A Review: Generative Adversarial Networks

Liang Gonog[1,2] and Yimin Zhou[1,*]
[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[2]School of Electrical Engineering, University of South China, Hengyang 421000, China
Email: {liang.gong, ym.zhou}@siat.ac.cn

***Abstract*—Deep learning has achieved great success in the field of artificial intelligence, and many deep learning models have been developed. Generative Adversarial Networks (GAN) is one of the deep learning model, which was proposed based on zero-sum game theory and has become a new research hotspot. The significance of the model variation is to obtain the data distribution through unsupervised learning and to generate more realistic/actual data. Currently, GANs have been widely studied due to the enormous application prospect, including image and vision computing, video and language processing, etc. In this paper, the background of the GAN, theoretic models and extensional variants of GANs are introduced, where the variants can further optimize the original GAN or change the basic structures. Then the typical applications of GANs are explained. Finally the existing problems of GANs are summarized and the future work of GANs models are given.**

***Index Terms*—Generative Adversarial Networks (GANs), Deep learning, Generative models, Application**

I. Introduction

The generative adversarial networks (GANs)[1] is an emerging generative model proposed by Ian Goodfellow of Google Brain scientists in 2014. As a new method of learning and generative model, GAN can avoid some deficiency in the practical application of some traditional generation models, and can subtly optimize some loss functions that are hardly to deal with through adversarial learning, and realize the semi-supervised and unsupervised learning technology by implicitly modeling the high dimensional distribution of data. GANs excels in various challenging tasks, such as realistic image generation[2][3], video frame generation[4], artistic style migration[5], etc.

In this paper, we are providing GANs proposal background and development from the algrithm prospective, including basic theory and extensional models of GAN in Section II. Then we introduce the variants of GANs for different application fields in Section III. In Section IV, problems and advantages of the GANs are summarized, and the future prospect of GANs is discussed in Section V.

II. Theory and extensional models of GANs

A. Basic Theory of GANs

GANs are structurally inspired by two-person zero-sum games in the game theory (i.e. the sum of the two people interests is zero, and the gain of one side is exactly what the other side loses). It sets up one generator and one discriminator for each participant in the game. The aim of the generator is to learn and capture the potential distribution in the actual data samples as much as possible, and generate new data samples. Discriminator is a binary classifier, and the aim is to determine whether the input data is from the actual data or from the generator. In order to win the game, the two players need to constantly improve their capability to generate and discriminate. So the process of optimal learning is a minimax game problem[6]. The aim is to find a Nash equilibrium[7] between the two sides, so that the generator can estimate the distribution of data samples. The computation procedure and structure of GANs are shown in Fig.1.

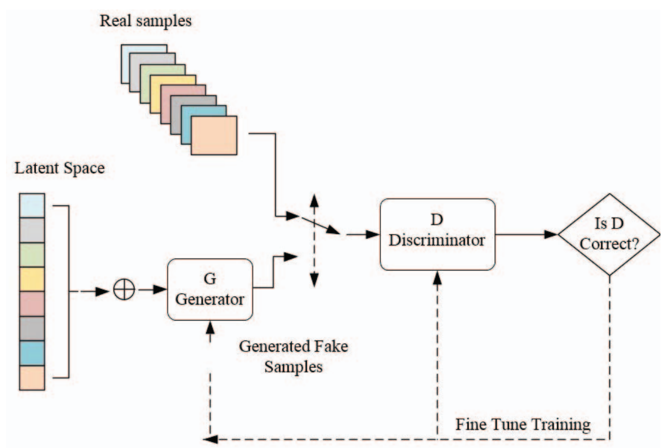

Fig. 1. The Structure of GAN

Any differentiable function can be used to represent the generator and discriminator of GANs, which means that the generator and discriminator can adopt the deep neural network. If the differentiable functions D and G are used to represent discriminators and generators respectively, their inputs are real data x and random variable z respectively. $G(z)$ is a sample generated by G, which obeys the distribution of real data. If the input of the discriminator comes from the real data, it is tagged as One. If the input sample is $G(z)$, it is tagged as $Zero$.The goal of D is to achieve a correct binary classification of data sources: true(from the distribution calssification of real data x) or false(from the fake data $G(z)$ of the generator)[8], while the goal of G is to make the selfgenerated false data $G(z)$ perform as the same as the real data x performing on $D(x)$, which are mutually antagonistic. The performance of D and G can be improved by the process and iteratively optimized[9]. When the discriminative ability of D is improved to a certain

978-1-5386-9490-9/19/$31.00 ©2019 IEEE

degree and the data source can not be correctly identified, the generator G can be considered to have learned the distribution of the actual data.

