Recent Advances of Generative Adversarial Networks

Zhelin Liu[1†]
[1]*Fuzhou No.8 High School Sino-US Program*
Fuzhou, Fujian, China
zhelin.liu@fz.lyndoninstitute.org.cn

Tingxu Yuan[2†]
[2]*American Heritage School, Boca Delray*
Delray Beach, FL, US
yuantingxu0424@sina.com

Yaxin Lin[3*†]
[3]*DuPont Manual High School*
Louisville, KT, US
[*]yaxin.lin@stu.jefferson.kyschools.us

Botao Zeng[4†]
[4]*College of Engineering*
Texas A&M University
College Station, TX, US
botter.tsang@gmail.com

[†] These authors contributed equally.

***Abstract*—Generative adversarial networks (GANs) have received great attention recently. GAN is an unsupervised learning method that learns representations without highly relying on annotations. GAN comprises two networks called generator and discriminator that training in a competitive process. GANs have made notable progress and promising performance in various applications including image inpainting, image super-resolution, and face synthesis. This paper aims to introduce an overview of GAN, including fundamentals, several up-to-date variants, applications, and challenges of GAN. Finally, we also provide readers with some solutions to mitigate issues existing in GANs.**

Keywords—deep learning, generative adversarial networks, unsupervised learning

I. INTRODUCTION

Machine learning (ML) algorithms have been adopted in a wide range of applications, including image recognition, drug discovery, email filtering, self-driving cars, and natural speech processing. Artificial neuron network (ANN) [1] is a famous algorithm of machine learning and there's another algorithm called generative adversarial networks (GANs) [2] based on it.

There are two modules in GANs: a discriminator and a generator. The generator tries to learn the distribution of the target dataset for new data samples generation. The discriminator is often a classification network, identifying samples generated by the generator from true samples as precisely as possible. In order words, these two networks compete with each other. Training two networks in competition with each other can be defined as an adversarial loss function that influences both generator and discriminator during the learning process. The adversarial loss function pushes the generator to learn the distribution of original data distribution and generate images that look like real data. GAN is widely used for generating human face images [3], image inpainting [4], face aging [5], image super-resolution [6], anime character generation [7], style transfer [8], and medical image analysis [9].

In this paper, we will present an overview of GAN. The rest of the article is organized as follows. Section 2 provides the fundamentals of GAN, Section 3 describes five up-to-date variants of GAN. Section 4 provides some application examples of GAN. Section 5 discusses some challenges of GAN. Finally, we conclude the paper in Section 6.

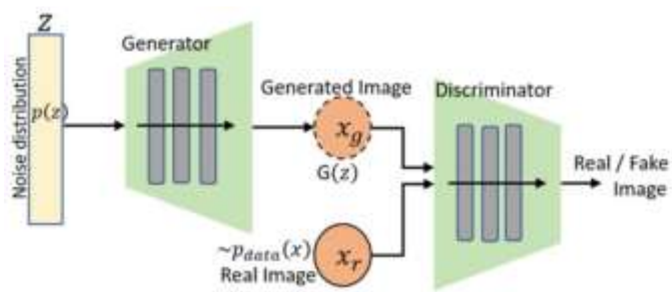

Figure 1. Typical structure of GAN [10].

II. THE FUNDAMENTALS OF GAN

Generative adversarial network (GAN) is a class of deep learning first designed by Ian Goodfellow et al. in 2014. GAN is mainly inspired by the zero-sum game theory. GAN has become one of the most effective approaches for image generation.

A. Generator & Dsicriminator

GAN is composed of two networks—a generator network and a discriminator network, as shown in Figure 1. The generator is a generative network that takes an input random noise *z* and outputs an image. In some applications such as image super-resolution, the generator receives low-resolution images and outputs high-resolution images. The discriminator is a discriminant network to identify whether an image is real or fake. The discriminator's input consists of two parts: images generated from a generator (also called fake images) and images sampled from the original training data set (also called real images). During training, the goal of the generator G is to generate real images to fool the discriminator. The goal of discriminator D is to try to distinguish the fake data generated by G from the true samples. Therefore, the generator and discriminator together form a dynamic game process that competes with each other, and ideally, the training process will end up in the Nash equilibrium point.

