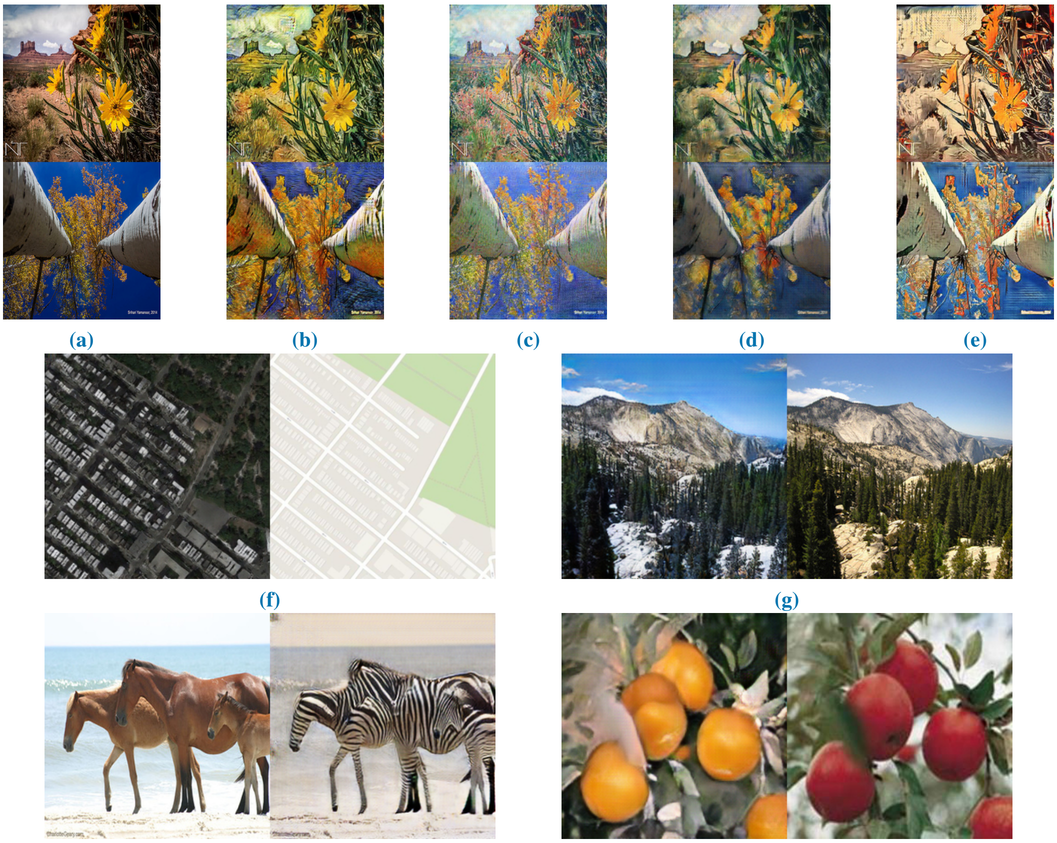
GANs are also used in other domains. In medicine, Schlegl et al. [65] proposed an AnoGAN for anomaly detection of medical images, and learned the characteristics of lesions by learning the characteristics of health data sets. Killoran

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2905015, IEEE Access



(a) (b) (c)

FIGURE 6: Image super-resolution generated by different GANs. (a) Ground truth; (b) SRGAN; (c) ESRGAN.



(a) (b) (c) (d) (e) (f) (g) (h) (i)

FIGURE 7: Image translation generated by CycleGAN. (a) Original; (b) Vangogh; (c) Monet (d) Cezanne; (e) Ukiyoe; (f) Satellite to map; (g) Winter to summer; (h) Horse to zebra; (i) Orange to apple.

et al. [66] generated the DNA sequence by using GANs to optimize protein binding. In the security domain, Hu et al. [67] used GANs to generate malware. Furthermore, for private product customization, Hwang et al. [68] utilized GANs to manufacture medical products.

