ISSN: 2664-9640

A review of the research on Generating Adversarial Networks

Mingyu Yang

Chengdu University of Technology, Chengdu 610059, China.

Abstract

Since Generating Adversarial Networks(GAN) was proposed, this direction has become a research hotspot of artificial intelligence.GAN the idea of using a zero-sum game method, composed of generator and the discriminant apparatus, generator are responsible for generating the sample distribution, a discriminant changer discriminant input is a real sample or samples, a generator and a constant interaction optimization criterion, finally achieve the optimal effect. GAN model proposed is very novel, but there are also many shortcomings, therefore with the deepening of the research, GAN for continuous improvement and optimization of GAN derivative model also emerge in endlessly, has been widely applied in the field, audio visual field, in the field of natural language and various other fields.Starting from traditional GAN, this paper summarizes the more prominent aspects of the research on GAN in recent years. First, it introduces the basic theory of traditional GAN, then analyzes the main derived models of GAN in recent years, and finally summarizes the main application results of GAN in different fields.

Keywords

Generating adversarial networks, Gradient disappearance, Image field, Style transfer.

1. Introduction

In recent years, with the improvement of computing power, the accumulation of data and the research on animal neural networks, the field of artificial intelligence has developed rapidly. With people's enthusiasm for deep learning research rising, deep learning has achieved good results in many fields. Various frameworks and models have been put forward and continuously improved.Among them, generating antagonistic network (GAN) is a generation model proposed by Goodfellow et al.[1] in 2014. GAN's proposal is inspired by a two-person zero-sum game (that is, the sum of the interests of two participants in the game is zero, and the interests of one party is the loss of the other party), and this model contains a generator and a discriminator. Generator is responsible for the capture of real data sample distribution and generate new data distribution, and the discriminant apparatus is a two splitters, discriminant of the input data is real data or data generated. GAN optimization is the minimax game "problems, to the best of the generator to generate data distribution could be close to the real data distribution, thus "confused" discrimination. In the past two years, GAN has become a popular research direction. More and more papers on GAN have been published, including the improvement of GAN theory, the improvement of GAN model and the application of research.LeCun Y, a famous scholar, praised the GAN model highly, calling it "the coolest thing ever" and "the most interesting idea in machine learning research in the past decade". At present, the generation of adversarial network is mainly applied in the field of computer images and vision. It can generate realistic images, generate faces, detect targets, and generate real scenes and apply them to driverless scenes. The image can also be repaired[2] and converted [3] according to the context of the image.In addition, GAN can also be applied to text generation, speech and language generation, video prediction, etc.

ISSN: 2664-9640

2. The Original GAN

As shown in Figure 1, when G outputs the real sample x_{real}, expect D to output a high probability. When the input of the D is some simple obey a certain distribution (e.g. gaussian distribution) of the random noise z, the output is a realistic generated sample $G(z)$ with x, it entered into the discriminant D, for D, when expected output low probability (close to zero), G to do quantity deceives tricked into D, D lose out high probable rate ($G(z)$ wrongly convicted for x_{real}). For D, x_{real} prints 1, and $x_{\text{fake}} = G(z)$ prints 0. Thus, G and D form competition and confrontation. Through this process, once G has enough understanding of the distribution of training samples, it can generate new samples with similar characteristics. D wants to achieve its goal by constantly optimizing itself to prevent x_{fake} from deceiving it. Conversely, G also optimizes itself to produce very realistic x_{fake}, making D as difficult to discern as possible. Finally, the argument generated by G is either x_{fake} or x_{real}, and for most of the time, the correct probability of D is stable.