B. *Objective Function*

The objective function that GAN tries to optimize is

$$E_x\left[\log\left(D(x)\right)\right] + E_z\left[\log\left(1 - D\left(G(z)\right)\right)\right] \tag{1}$$

where x is the real sample from the training data set, z is the input random noise, $D(x)$ is the classification result for the real sample, E_x is the expectation across all real data samples, $G(z)$ is generated sample given the random noise z. $D\left(G(z)\right)$ probability estimated by the discriminator D, E_z is the expected expectation across all input random noises. In order words, the formula is defined as a cross-entropy loss between the generated and real sample distributions. Note that the generator is unable to influence the $\log\left(D(x)\right)$ term in equation (1). Therefore, minimizing equation (1) for the generator is equivalent to minimize equation as follows

$$\log\left(1 - D\left(G(z)\right)\right) \tag{2}$$

Overall, the training goal for the generator is to minimize the equation (1) while for the discriminator is to maximize it.

C. *Training GAN*

The generator and the discriminator are training in an alternating fashion as:

1) Train the discriminator in one or more epochs.

2) Train the generator in one or more epochs.

3) Repeat steps 1) and 2) until the model convergence.

The parameters of the discriminator are treated as constant during training the generator and vice versa. The training goal of the discriminator is to figure out how to recognize real data from fake, which encourages the generator to improve the realism of the generated data.

III. EXTENSIONS OF GAN

In this section, we introduce five popular extensions of GAN as follows.

A. *Conditional Generative Adversarial Networks*

Conditional generative adversarial network (CGAN) [11] is a GAN training with conditions, which can be built by adding user-defined conditions y such as class labels into both the discriminator and generator. The network architecture of GAN is shown in Figure 2. CGAN is an extension of GAN with extra information y as input. Concretely, y can be annotations or text descriptions for the input images. The input of the generator in CGAN is a joint hidden representation that consists of random noise $p_z(z)$ and auxiliary information y. Similarly, the input for the discriminator is also a combination of x and y. The loss fucntion for training CGAN can be defined as follows

$$\begin{aligned} L_{CGAN} = {} & E_{x\sim p_{data}(x)}\left[\log\left(D(x|y)\right)\right] \\ & + E_{z\sim p_z(x)}[\log\left(1 - D(G(z|y))\right)] \end{aligned} \tag{3}$$

where $D(x|y)$ is estimation probability that an input x is from real conditioning by auxiliary information y. $G(z|y)$ is the generated data using input random noise z conditioning by auxiliary information y. Compared to vanilla GAN, the advantage of CGAN is that it can generate data with user-defined properties.

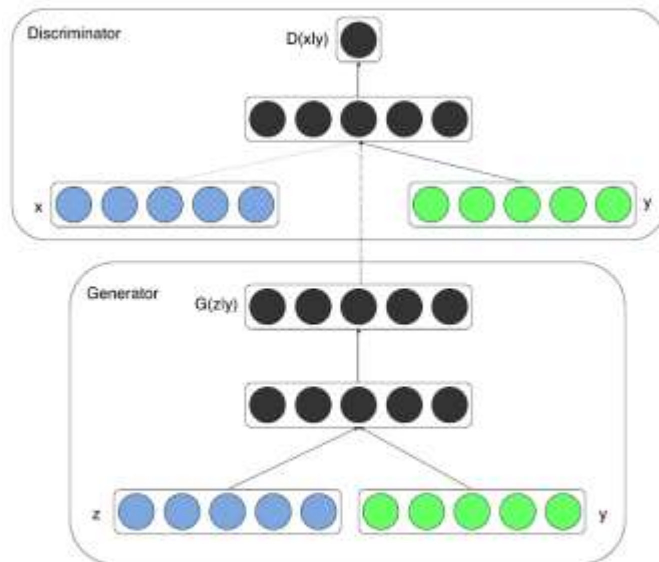

Figure 2. Conditional adversarial net [11].

B. *Deep Convolutional Generative Adversarial Networks*

Deep convolutional generative adversarial network (DCGAN) was the first structure that adopted convolutional networks into GAN, as shown in Figure 3. While the convolutional networks used in DCGAN have several differences compared with standard convolutional networks:

1) DCGAN replaces the pool layers with stride convolutions, in order to reduce information loss caused by polling functions.

2) Add batch norm into GAN after each of the convolutional layers but remove batch norm after the output layer of the generator and the input layer of the discriminator.

3) Replace ReLU with Leaky-ReLU for the discriminator.

These modifications have been proved helpful to stabilize GAN training.