VII. DISCUSSION

A. THE PROBLEMS OF GANS

The main problem in the development of GANs is that it always gets into mode collapse. The generated samples of GANs are always concentrated into a few or even a single model, which will result in a lack of diversity in the generated samples. Therefore, how to increase the diversity of the generated samples is still a problem that need to be addressed. One solution is to use batches of samples to increase the diversity of assessment. In Section 4, mini-batch discriminator is one of the technologies in this category. In addition, using multiple generators to get multiple models is another solution. Ref. [69] combined samples generated by different models to address mode collapse. Furthermore, optimizing the objective function can overcome the problem as well. WGAN [27] and unrolled GAN [21] are representative of this

FIGURE 8: Image synthesis generated by BigGAN.

solution.

The instability in the training process is also a challenge that researchers must be addressed. GANs needs to reach the Nash equilibrium during the training, but it is proved difficult in [27]. In Sections 3 and 4, [27]–[29], [31]–[33] put forward their own solutions for more stable training. In the future, more solutions should be proposed to make GAN training more stable and converge to Nash equilibrium.

Compared with other generative models, the evaluation problem of GAN is more difficult. Section 5 has given several evaluation metrics widely used at present and suggestions on how to select them. Therefore, this is one of the directions that still need to be addressed in the future.

B. THE FUTURE OF GANS

By improving the architecture of network and algorithms, one hopes to design a more powerful generative model which can generate images, audios, videos and texts that are difficult for humans to distinguish. Especially the use of GANs in the field of text, there still exist areas for continued development in natural language processing (NLP) and information retrieval (IR).

Moreover, the relationship between GANs and reinforcement learning (RL) is also a promising research direction in recent years. For instance, [70] embedded GANs into imitation learning; [61] combined GANs with policy gradient; [71] described the relationship between GANs and actor-critic algorithms, etc.

In the security domain, GANs will also have great use. Aversarial attacks on neural networks are one of the popular directions which need people to research. If there is slight perturbation in the input samples, the neural network will be deceived and make wrong classification and prediction. Currently, there are already some works of attack on Convolutional Neural Network (CNN) [72]–[74], Recurrent Neural Networks (RNN) [75] and Deep Reinforcement Learning (DeepRL) [76], [77] and GANs [78]. At present, the vulnerability of deep learning to subtle adversarial perturbations is a common phenomenon [79]. To against adversarial attacks, [80] and [81] use GANs to make the right defense. In the future, we hope that GANs will be more robust in adversarial attacks. Furthermore, in information hiding, Hayes et al. [82] introduced the adversarial technology into steganography, which provides the novel idea for researchers to address such problems.

In addition to the research direction proposed above, GAN is proposed as unsupervised learning, but in practical applications, adding a certain number of labels can greatly improve its generating ability. It is difficult to obtain a large number of data labels, but a small number of labels can be obtained. Therefore, how to better combine GAN and semi-supervised learning is also one of the future directions.

VIII. CONCLUSIONS

This paper summarizes the research background of GANs, expounds its basic principles, and introduces its derived model and its application in various fields. In addition, the evaluation metrics and training tricks are also discussed. Finally, the existing problems of GANs are summarized, and the future research directions are pointed out.

REFERENCES

[1] K. P. Murphy, “Machine learning: a probabilistic perspective,” Cambridge, MA: MIT Press, 2012.
[2] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798–1828, Aug 2013.
[3] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” nature, vol. 521, no. 7553, pp. 436–444, 2015.
[4] P. Smolensky, “Information processing in dynamical systems: foundations of harmony theory,” Cambridge, MA: MIT Press, 1986.
[5] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” Neural Comput., vol. 18, no. 7, pp. 1527–1554, Jul. 2006. [Online]. Available: http://dx.doi.org/10.1162/neco.2006.18.7.1527
[6] R. Salakhutdinov and G. Hinton, “Deep boltzmann machines,” in Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics, vol. 5, Apr 2009, pp. 448–455. [Online]. Available: http://proceedings.mlr.press/v5/salakhutdinov09a.html
[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in

Advances in Neural Information Processing Systems 27, 2014, pp. 2672–2680. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[8] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, pp. 533–536, 1986.