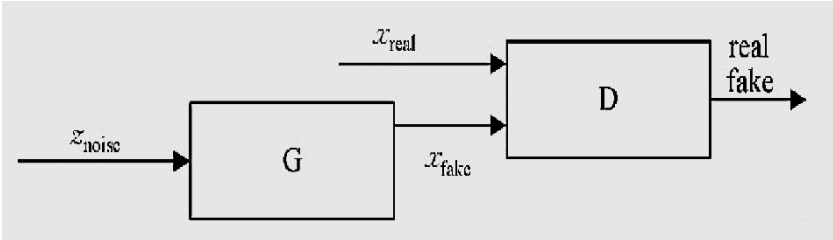

Figure 1: GAN model structure

GAN's idea comes from the two-person zero-sum game in game theory[4]. Generator and discriminator can be regarded as two players in the game. In the process of model training, the generator and discriminator will update their own parameters to minimize the loss, and finally reach a Nash equilibrium state through continuous iterative optimization, at which time the model reaches the optimal state. The objective function of GAN is defined as:

$$\min_G \max_D V(D,G) = E_{x\sim p_{\text{data}}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\lg(1 - D(G(z)))] \tag{1}$$

3. The Improved GAN

In view of the training difficulties of GAN and other problems, researchers finally proposed a lot of variants based on GAN through continuous exploration. This chapter mainly discusses some typical GAN variants.

3.1. CGAN

Conditions was born into a type of resistance networks (Conditional Adversarial Nets, CGAN) [5] model for G and D, C_{class} adds additional information as a condition, used to guide the production of samples. If the conditional variable C_{class} is a category label, then CGAN is to turn an unsupervised GAN into a supervised GAN. The model structure of CGAN is shown in Figure 2.

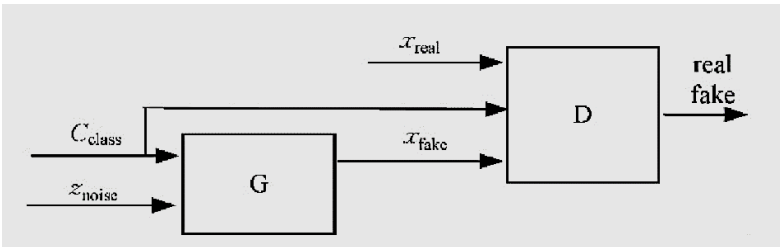

Figure 2: CGAN model structure

ISSN: 2664-9640

3.2. InfoGAN

GAN's strong learning ability can finally learn the distribution of real samples, but it is not clear about the correspondence between input noise signal z and semantic features of data. An ideal situation is to have a clear correspondence between them so that the corresponding changes can be achieved by controlling the corresponding dimensional variables. For example, for MNIST handwritten numeral recognition project, if the corresponding relationship is known, the light, stroke weight and font inclination of the output image can be controlled. InfoGAN[6] solved this problem by dividing input noise z into two parts, one is noise signal z and the other is interpretable and implicit signal c . The structure of InfoGAN model is shown in Figure 3.

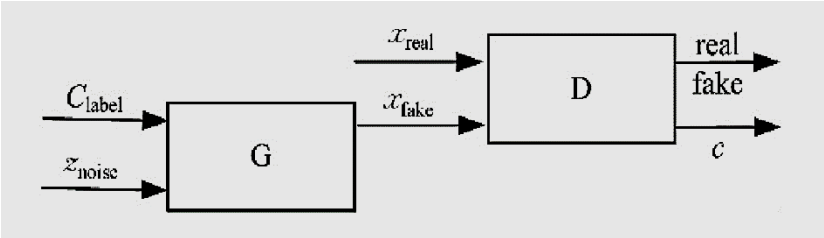

Figure 3: InfoGAN model structure

3.3. CatGAN

In the clustering method of discriminant type, a certain distance is generally used as the criterion of degree, while the CatGAN model proposed in the paper[7] uses the sample entropy as the criterion to construct GAN. Specifically, for real samples, CatGAN hopes that D can be divided into real samples with high confidence, but it is uncertain to divide the generated samples into which category. The goal of G , therefore, is to generate samples with a high degree of certainty that "categorize them into a certain category" and try to fool D . The target of CatGAN can be obtained by combining the confidence target with the optimal target of GAN's authenticity identification. The CatGAN model structure is shown in Figure 4.