B. Mathematical Model and Training Method

Here, we introduce the mathematical model and the training method of GANs. First, we describe the optimization of the discriminator D given generator G. Similar to the training of Sigmod functionbased classifiers, training the discriminator involves minimizing the cross entropy[10]. The loss function is formulated as,

$$\begin{aligned} Obj^D(\theta_D,\theta_G) = &-\frac{1}{2}E_{x\sim p_{data}}[logD(x)] \\ &-\frac{1}{2}E_{z\sim p_z(z)}[log(1-D(g(z)))] \end{aligned} \tag{1}$$

where x is sampled from the real data distribution $p_{data}(x)$, z is sampled from the prior distribution $p_z(z)$ such as uniform or Gaussian distribution, and $E(\cdot)$ represents the expectation. It should be noted that the training data consists of two parts: one part from the real data distribution $p_{data}(x)$ and another part from the generated data distribution $p_g(x)$. This is slightly different from the conventional methods for binary classification. Given the generator, we need to minimize (1) to obtain the optimal solution[11]. In the continuous time space, Eq.(1) can be reformulated as,

$$\begin{aligned} O_{b_{jD}}(\theta_D,\theta_G) = &-\frac{1}{2}\int_x P_{data}(x)log(D(x))d_x \\ &-\frac{1}{2}\int_z P_z(z)log(1-D(g(z)))d_z \\ = &-\frac{1}{2}\int_x [P_{data}(x)log(D(x))]d_x \\ &+\frac{1}{2}\int_z [P_g(z)log(1-D(z))]d_z \end{aligned} \tag{2}$$

For any $(m,n)\epsilon R^2$, and $y\epsilon[0,1]$, the expression,

$$-mlog(y)-nlog(1-y) \tag{3}$$

can achieve its minimum value at $y = m/(m+n)$. When giving generator G, the objective function $Eq.(2)$ can achieve its minimum value, $D^*_G(x)$, as,

$$D^*_G(x) = \frac{p_{data}(x)}{p_{data}(x)+p_g(x)} \tag{4}$$

This is the optimal solution of the discriminator D. From Eq(4), the discriminator of the GAN can estimate the ratio of the two probability densities, which is the key difference from Markov chain or lower bound based methods. On the other hand, $D(x)$ denotes the probability of x sampled from the real data rather than the generated data. If the input data is from the real data x, the discriminator strives to make $D(x)$ close to 1. If the input data is from the generated data $G(z)$, the discriminator strives to make $D(G(z))$ close to 0 while the generator G tries to make it close to 1. Since this is a zero-sum game between G and D, the optimization of GAN can be formulated as a min-imax problem:

$$\begin{aligned} min_G max_D f(D,G) = nE_{x\sim P_{data}}(logD(x)) \\ +E_{z\sim P_z}(log(1-D(g(z)))) \end{aligned} \tag{5}$$

In summary, for learning the parameters of the GAN, we need to train the discriminator D to maximize the accuracy of the input data from the real data x or the generated data $G(z)$. In addition, we should train the generator G to minimize $log(1-D(G(z)))$. Here, an alternative training method can be used. First, G is fixed to maximize the discrimination accuracy of D. Then, D is fixed to minimize the discrimination accuracy of D[14]. This process alternates and we could achieve the global optimal solution if and only if $p_{data} = p_g$. In the training process, the parameters of D are updated k times and the parameters of G is only updated once.