[9] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in International Conference on Learning Representations, 2014.

[10] H. Akaike, "Fitting autoregressive models for prediction," Annals of the Institute of Statistical Mathematics, vol. 21, no. 1, pp. 243–247, Dec 1969. [Online]. Available: https://doi.org/10.1007/BF02532251

[11] A. Van Den Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," in Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48, 2016, pp. 1747–1756. [Online]. Available: http://dl.acm.org/citation.cfm?id=3045390.3045575

[12] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in International Conference on Learning Representations, 2016. [Online]. Available: https://arxiv.org/abs/1511.06434

[13] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," arXiv e-prints, p. arXiv:1411.1784, Nov. 2014. [Online]. Available: https://arxiv.org/abs/1411.1784

[14] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," in Advances in Neural Information Processing Systems 29, 2016, pp. 2172–2180. [Online]. Available: http://papers.nips.cc/paper/6399-infogan-interpretable-representation-learning-by-information-maximizing-generative-adversarial-nets.pdf

[15] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in Proceedings of the 34th International Conference on Machine Learning, vol. 70. Aug 2017, pp. 2642–2651. [Online]. Available: http://proceedings.mlr.press/v70/odena17a.html

[16] A. Makhzani, J. Shlens, N. Jaitly, and I. Goodfellow, "Adversarial autoencoders," in International Conference on Learning Representations, 2016. [Online]. Available: http://arxiv.org/abs/1511.05644

[17] J. Donahue, P. Kr?henbÍźhl, and T. Darrell, "Adversarial feature learning," in International Conference on Learning Representations, 2017.

[18] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, and A. Courville, "Adversarially learned inference," in International Conference on Learning Representations, 2017. [Online]. Available: https://arxiv.org/abs/1606.00704

[19] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "It takes (only) two: Adversarial generator-encoder networks," in AAAI Conference on Artificial Intelligence, 2018. [Online]. Available: https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16568

[20] A. B. L. Larsen, S. K. S?nderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," in Proceedings of The 33rd International Conference on Machine Learning, Jun 2016, pp. 1558–1566. [Online]. Available: http://proceedings.mlr.press/v48/larsen16.html

[21] L. Metz, B. Poole, D. Pfau, and J. Sohl-Dickstein, "Unrolled generative adversarial networks," in International Conference on Learning Representations, 2017.

[22] S. Nowozin, B. Cseke, and R. Tomioka, "f-gan: Training generative neural samplers using variational divergence minimization," in Advances in Neural Information Processing Systems 29, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. Curran Associates, Inc., 2016, pp. 271–279. [Online]. Available: http://papers.nips.cc/paper/6066-f-gan-training-generative-neural-samplers-using-variational-divergence-minimization.pdf

[23] T. Che, Y. Li, A. P. Jacob, Y. Bengio, and W. Li, "Mode regularized generative adversarial networks," in International Conference on Learning Representations, 2017. [Online]. Available: https://arxiv.org/abs/1612.02136

[24] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "On the effectiveness of least squares generative adversarial networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.

[25] G.-J. Qi, "Loss-Sensitive Generative Adversarial Networks on Lipschitz Densities," arXiv e-prints, arXiv:1701.06264, Jan. 2017. [Online]. Available: https://arxiv.org/abs/1701.06264

[26] J. Zhao, M. Mathieu, and Y. LeCun, "Energy-based generative adversarial network," in International Conference on Learning Representations, 2017.

[27] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in Proceedings of the 34th International Conference on Machine Learning, vol. 70, Aug 2017, pp. 214–223. [Online]. Available: http://proceedings.mlr.press/v70/arjovsky17a.html

[28] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in Advances in Neural Information Processing Systems 30, 2017, pp. 5767–5777. [Online]. Available: http://papers.nips.cc/paper/7159-improved-training-of-wasserstein-gans.pdf

[29] H. Petzka, A. Fischer, and D. Lukovnicov, "On the regularization of wasserstein gans," in International Conference on Learning Representations, 2018. [Online]. Available: https://arxiv.org/abs/1709.08894

[30] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural Computation, vol. 1, no. 4, pp. 541–551, Dec 1989.