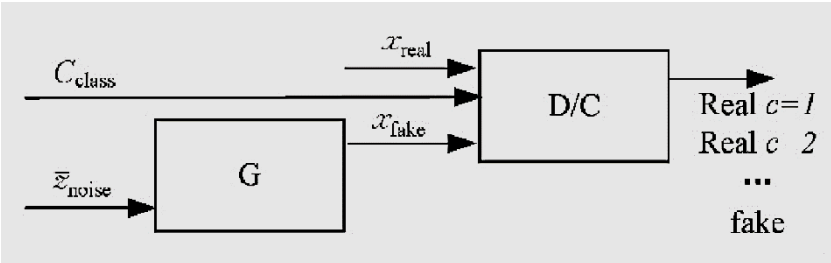

Figure 4: CatGAN model structure

3.4. LAPGAN

The main operations of LAPGAN model are up-sampling and down-sampling, and the advantage of LAPGAN model is that only the residual difference between the sample and the generated sample is considered each time[8]. To some extent, LAPGAN is similar to the residual network, and its learning process is shown in Figure 5.

ISSN: 2664-9640

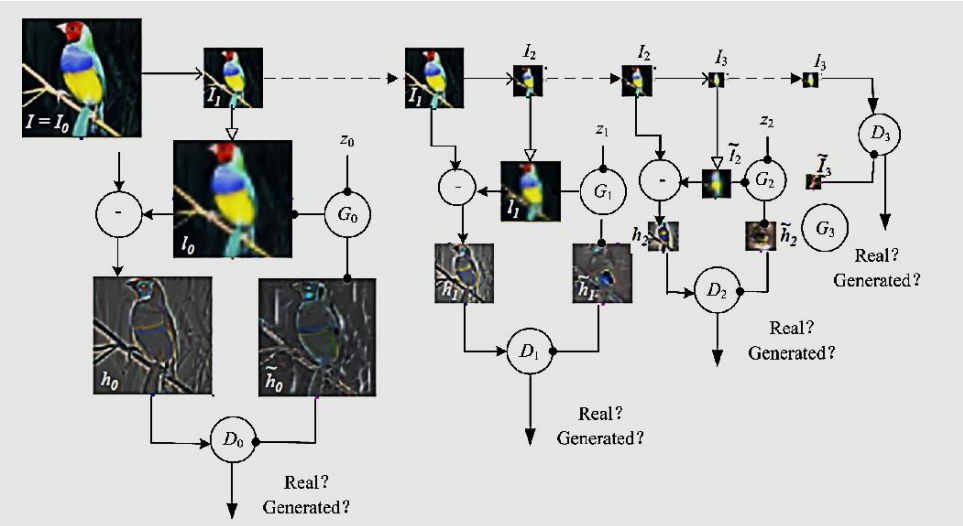

Figure 5: Learning process of LAPGAN model

3.5. f-GAN

Literature[9] proves that GAN is a special case when f-divergence takes a specific measure. Among them, f-divergence refers to the distance of probability distribution, such as KL divergence, Pear-son divergence, etc. Specifically, the author decomposed the optimization steps of GAN, and transformed the estimation problem of real sample distribution into the minimization problem of f-divergence, namely, the f-GAN model.

3.6. EBGAN

Reference[10] expands GAN from the Angle of energy mode, and puts forward EBGAN model. In this model, D is regarded as an energy function, which has a smaller energy value in the area near the real sample domain and a higher energy value in the area outside the real sample domain. Therefore, in EBGAN, an interpretation of the energy model is given to GAN, that is, G aims to generate samples with the minimum energy, while D aims to give these generated samples with the higher energy.

The advantage of EBGAN model is that more and broader structures and loss functions can be used to train GAN structures.

The structure of EBGAN model is shown in Figure 6. On the stability side of the model, EBGAN is better than GAN and can produce clearer and more realistic images[11].