C. Extentional Models of GANs

Compared to other generative models, higher quality samples (sharper and clearer images) can be produced by GAN than other models. Shorter runtime will consumed by GAN to generate a sample compared to PixelRNN[12]. GAN can approximate arbitrary probability distributions in a theoretical mode to overcome the problem that VAE's[13] (Variational Auto Encoders) final simulation results would be biased. However, GAN still has the problem of divergence for the model is too free and uncontrollable. Therefore, some improvements have been made to improve the performance of GANs.

1) Optimization of GAN: We have introduced the basic structure of the original GAN, whose generator uses a distribution to directly sample without the requirement for pre-modeling, thus achieving theoretically complete fitting of the actual data distribution, which is the biggest advantage of GAN. Many GANs models can realize the optimization of the GAN without changing the original structure.

DCGAN(deep convolutional GAN)[14] is one of the extension of the GAN. It replaces the G and D in the original GAN with two CNN(convoluitnal neural network) without changing the basic structure of the GAN, a step size convolution is used instead of the upsampling layer, and a convolutional layer is used to replace the full connection layer to increase the stability of the training.

Unlike DCGAN, the permutation on the fully connected layer can be improved via WGAN (Wasserstein GAN) [15]. The Jensen-Shannon divergence is not suitable for measuring the distance of the distribution of the disjoint parts, and Wasserstein distance is used instead to measure the distance between the generated data distribution and the real data distritution. So the problem of instability in training and model collapse can be solved partially. In fact, Lipschitz restriction in WGAN requires to cut off the absolute value of the discriminator parameters without exceeding the fixed constant c. Hence, Gulrajani proposed WGAN-GP (WGAN-gradient penalty)[16] with the gradient penalty replacement so that the discriminator can learn reasonable parameters to solve

TABLE I
THE CLASSIFICATION OF THE GANS

Category	GANs Networks
Optimazation of GAN	DCGAN[14] WGAN[15] LSGAN[17] EBGAAN[19]
Different Structure from GAN	CGAN[21] Semi-supervised-GAN[22] InfoGAN[23] BigGAN[2] CycleGAN[24]

the slow convergence problem of the WGAN.

Although WGAN and WGAN-GP have basically solved the problem of the training failure, both the training process and the convergence speed are slower than that of the conventional GAN. Inspired by WGAN, Mao et al. proposed the least square GAN (LSGAN)[17], one of the starting points of LSGAN is to improve the quality of the picture. Its main idea is to provide discriminator D with a loss function for smooth and unsaturated gradients. There is another LSGAN (Loss-sensitive GAN) proposed by Qi[18], the loss function obtained by the minimum objective function is limited to satisfy the Lipschitz continuity function class, in order to limit the modeling ability of the model and solve the over-fitting problem. Similar to WGAN, EBGAN (Energy-Based GAN)[19] uses energy values as a measure from the energy model; BEGAN (Boundary Equilibrium GAN)[20] also proposes a "boundary balance" architecture based on EBGAN and WGAN, using standard training steps to achieve fast and stable convergence.

2) Different Structure from the Original GAN: The above GANs are the improvement on the GAN foundation. However, the underlying GAN is sometimes insufficient in practice to meet our requirements for data generation. For example, sometimes it is necessary to generate a certain type of images instead of randomly simulating sample data, such as generating a certain text; sometimes it is required some parts of the image to be generated instead of generating all images, such as mosaic. Based on these real life requirements, GAN also requires to adjust the structure of the original model to meet the data to be generated.

In the application, the vast majority of the data is the multi-labelled data, and how to generate the data of the specified label is the contribution of the conditional GAN(cGAN)[21] on the GAN model. In the basic GAN model, the generator is implemented by inputting a string of random numbers that satisfy a certain distribution. In CGAN, not only the random number but also the label category is spliced and input, so the generator can generates the required data. In addition, for the discriminator, it is also necessary to splice the real data or the generated data with the corresponding tag category, and then inputting the neural network of the discriminator for identification and judgment.

Since cGAN has been proposed, many scholars have used cGAN in the application or improvement for the follow-up work of cGAN. For example, Laplacian Generative Adversarial Networks (LAPGAN)[25] combines the principles of GAN and cGAN using a tandem network, and the pictures generated by the above level are used as condition variables to form a laplacian pyramid to generate images from rough to precise mode.