[31] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, and X. Chen, "Improved techniques for training gans," in Advances in Neural Information Processing Systems 29, 2016, pp. 2234–2242. [Online]. Available: http://papers.nips.cc/paper/6125-improved-techniques-for-training-gans.pdf

[32] 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," in Advances in Neural Information Processing Systems , 2017, pp. 6626–6637. [Online]. Available: http://papers.nips.cc/paper/7240-gans-trained-by-a-two-time-scale-update-rule-converge-to-a-local-nash-equilibrium.pdf

[33] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in International Conference on Learning Representations, 2018. [Online]. Available: https://openreview.net/forum?id=B1QRgziT-

[34] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-Attention Generative Adversarial Networks," arXiv e-prints, p. arXiv:1805.08318, May 2018. [Online]. Available: https://arxiv.org/abs/1805.08318

[35] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, "Demystifying MMD GANs," in International Conference on Learning Representations, 2018. [Online]. Available: https://openreview.net/forum?id=r1lUOzWCW

[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, April 2004.

[37] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, vol. 2, Nov 2003, pp. 1398–1402.

[38] W. Fedus, M. Rosca, B. Lakshminarayanan, A. M. Dai, S. Mohamed, and I. Goodfellow, "Many paths to equilibrium: Gans do not need to decrease a divergence at every step," in International Conference on Learning Representations, 2018. [Online]. Available: https://arxiv.org/abs/1710.08446

[39] K. Kurach, M. Lucic, X. Zhai, M. Michalski, and S. Gelly, "The gan landscape: Losses, architectures, regularization, and normalization," in International Conference on Machine Learning, 2018. [Online]. Available: https://arxiv.org/abs/1807.04720

[40] E. L. Lehmann and J. P. Romano, "Testing statistical hypotheses," Springer Science and Business Media, 2006.

[41] D. Lopez-Paz and M. Oquab, "Revisiting classifier two-sample tests," in International Conference on Learning Representations, 2017. [Online]. Available: https://arxiv.org/abs/1610.06545

[42] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," Journal of Machine Learning Research, vol. 13, no. 1, pp. 723–773, Mar. 2012. [Online]. Available: http://dl.acm.org/citation.cfm?id=2503308.2188410

[43] A. Borji, "Pros and Cons of GAN Evaluation Measures," arXiv e-prints, p. arXiv:1802.03446, Feb. 2018. [Online]. Available: https://arxiv.org/abs/1802.03446v1

[44] C. Ledig, L. Theis, F. HuszÍćr, 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 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017, pp. 105–114.

[45] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in The European Conference on Computer Vision Workshops (ECCVW), September 2018.

[46] A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard GAN," arXiv e-prints, p. arXiv:1807.00734, Jul. 2018. [Online]. Available: https://arxiv.org/abs/1807.00734?context=stat.ML

[47] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017, pp. 5967–5976.

[48] 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 The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.

[49] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in The IEEE International Conference on Computer Vision (ICCV), Oct 2017.

[50] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in Proceedings of the 34th International Conference on Machine Learning, Aug 2017, pp. 1857–1865. [Online]. Available: http://proceedings.mlr.press/v70/kim17a.html

[51] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in The IEEE International Conference on Computer Vision (ICCV), Oct 2017.

[52] Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018, pp. 8789–8797.

[53] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in European Conference on Computer Vision, 2016, pp. 702–716.

[54] N. Jetchev, U. Bergmann, and R. Vollgraf, "Texture synthesis with spatial generative adversarial networks," in Advances in Neural Information Processing Systems, Conference Proceedings.