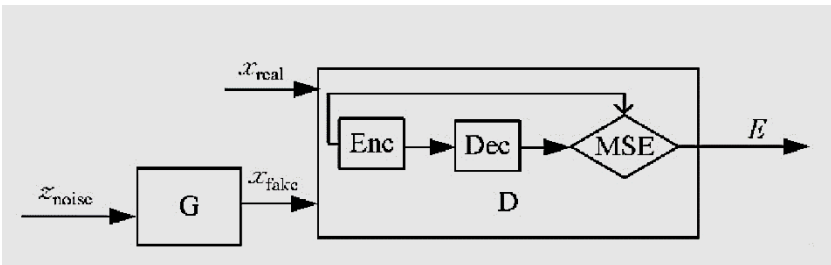

Figure 6: EBGAN model structure

ISSN: 2664-9640

3.7. DCGAN

DCGAN is a relatively good improvement after GAN, and its main improvement is mainly in the network structure. So far, the network structure of DCGAN has been widely used, and DCGAN has greatly improved the stability of GAN training and the quality of generated results. DCGAN model structure as shown in figure[12], compared with the original GAN, almost completely use the convolution DCGAN layer instead of the link layer, the discriminant is almost and generator symmetrical, we can see from the above, the entire network without the existence of pooling and sampling on the layers, in fact is the use of the convolution with step (fractional - strided) instead of the sample, in order to increase the stability of the training.

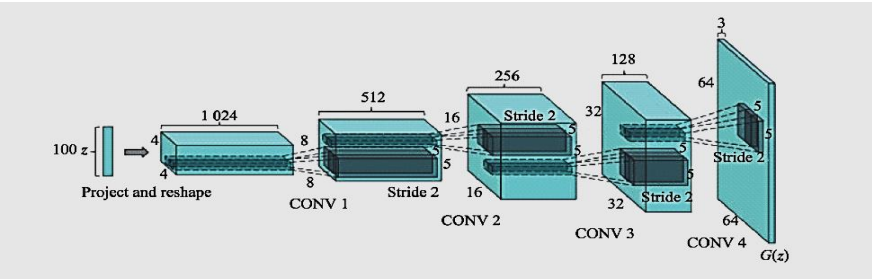

Figure 7: DCGAN model structure

4. The Application of GAN

Since the advantage of GAN network is that it does not need to design special loss functions for application problems and does not need to explicitly model data distribution, it is widely used in image, text, speech and other fields.

4.1. High resolution image

High-resolution technology refers to the reconstruction of the corresponding high-resolution image from the low-resolution image. When the clear image cannot be acquired, it has important application value, such as monitoring equipment, medical image, etc. However, the traditional deep learn-based method is to collect information from the original image through the convolutional network, and the generated high-resolution image lacks specific texture details and is prone to blur. GAN, as a generation model, can solve this problem skillfully. SRGAN [13], based on the original loss of GAN, uses residual network and perceptual similarity loss to generate images with rich details. The perceptual loss focuses on the feature error of the middle layer of the discriminator rather than the pixel error of the output results. The peak signal-to-noise ratio(PSNR) is often used to measure the image quality of high-resolution reconstruction. The larger the PSNR value is, the better the image quality will be. The value greater than 20 dB meets the reconstructed image standard. Take truncation thought in literature[13], under the same network structure, by against loss training SRGAN image PSNR of 29 dB above, slightly lower than not taken against losses on numerical training image produced by a network, but the latter will be vague concepts, the former can produce very fine texture.

4.2. Target detection and variation

The problem in image detection of small target objects often lies in the low resolution of small target objects, so the intuitive solution is to enlarge the low-resolution image to the high-resolution image to enhance the discriminability. Literature [14] divides discriminators into antagonistic branch and perceptual branch. The antagonistic branch is responsible for the traditional task of generating large images, while the perceptual branch is responsible for ensuring the effectiveness of large images in the test.SeGAN uses a segmentator, a generator

and a discriminator to reconstruct hidden objects [15]. In contrast to scene transformation, object deformations replace objects in an image with specific conditions without changing the background. GeneGAN (Generated GAN) uses an encoder decoder structure to generate GAN. The encoder splits the image into background features and object features, and the decoder reintegrates the background features and object features to be deformed to reconstruct the image [16]. Importantly, in order to separate the feature space, two separate training sets are needed, one is the image set with the object, and the other is the image set without the object. In addition, GAN can also be applied to image blending tasks, which implant an object into the background of another image. GP - Poisson GAN [17] proposed to combine GAN - based image fusion with traditional gradion-based image fusion. GP-GAN attempted to generate high resolution good fusion images by optimizing The Gaussian Poisson equation [18].

4.3. Image translation and style conversion

Image translation and Style Conversion Image style migration is the process of "migrating" the style of one image to another.Deep learning was first realized by using CNN framework [19], but such a model has the problems of slow training speed and too high requirements on training samples. Due to the advantages of GAN's autonomous learning and random sample generation, as well as the reduced requirements for training samples, GAN has achieved fruitful research results in the field of image style transfer [20-26].