InfoGAN[23] can also be seen as a kind of cGAN. From the starting point, InfoGAN is based on the simplicity of GAN. It can decompose the input z on the original singer, and decomposes an implicit code c from the original noise z. In addition to c, it contain a variety of variables. Taking the MNIST dataset as an example, it is possible to indicate such the direction of the illumination, the tilt angle of the font, the thickness of the stroke, and so on. The basic idea of the InfoGAN is that the c can explain the generated $G(z, c)$, then c should be highly correlated with $G(z, c)$. The important significance of InfoGAN is that it breaks out the structured implicit code c from the noise z, which makes the generation process have certain a degree of controllability, and the generated result also has certain interpretability.

Pix2Pix[26] based on CGAN treats the generator as a kind of mapping, which maps the image to another desired image pixels to pixels, so as to realize the different image translation function through the same model. This inspires the researchers to explore further. However, the fatal flaw of Pix2Pix is that its training requires pictures x and y that are paired with each other. This type of data is extremely lacking and greatly limits the application of the model. In this regard, CycleGAN[24] proposes an image translation method that does not require pairing data.

As a promotion of CycleGAN, StarGAN[27] has turned the two-two mapping into a mapping between multiple domains, which is another major breakthrough in the field of image translation. In addition, StarGAN can train the same model by implementing joint training between multiple datasets (such as CelebA datasets with labels such as skin color, age, and RaFD datasets with angry, scary, and other expression tags). Completing the compression of the model is a major success in the field of image translation.

At the same time, we can classify GANs as unsupervised learning, semi-supervised learning and supervised learning through the presence or absence of the category label information in the training data. Table II demonstrates the different attribute category of some GANs.

TABLE II
CLASSFICATION OF THE GANS BASE ON THE LEARNING METHOD

Category	Characteristics	Networks
Unsupervised Learning	Tagged data	InfoGAN, CycleGAN, GAN
Semi-supervised Learning	Partially tagged data	Improved-GAN[28], SGAN
Supervised Learning	Untagged data	CGAN, ACGAN, AM-GAN

III. APPLICATION

The most important power of GANs is the networks that can generate samples with the same distribution as the real data, such as generating photorealistic images. GANs can also be used to tackle the problem of insufficient training samples for supervised or semi-supervised learning. Currently, successful applications of GAN are computer vision, including images and video, such as image translation, image coloring, image restoration, and video generation. In addition, GANs have been applied in speech and language processing, such as generating dialogues. In this section, we discuss the application range of GANs.

A. The Application in the Image

1) image-to-image Translation: CycleGAN is an important application model of GAN in the field of image. CycleGAN is based on two types of images that requires no pairing. You can turn it into a smile by typing a crying face or change the zebra to horse as shown in Fig.2. StarGAN is a further extension of CycleGAN, where it takes strength to train one category to the other category. The smile face can be transformed into a crying face via the StarGAN, along with variety of the expressions such as surprise, frustration, etc.

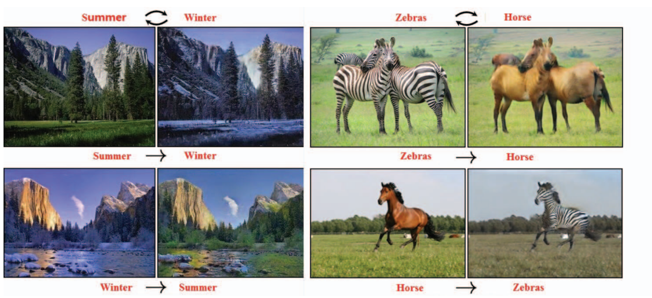

Fig. 2. Imge to image translation result of CycleGAN

2) Image Super-Resolution Reconstruction: The use of GAN for super-resolution is to solve the shortcomings of the conventional methods including the deep learning methods, which lacks of high-frequency information. The traditional deep CNN can only improve this defect by selecting the objective function. On the other hand, GAN can also solve this problem and obtain satisfying perception. SRGAN[29] uses perceptual loss and adversarial loss to enhance the realism of the recovered picture and realize the 4x resolution reconstruction. Perceptual loss is a feature extracted by convolutional neural network. By comparing the features of the generated image after convolutional neural network and the characteristics of the target image after convolutional neural network, the generated image and the target image are more similar in semantics and style.