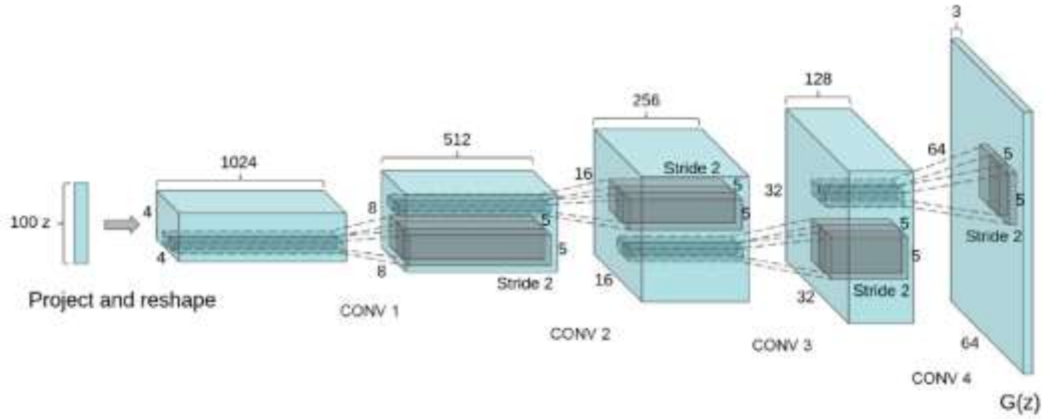

Figure 3. DCGAN generator.

C. Wasserstein Generative Adversarial Networks

The Wasserstein generative adversarial network (WGAN) [12] is a variant of GAN that modifies the training procedure in order to solve the training instability problems. Overall, WGAN can increase the stability of model training, alleviate the mode collapse problem, and provide a strategy that can gauge the discrepancy of distributions between generated samples and real samples.

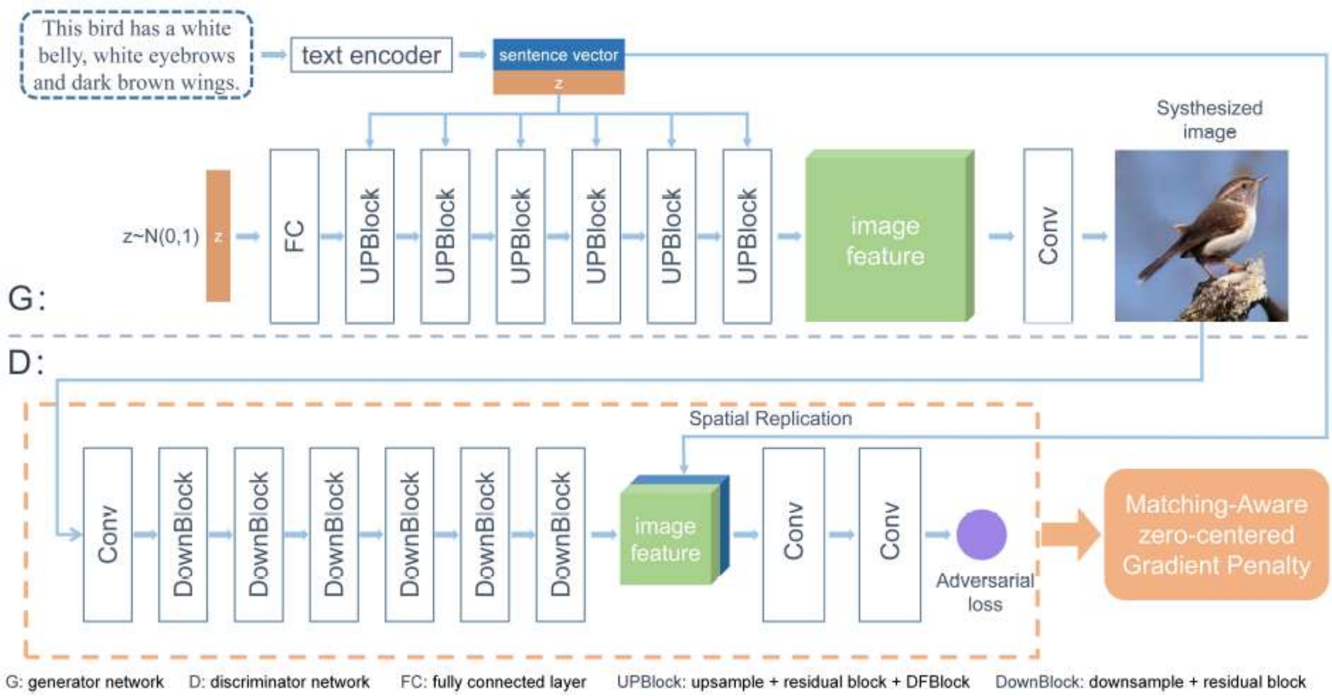

Figure 4. The network architecture of DF-GAN [13].