5. Summary and Prospect

Since Goodfellow et al. first proposed the model structure of GAN network in 2014, papers have mushroomed like mushrooms to solve various problems and solve the training difficulties of the network by utilizing the exquisite zero-sum game thinking of the network. Theory, the basis of the purpose of this paper is to explain a popular GAN finishing explains the commonly used ones are GAN training model, and the time line according to the logic of theory development of GAN under the conditions in the development of image generation, and a more macro inductive GAN applications in a variety of different scenarios, and summarizes some improve training skills. The author believes that this technology will greatly improve the limitations of deep learning in some application scenarios, especially in the generation of models. From zero to creation has always been a difficult problem to solve, and GAN is just an excellent improvement on the generation of models. At present, relevant applications are still in the initial stage, and there will be more effective and extensive attempts in the future. As for the landing of GAN in application, some prospects are proposed as follows:

(1) GAN truly embodies creativity in image generation and style transformation. With the development of the society, cultural entertainment carries an increasing proportion of life, and the most important thing of the cultural and entertainment industry is rapid and innovative. If the distribution breadth of GAN in generation can be used to create some novel cultural and entertainment products, so as to stimulate people's creativity, the productivity of this industry will be greatly improved. For example, GAN controlled by conditions can automatically generate a series of animation works with a given story background, and the style of animation works may be unique.

(2) GAN's application in the field of image restoration. Not only in the removal and filling of the target in the image, but also in the super-resolution analysis of the image, the deep neural network has a powerful approximation ability for the high-dimensional complex mapping, which can effectively extract the semantic meaning in the image. Combining semantics and texture can help repair the details of works of art, such as mural repair, calligraphy rubbings restoration, etc.

(3) GAN's application in medical image generation. Artificial intelligence to assist doctors in image data analysis is becoming more and more common, and the more diverse the training data samples these ARTIFICIAL intelligence need, the better. But researchers had trouble getting enough images of the lesions to compare with healthy samples. In the similar data enhancement research of various forms(image, speech, language, etc.), GAN has a broad development prospect.

References

[1] Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generativeadversarial nets. Proceedings of the 27th InternationalConference on Neural Information Processing Systems.Montreal, QC, Canada. 2014. 2672–2680.

[2] Denton E, Gross S, Fergus R. Semi-supervised learning withcontext-conditional generative adversarial networks. arXivpreprint arXiv: 1611.06430, 2014.

[3] Zhu JY, Park T, Isola P, et al. Unpaired image-to-imagetranslation using cycle-consistent adversarial networks.Proceedings of 2017 International Conference on ComputerVision. Venice, Italy. 2017. 2223–2232.

[4] Mertens J F, Zamir S. The value of two- person zero- sumrepeated games with lack of information on both sides[J].International Journal of Games Theory, 1971, 1(1): 39-64.

[5] Mirza M, Osindero S. Conditional Generative Adversarial Nets[J]. Computer Science, 2014:2672-2680.

[6] Chen X, Duan Y, Houthooft R, et al. InfoGAN: interpre-table representation learning by information maximizinggenerative adversarial nets[C]//Proceedings of the 2016 Inter-national Conference on Neural Information ProcessingSystems, Barcelona, Dec 9, 2016. Cambridge: MIT Press,2016: 1-8.

[7] TOBIAS J. Unsupervised and semi-supervised learning with cat-egorical generative adversarial networks[C]//ICLR-2016,Spri-ngenberg.2016;876-884.

[8] DENTON E L, CHINTALA S, FERGUS R. Deep GenerativeImage Models using a Laplacian Pyramid of Adversarial Net-works[C]// Advances in Neural Information Processing Sys-tems.2015:1486-1494.

[9] Nowozin S, Cseke B, Tomioka R. f-GAN:Training generative neural samplers using variational divergenceminimization[C]//Advances in Neural Information Processing Systems. 2016:271-279.

[10] ZHAO J B,MICHAELM，YANN L C. Energy-based Genera-tive Adversarial Network[J/OL]. https://arxiv.org/abs/1609.03126.