3) Style Transformation: Based on the characteristics of the GAN, autonomous learning and random samples generation, combined with the method of adding conditional variables to generate specific samples, a number of GANs have been developed to improve to the images learning through unpaired training data under unsupervised conditions. MC-GAN[30] was proposed by BAIR with a multi-content architecture to re-customize the training network for each observed character set, rather than decorating a single training network for all possible fonts. The multi-content GAN model includes a stacked cGAN architecture for predicting rough glyph shapes and an ornamental network for predicting the final glyph colors and textures. The first network, called GlyphNet, is used to predict the glyph mask; the second, called OrnaNet, is used to fine-tune the colors and decorations that generate glyphs from the first network. Each sub-network follows the structure of cGAN and modifies the structure to achieve the specific purpose of stylizing glyphs or decorating predictions.

4) Image Generation: The Composite GAN[31] proposed by Hanock et al. has generated partial images through multiple generators and finally synthesized the entire image. For example, in face generation, the background and the face are divided into two parts and generated by different generators into the RNN[32], can produce more realistic images. TP-GAN[33], proposed by the CASIA, synthesized the frontal face image through the side face photos, to achieve the state-of-art results. The generative network of the model is characterized by dual path generation, in which one path is generated by partial facial features extracted from the side face. The facial features can be completed, while the other path can produce a vague full face, to locate patch networks to deal with facial features so as to finally form a complete realistic face image. BigGAN[2] applies orthogonal regularization to the generator to obey a simple truncation, allowing fine-tuning of sample fidelity and diversity tradeoffs by truncating hidden spaces. This modification allows the model to achieve the best performance in the image synthesis of class conditions. When training on ImageNet with 128x128 resolution, the model of BigGAN can achieve an Inception score (IS) of 166.3 and a Frechet Inception Distance (FID) of 9.6, while the previous best IS and FID were only 52.52 and 18.65.

B. The Application with the Video

1) Video Frame Prediction: Mathieu et al.[34] first applied GAN training to video prediction, that is, the generator can generate the last frame of the video based on the previous series of the frames, and the discriminator is used to conclude the frame. All the frames except the last frame are real pictures. The advantage is that the discriminator can effectively use the information of the time dimension, and also assists to make the generated frame consistent with all the previous frames. Experimental results show that the frames generated by confrontation training are more clear than the other algorithms.

2) Video Generation: Vondrick et al.[4] have made great progress in the video field, 32-frame resolution 64×64 realistic video can be generated, depicting golf courses, beaches, train stations and newborns. After testing, 20 of the markers did not recognize the authenticity of these videos. MD-GAN(Multi-stage Dynamic GAN)[35] predicts future video frames with the proposed model through a given first frame image. In its two-stage model, the first stage a time-lapse video with realistic content can be generated; the second stage the results of the first stage is optimized, mainly in adding dynamic motion information to increase the vraisemblance. In order to have vivid motion information for the resulting video, a Gram matrix is introduced to describe the motion information more accurately. Moreover a large-scale time-lapse photography video dataset was built and tested on this dataset. By using this model, realistic time-lapse photographic video with a resolution of 128×128 up to 32 frames can be generated. Both qualitative and quantitative experiments demonstrate the superiority of the method outperforming available models.

C. The Application of Human-Computer Interaction

1) Auxiliary Automatic Driving: Santana et al.[36] implemented the assisted autonomous driving with GAN. First, an image is generated, that is consistent with the distribution of the official traffic scene image, and then a transition model is trained based on the cyclic neural network to predict the next traffic scene.