D. Deep Fusion Generative Adversarial Networks

Deep fusion generative adversarial networks (DF-GAN) [13] is an extension of GAN used to synthesize realistic images from given text descriptions. DF-GAN contains a pre-trained text encoder, a generator, and a discriminator, as shown in Figure 4. DF-GAN develop three novel modules to reduce the computational cost and improve computational efficiency. First, DF-GAN proposed a novel one-stage backbone that only relies on one pair of discriminators and generators for text-to-image translation. The one-stage method can be expressed as follows:

$$\begin{aligned} L_D = & -E_{x\sim P_r}\left[\min\left(0,-1+D(x,e)\right)\right] \\ & -\frac{1}{2}E_{G(z)\sim P_g}\left[\min\left(0,-1 - D(G(z),e)\right)\right] \\ & -\frac{1}{2}E_{x\sim P_{mis}}\left[\min\left(0,-1 - D(x,e)\right)\right] \end{aligned} \tag{4}$$

$$L_G = -E_{G(z)\sim P_g}[D(G(z),e)] \tag{5}$$

where z is the random noise sampled from the Gaussian distribution. Second, a new fusion module is proposed to strengthen the text-image fusion process of the generator. Third, DF-GAN uses a target-aware discriminator that contains gradient penalty and one-way output to boost the generator to synthesize more realistic images without the need of using extra networks. The experimental results show that DF-GAN can not only outperformance other methods but also improve the efficiency of image synthesis.

E. Classification Enhancement Generative Adversarial Networks

Classification enhancement generative adversarial network (CEGAN) [14] aims to generate minority class data for data augmentation. CEGAN is trained with both real and generated data for preventing generated minority data from overfitting to majority. Specifically, two techniques including classification enhancement and ambiguity reduction are introduced to improve the quality of synthetic data while subsequently increase the performance of classification in data imbalanced conditions for classification tasks. CEGAN is composed of three subnetworks: a classifier, a discriminator, and a generator. Training CEGAN includes two steps as shown in Figures 5 and 6. Step 1 is to train the GAN with a classification enhancement strategy as shown in Figure 5 while step 2 is to train a classifier with augmented data as shown in Figure 6. Moreover, to address the problem that traditional CGAN ignores ambiguous relationships between the data, CEGAN develops an ambiguity reduction method consisting of three steps. Firstly, the feature embeddings of the dataset are reduced to two-dimension via principal component analysis (PCA) followed by t-Distributed Stochastic Neighbor Embedding (t-SNE). Then, the two-dimensional features are cluster into the number of the classes using Kmeans++ [15]. Finally, the class ambiguity can be calculated by counting misclassified two-dimensional features. CEGAN used five datasets to evaluate different models. The experimental result shows that CEGAN can approximate real data distribution and boost the classification performance significantly, especially in class imbalance situations, compared with previous data augmentation approaches.

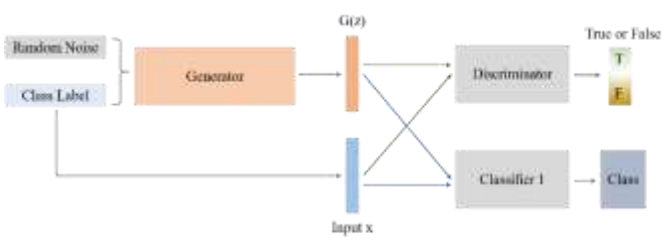

Figure 5. Step 1 for training CEGAN.

Figure 6. Step 2 for training CEGAN.

IV. Application of GAN

GAN has been widely adopted in many image processing fields such as image inpainting, image super-resolution, and person re-identification.

A. Image Inpainting

Image inpainting is a technique that can be harnessed to eliminate useless objects from images or to repair old photos. Recently GAN-based approaches have shown outstanding results for image inpainting. For instance, Yu et al. [4] proposed a GAN-based method that references surrounding image features and is able to produce novel image structures to improve training efficiency. The model can process multiple holes with variable sizes at any location in the image during model testing. Experimental results on three types of dataset prove that can produce superior inpainting results than previous methods. SN-PatchGAN [16] is another famous GAN-based approach for image inpainting. SN-PatchGAN leverages a gated convolution to address the problem that vanilla convolution considering all pixels of the input image is valid by designing a feature selection module. Also, a path-based loss is proposed to stable the training process. Experimental results on image inpainting indicate that the SN-PatchGAN can generate more flexible results compared to previous methods.

B. Image Super-Resolution

Image super-resolution receives a low-resolution image and converts it high-resolution one. Progressive growing GAN [17] increases the layers of generator and discriminator gradually, starting from few layers that can only produce low-resolution images and adding new layers to the model progressively to enhance the fine details during training. The strategy can not only speed up but also stabilize the training process, allowing GAN to generate high-quality images. The SRGAN [18] proposed to add an additional loss to constrain generated images to be natural images. Single image super-resolution, especially SRGAN, can generate photorealistic images from down-sampled images. SRGAN can produce a 4× up-scaling image from a given low-resolution image. The training loss function for SRGAN includes adversarial loss similar to previous GANs, a perceptual loss calculating via a pre-trained classifier, and a content loss that is more invariant to changes in pixel space. The experimental results show that SRGAN can recover textures from low-resolution images on benchmark datasets.