[55] U. Bergmann, N. Jetchev, and R. Vollgraf, "Learning texture manifolds with the periodic spatial GAN," in Proceedings of the 34th International Conference on Machine Learning, Aug 2017, pp. 469–477. [Online]. Available: http://proceedings.mlr.press/v70/bergmann17a.html

[56] 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 The IEEE International Conference on Computer Vision (ICCV), Oct 2017.

[57] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in International Conference on Learning Representations, 2019. [Online]. Available: https://openreview.net/forum?id=B1xsqj09Fm

[58] S. Tulyakov, M. Liu, X. Yang, and J. Kautz, "Mocogan: Decomposing motion and content for video generation," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018, pp. 1526–1535.

[59] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in Proceedings of The 33rd International Conference on Machine Learning, 20–22 Jun 2016, pp. 1060–1069. [Online]. Available: http://proceedings.mlr.press/v48/reed16.html

[60] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks," in The IEEE International Conference on Computer Vision (ICCV), Oct 2017.

[61] L. Yu, W. Zhang, J. Wang, and Y. Yu, "Seqgan: Sequence generative adversarial nets with policy gradient." in AAAI Conference on Artificial Intelligence, 2017, pp. 2852–2858.

[62] K. Lin, D. Li, X. He, Z. Zhang, and M.-t. Sun, "Adversarial ranking for language generation," in Advances in Neural Information Processing Systems 30, 2017, pp. 3155–3165. [Online]. Available: http://papers.nips.cc/paper/6908-adversarial-ranking-for-language-generation.pdf

[63] J. Li, W. Monroe, T. Shi, S. Jean, A. Ritter, and D. Jurafsky, "Adversarial learning for neural dialogue generation," in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017, pp. 2157–2169. [Online]. Available: http://aclweb.org/anthology/D17-1230

[64] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," Nature, vol. 518, pp. 529–533, 2015.

[65] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in Information Processing in Medical Imaging, 2017, pp. 146–157.

[66] N. Killoran, L. J. Lee, A. Delong, D. Duvenaud, and B. J. Frey, "Generating and designing dna with deep generative models," in Advances in Neural Information Processing Systems, 2017.

[67] W. Hu and Y. Tan, "Generating Adversarial Malware Examples for Black-Box Attacks Based on GAN," arXiv e-prints, p. arXiv:1702.05983, Feb. 2017. [Online]. Available: https://arxiv.org/abs/1702.05983

[68] J.-J. Hwang, S. Azernikov, A. A. Efros, and S. X. Yu, "Learning Beyond Human Expertise with Generative Models for Dental Restorations," arXiv e-prints, p. arXiv:1804.00064, Mar. 2018. [Online]. Available: https://arxiv.org/abs/1804.00064

[69] A. Ghosh, V. Kulharia, V. Namboodiri, P. H. S. Torr, and P. K. Dokania, "Multi-agent diverse generative adversarial networks," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018, pp. 8513–8521.

[70] J. Ho and S. Ermon, "Generative adversarial imitation learning," in Advances in Neural Information Processing Systems 29, 2016, pp. 4565–4573. [Online]. Available: http://papers.nips.cc/paper/6391-generative-adversarial-imitation-learning.pdf

[71] D. Pfau and O. Vinyals, "Connecting generative adversarial networks and actor-critic methods," in NIPS Workshop on Adversarial Training, 2016.

[72] Y. Liu, X. Chen, C. Liu, and D. Song, "Delving into transferable adversarial examples and black-box attacks," International Conference on Learning Representations, 2017.

[73] S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard, "Universal adversarial perturbations," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017, pp. 86–94.

[74] J. Su, D. V. Vargas, and K. Sakurai, "One pixel attack for fooling deep neural networks," IEEE Transactions on Evolutionary Computation, 2019.

[75] N. Papernot, P. McDaniel, A. Swami, and R. Harang, "Crafting adversarial input sequences for recurrent neural networks," in MILCOM 2016 - 2016 IEEE Military Communications Conference, pp. 49–54, 2016.