2) Text to Image: This field is the result of a collision between NLP (Natural Language Processing) and CV (Computer Vision). The task is described as: to generate a picture that matches the image text from a given textual description. StarkGAN[37] generates high resolution images from text descriptions. The model decomposes the generation process into two more controllable steps: to draw the basic shape and color of the object; to correct the shortcomings of the first stage results and add more detail. Extensive experiments are performed to prove that the method is more effective. By introducing an attentional generative network, AttnGAN[38] can synthesize fine-grained details of different sub-regions of an image by focusing on the related words in natural language description. In addition, a deep attentional multimodal similarity model is proposed to calculate fine-grained image-text matching loss for generator training. It is the first time that the layered attentional GAN can automatically select word-level conditions to generate different parts of the image.

In addition, some researchers expect to use GAN learning methods in the fields of pharmaceutical molecules and materials science to generate pharmaceutical molecular structures and synthetic new material formulations. The idea is quite creative, if it can be realized in reality, and the Artifical Intelligence will be omnipotent.

IV. Discussion

Generative adversarial networks are quite important in the generative models. GANs have great power to deal with the problem of generating data which can be interpreted naturally. In this section, we will analyze the main avantages, disadvantages and prospect of the generative adversarial networks.

A. Advantages

Compared with other generative models, the generative adversarial networks has the following three advantages. 1) It can produce better samples than other models (the image is sharper and clearer). GANs can train any kind of the generative network. Most other frameworks require the generative network to have some specific functional form. It is important that all other frameworks require generative networks to be distributed over non-zero mass. 2) Generative adversarial networks can learn to generate points only on thin manifolds that are close to the data.The training does not rely on the inefficient Markov chain method, nor the approximate inference. 3) There is no complex variational lower bound, which can greatly reduce the training difficulty and improve the training efficiency.

B. Disadvantages

Although GANs can perform well in Nash equilibrium, the gradient descent can guarantee Nash equilibrium only in the case of convex function. The training process requires to ensure balance and synchronization of the two adversarial networks, otherwise it can not achieve ideal performance. However, it is difficult to control the synchronization of the two adversarial networks, so the training process may be unstable. In addition, the GAN model is defined as a minimax problem with no loss function. It is difficult to distinguish whether the progress is being made in the training process. GANs learning process may cause collapse problem, i.e., the generator degeneration, continuous the same sample points generation, unable continue learning. When the generative model collapses, the discriminative model also points to the same direction for similar sample points, and the training can not continue. Furthermore, although the samples generated by GANs are diverse, there exists the collapse mode problem. Mode collapse refers to the scenarios in which the generator makes multiple images that contain the same color or texture themes, thereby having little difference for human understanding.

C. Prospect

In spite of various disadvantages and limitations, it is worth investigating the future applications and progress of the GANs. For example, Wasserstein GAN (WGAN) have two great improvements compared with the initial generative adversarial network. The first improvement is WGAN which can greatly overcome the training instability problem, and the second improvement is that it can partially solve the collapse mode problem at the same time. How to completely avoid collapse mode and further optimize the training process remains a research direction of the GANs. Furthermore, the theory about model convergence and the existence of equilibrium point remain important research directionnin the future. How to generate a variety of data that can interact with humans from simple random inputs is an important research direction. From the perspective of combining GANs and other methods, how to integrate GANs with feature learning, imitation learning, and reinforcement learning to develop new AI applications and promote the development of these methods is quite meaningful. In the future, GANs could be used to accelerate the development and application of AI, make the AI have the ability to understand our human beings and explore the world.

V. Conclution

As an unsupervised learning method, GANs is one of the most important research directions in deep learning. The explosion of interest in GANs is driven not only by their potential to learn deep, highly nonlinear mappings from a latent space into a data space and also it has potential to make use of the vast quantities of unlabeled data. Although our world is almost overwhelmed by the data, a large part are unlabeled, which means that the data is not available for most current supervised learning. Generative adversarial networks, which rely on the internal confrontation between real data and models to achieve unsupervised learning, is just a glimmer of light for AIs self-learning ability. Therefore, there are many opportunities for the developments in both theory and algorithms, and by using deep networks, there are vast opportunities for new applications.