C. Reenter private information

Person re-identification aims to identify the same person in pictures captured by the camera. It is usually used to track a particular person in videos recorded by cameras, which is important in finding missing people and security applications. Two major challenges have been identified for the person re-identification problems: the lack of cross-view paired training, and the drastically changing appearance and pose of a person across views. Various GAN-based approaches such as PTGAN [19], PN-GAN [20], and IPGAN [21] address these issues by producing realistic images for data augmentation to enhance the generalization ability of the person re-identification model.

V. Challenges

Although GANs are very helpful in some application fields, several challenges can also easily be found in practice applications. We can divide the GAN challenge into three main problems: mode collapse, non-convergence and instability, and high sensitivity to hyperparameters. and performance indicators. Collapse mode meant that the generator can only produce few representative samples. Non-convergence and instability issues are mainly caused by the unbalanced training of GAN.

VI. Conclusion

In this study, we investigate the fundamentals, variants, as well as applications of GAN. GAN is a deep learning model that receives random noise or other types of input and then generates data similar to real ones. While GAN is a very promising deep learning model, several disadvantages such as mode collapse, unstable training process, and high sensitivity to hyperparameters can also be found. We suggest three solutions to mitigate these problems: designing novel network architecture, using advanced loss functions, and introducing alternative optimization algorithms, also, A potential solution to the problem of finding good GAN hyperparameters is to use Weights and Biases to track your experiments.

References

[1] W. Pan, L. Zhang, and C. Shen, "Data-driven time series prediction based on multiplicative neuron model artificial neuron network," *Appl. Soft Comput.*, vol. 104, p. 107179, 2021.

[2] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[3] H. Zhang, V. M. Patel, B. S. Riggan, and S. Hu, "Generative adversarial network-based synthesis of visible faces from polarimetrie thermal faces," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 100–107.

[4] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5505–5514.

[5] G. Antipov, M. Baccouche, and J.-L. Dugelay, "Face aging with conditional generative adversarial networks," in *2017 IEEE international conference on image processing (ICIP)*, 2017, pp. 2089–2093.

[6] X. Wang *et al.*, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, p. 0.

[7] K. Hamada, K. Tachibana, T. Li, H. Honda, and Y. Uchida, "Full-body high-resolution anime generation with progressive structure-conditional generative adversarial networks," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, p. 0. https://eccv2020.eu/

[8] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[9] Z. Yang, L. Zhao, S. Wu, and Y. C. Chen, "Lung Lesion Localization of COVID-19 from Chest CT Image: A Novel Weakly Supervised Learning Method," *IEEE J. Biomed. Heal. Informatics*, 2021, doi: 10.1109/JBHI.2021.3067465.

[10] N. K. Singh and K. Raza, "Medical Image Generation Using Generative Adversarial Networks: A Review," *Heal. Informatics A Comput. Perspect. Healthc.*, pp. 77–96, 2021.

[11] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," *arXiv Prepr. arXiv1411.1784*, 2014, [Online]. Available: http://arxiv.org/abs/1411.1784.

[12] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*, 2017, pp. 214–223.

[13] M. Tao *et al.*, "DF-GAN: Deep Fusion Generative Adversarial Networks for Text-to-Image Synthesis," *arXiv Prepr. arXiv2008.05865*, 2020, [Online]. Available: http://arxiv.org/abs/2008.05865.

[14] S. Suh, H. Lee, P. Lukowicz, and Y. O. Lee, "CEGAN: Classification Enhancement Generative Adversarial Networks for unraveling data imbalance problems," *Neural Networks*, vol. 133, pp. 69–86, 2021.

[15] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," 2006.

[16] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Free-form image inpainting with gated convolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4471–4480.

[17] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2018, [Online]. Available: https://openreview.net/forum?id=Hk99zCeAb.

[18] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[19] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer gan to bridge domain gap for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 79–88.

https://ieeexplore.ieee.org/xpl/conhome/1000147/all-proceedings

[20] X. Qian *et al.*, “Pose-normalized image generation for person re-identification,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 650–667.

[21] J. Liu *et al.*, “Identity preserving generative adversarial network for cross-domain person re-identification,” *IEEE Access*, vol. 7, pp. 114021–114032, 2019.