[76] S. Huang, N. Papernot, I. Goodfellow, Y. Duan, and P. Abbeel, "Adversarial Attacks on Neural Network Policies," International Conference on Learning Representations, 2017.

[77] Y.-C. Lin, Z.-W. Hong, Y.-H. Liao, M.-L. Shih, M.-Y. Liu, and M. Sun, "Tactics of Adversarial Attack on Deep Reinforcement Learning Agents," International Joint Conference on Artificial Intelligence (IJCAI), 2017.

[78] J. Kos, I. Fischer, and D. Song, "Adversarial examples for generative models," in 2018 IEEE Security and Privacy Workshops (SPW), pp. 36–42, 2018.

[79] 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.

[80] H. Lee, S. Han, and J. Lee, "Generative Adversarial Trainer: Defense to Adversarial Perturbations with GAN," arXiv e-prints, arXiv:1705.03387, 2017. [Online]. Available: https://arxiv.org/abs/1708.05207v3

[81] S. Shen, G. Jin, K. Gao, and Y. Zhang, "APE-GAN: Adversarial Perturbation Elimination with GAN," arXiv e-prints, arXiv:1707.05474, Jul 2017. [Online]. Available: https://arxiv.org/abs/1707.05474v3

[82] J. Hayes and G. Danezis, "Generating steganographic images via adversarial training," in Advances in Neural Information Processing Systems, 2017, pp. 1954–1963. [Online]. Available: http://papers.nips.cc/paper/6791-generating-steganographic-images-via-adversarial-training.pdf

ZHAOQING PAN (M'14) received the B.S. degree in computer science and technology from Yancheng Normal University, Yancheng, China, in 2009, and the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, in 2014. In 2013, he was a Visiting Scholar with the Department of Electrical Engineering, University of Washington, Seattle, WA, USA, for six months. He is currently a Full Professor with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China. He is an Associate Editor of IEEE Access, and an Editorial Board Member of PLOS One. He also was a Lead Guest Editor for Advances in Multimedia, and Security and Communication Networks. His research interest focuses on video coding and machine learning.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2905015, IEEE Access

WEIJIE YU received the B.S. degree in software engineering from Nanjing University of Information Science and Technology, Nanjing, China, in 2018. He is currently working toward the M.S. degree in software engineering at the Nanjing University of Information Science and Technology, Nanjing, China. His research interests focus on video coding, computer vision, and machine learning.

XIAOKAI YI received the B.S. degree in computer science and technology from Nanjing University of Information Science and Technology, Nanjing, China, in 2016. He is currently working toward the M.S. degree in computer science and technology at the Nanjing University of Information Science and Technology, Nanjing, China. His research interests focus on video coding, and machine learning.

ASIFULLAH KHAN received the M.S. and Ph.D. degrees in computer systems engineering from the Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Pakistan, in 2003 and 2006, respectively. He has completed two years of postdoctoral research at the Signal and Image Processing Lab, Department of Mechatronics, Gwangju Institute of Science and Technology, Korea. He has more than 15 years of research experience and is working as an associate professor in the Department of Computer and Information Sciences at Pakistan Institute of Engineering and Applied Sciences, Nilore, Pakistan. His research interests include digital watermarking, pattern recognition, image processing, and evolutionary algorithms.

FENG YUAN received the B.S. degree in computer science and technology from Nanjing University of Information Science and Technology, Nanjing, China, in 2018. He is currently working toward the M.S. degree in computer science and technology at the Nanjing University of Information Science and Technology, Nanjing, China. His research interests focus on video coding, computer vision, and machine learning.

YUHUI ZHENG (M'15) received the B.S. degree in chemistry, and the Ph.D. degree in computer science from the Nanjing University of Science and Technology, Nanjing, China, in 2004 and 2009, respectively. From 2014 to 2015, he was a Visiting Scholar with the Digital Media Laboratory, School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon, South Korea. He is currently a Professor with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing. His current research interests include inverse problems in image processing, scientific computing, data mining, and pattern recognition.