Acknowledgment

This work is supported under the Shenzhen Science and Technology Innovation Commission Project Grant Ref. JCYJ20160510154736343 and Ref.JCYJJCYJ201708181536 35759, and Science and Technology Planning Project of Guangdong Province Ref. 2017B010117009, and Guangdong Provincial Engineering Technology Research Center of Intelligent Unmanned System and Autonomous Environmental Perception.

References

[1] 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*, 2014, pp. 2672–2680.

[2] A. Brock, J. Donahue, and K. Simonyan, "Large scale gan training for high fidelity natural image synthesis," *arXiv preprint arXiv:1809.11096*, 2018.

[3] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017.

[4] C. Vondrick, H. Pirsiavash, and A. Torralba, "Generating videos with scene dynamics," in *Advances In Neural Information Processing Systems*, 2016, pp. 613–621.

[5] H. Chang, J. Lu, F. Yu, and A. Finkelstein, "Pairedcyclegan: Asymmetric style transfer for applying and removing makeup," in *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[6] L. Li, Y.-L. Lin, D.-P. Cao, N.-N. Zheng, and F. Wang, "Parallel learning-a new framework for machine learning," *Acta Automatica Sinica*, vol. 43, no. 1, pp. 1–8, 2017.

[7] J. Nash, "Two person cooperative games," *Econometrica*, vol. 21, no. 1, pp. 128–140, 1953.

[8] F.-Y. Wang, J. Zhang, Q. Wei, X. Zheng, and L. Li, "Pdp: parallel dynamic programming," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 1, pp. 1–5, 2017.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[10] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal processing magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[11] Y. Zhang, Z. Gan, and L. Carin, "Generating text via adversarial training," in *NIPS workshop on Adversarial Training*, vol. 21, 2016.

[12] A. v. d. Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," *arXiv preprint arXiv:1601.06759*, 2016.

[13] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[14] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *Computer Science*, 2015.

[15] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.

[16] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Advances in Neural Information Processing Systems*, 2017, pp. 5767–5777.

[17] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," 2016.

[18] G. J. Qi, "Loss-sensitive generative adversarial networks on lipschitz densities," 2017.

[19] J. Zhao, M. Mathieu, and Y. Lecun, "Energy-based generative adversarial network," 2017.

[20] D. Berthelot, T. Schumm, and L. Metz, "Began: Boundary equilibrium generative adversarial networks," 2017.

[21] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *Computer Science*, pp. 2672–2680, 2014.

[22] T. Chavdarova and F. Fleuret, "Sgan: An alternative training of generative adversarial networks," in *Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition*, 2018.

[23] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," 2016.

[24] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," pp. 2242–2251, 2017.

[25] R. Fergus, R. Fergus, R. Fergus, and R. Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks," in *International Conference on Neural Information Processing Systems*, 2015.

[26] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," pp. 5967–5976, 2016.

[27] Y. Choi, M. Choi, M. Kim, J. W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," 2017.

[28] A. Odena, "Semi-supervised learning with generative adversarial networks," *arXiv preprint arXiv:1606.01583*, 2016.

[29] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network." in *CVPR*, vol. 2, no. 3, 2017, p. 4.

[30] H. Park, Y. Yoo, and N. Kwak, "Mc-gan: Multi-conditional generative adversarial network for image synthesis," 2018.

[31] H. Kwak and B. T. Zhang, "Generating images part by part with composite generative adversarial networks," 2016.

[32] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," *Computer Science*, 2015.

[33] 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," pp. 2458–2467, 2017.

[34] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," *arXiv preprint arXiv:1511.05440*, 2015.

[35] W. Xiong, W. Luo, L. Ma, W. Liu, and J. Luo, "Learning to generate time-lapse videos using multi-stage dynamic generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2364–2373.

[36] E. Santana and G. Hotz, "Learning a driving simulator," *arXiv preprint arXiv:1608.01230*, 2016.

[37] H. Zhang, T. Xu, H. Li, S. Zhang, X. Huang, X. Wang, and D. Metaxas, "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks," *arXiv preprint*, 2017.

[38] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He, "Attngan: Fine-grained text to image generation with attentional generative adversarial networks," *arXiv preprint*, 2017.