[11] Shuang M, Jian L F, Chang W C, et al. DA-GAN: instance-level image translation by deep attention generative adve-rsarial networks[C]//Proceedings of the 2018 IEEE Confer-ence on Computer Vision and Pattern Recognition, SaltLake City, Jun 18- 22, 2018. Washington: IEEE ComputerSociety, 2018: 5657-5666.

[12] Zhu W, Miao J, Qing L, et al. Unsupervised representationlearning with deep convolutional generative adversarialnetworks computer science[J]. arXiv:1511.06434, 2015.

[13] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realisticsingle image super-resolution using a generative adversarialnetwork [C]// Proceedings of the IEEE Conference onComputer Vision and Pattern Recognition. Hawaii: IEEE, 2017:4681-4690.

[14] LI J, LIANG X, WEI Y, et al. Perceptual generativeadversarial networks for small object detection [C]//Proceedings of the IEEE Conference on Computer Vision andPattern Recognition. Hawaii: IEEE, 2017:1222-1230.

[15] EHSANI K, MOTTAGHI R, FARHADI A. Segan: segmenting and generating the invisible[C]// Proceedings ofthe IEEE Conference on Computer Vision and PatternRecognition. Utah:IEEE，2018: 6144-6153.

[16] ZHOU S, XIAO T, YANG Y, et al. Genegan:learningobject transfiguration and attribute subspace from unpaired data[EB/OL].[2017-05-14]. https ://arxiv.org/abs/1705.04932.

[17] WU H, ZHENG S, ZHANG J, et al. Gp-gan:towardsrealistic high-resolution image blending[EB/OL]. [2017-03-21]. https://arxiv.org/abs/1703.07195.

[18] Guo Z P, Shang F W. Weakly supervised facial action unitrecognition through adversarial training[C]//Proceedings ofthe 2018 IEEE Conference on Computer Vision and Pat-tern Recognition, Salt Lake City, Jun 18- 22, 2018. Wash-ington: IEEE Computer Society, 2018: 2188-2196.

[19] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks[C]//Proceedings of the 34th International Conference on Mach-ine Learning, Sydney, Dec 18-21, 2017: 1730-1741.

[20] Taigman Y, Polyak A, Wolf L. Unsupervised cross domainimage generation[J]. arXiv:1611.02200, 2016.

[21] Chen Y, Lai Y K, Liu Y J. CartoonGAN: generative adver-sarial networks for photo cartoonization[C]//Proceedings ofthe 2018 IEEE Conference on Computer Vision and PatternRecognition, Salt Lake City, Jun 18-22, 2018. Washington:IEEE Computer Society, 2018: 9465-9474.

[22] Chang H, Lu J, Yu F, et al. PairedCycleGAN: asymmetricstyle transfer for applying and removing makeup[C]//Proceedings of the 2018 IEEE Conference on ComputerVision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 40-48.

[23] Choi Y, Choi M, Kim M, et al. StarGAN: unified generativeadversarial networks for multidomain image-to-image tran-slation[C]//Proceedings of the 2018 IEEE Conference onComputer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington:IEEE Computer Society, 2018:8789-8797.

[24] Shuang M, Jian L F, Chang W C, et al. DA-GAN: instance-level image translation by deep attention generative adve-rsarial networks[C]//Proceedings of the 2018 IEEE Confer-ence on Computer Vision and Pattern Recognition, SaltLake City, Jun 18- 22, 2018. Washington: IEEE ComputerSociety, 2018: 5657-5666.

[25] Azadi S, Fisher M, Kim V, et al. Multicontent GAN for few-shot font style transfer[C]//Proceedings of the 2018 IEEEConference on Computer Vision and Pattern Recognition,Salt Lake City, Jun 18- 22, 2018. Washington: IEEE Com-puter Society, 2018: 7564-7573.

[26] Hong Y Y, Di H, Yun H W, et al. Learning face ageprogression: a pyramid architecture of GANs[C]//Proceed-ings of the 2018 IEEE Conference on Computer Vision andPattern Recognition, Salt Lake City, Jun 18-22, 2018. Was-hington: IEEE Computer Society, 2018: 31-39.