International Journal of Engineering Research And Management (IJERM)
ISSN: 2349- 2058, Volume-06, Issue-12, December 2019

A Review of Generative Adversarial Network Applications

Gao Zengbin

Abstract— **In recent years, deep learning has achieved good results, and each born technology has been tried in different fields. The generative adversarial network has received extensive research and high attention from the artificial intelligence academia and industry since its introduction. With the development of deep learning technology, the generative adversarial model has been continuously promoted in theory and application. Generative Adversarial Network (GAN) has become a research hotspot in the field of artificial intelligence. It brings new vitality to the research of unsupervised learning and is of great significance to the development of generative models. Aiming at the hot model of generative adversarial network, the basic principles of generative adversarial network and several typical variations improved on it are introduced. At present, generative adversarial networks have been successfully applied in the fields of computer vision, speech, and natural language processing.**

Index Terms—**Deep Learning, GAN, Computer Vision, NLP.**

I. INTRODUCTION

Artificial intelligence is a copy of human intelligence on a computer. Machine learning refers to the ability of a machine to learn using large data sets rather than hard-coded rules. As the most important branch of machine learning, deep learning has developed rapidly in recent years. Deep learning [1] is reforming machine learning through excellent performance in certain tasks. It uses a large amount of data to train computers to do something that only humans could do before. As a result of the development of machine learning to a certain stage, in recent years, the deep learning technology has attracted wide attention from all walks of life, not only in the academic world, but also in the industrial world, it has achieved major breakthroughs and widespread applications. The most widely used research fields are natural language processing, speech recognition and image processing. Things like how to tell what objects are in the image, what people talk to when they call, translate documents into another language, help robots explore the world and respond to things in a timely manner, and so on.

The explosive development of deep learning relies on various excellent frameworks and algorithm models to be successively proposed, and gradually improved and improved. In June 2014, Ian Goodfellow of the University of Montreal and other researchers (including Joshua Bengio) proposed a new neural structure generating adversarial neural network (GAN) [2]. In the past two years, GAN has become a popular research direction, and more and more papers on GAN have been published, which include the improvement of GAN theory, the improvement and application of GAN models. The well-known scholar LeCun Y evaluated the GAN model quite well. Gao, calling it "the coolest thing ever" and "the most interesting idea in machine learning research in the past decade". At present, generative adversarial networks are mainly used in the field of computer images and vision, which can generate realistic images, Faces can be generated, targets can be detected, real scenes can be generated and applied to driverless scenes; image repair can also be performed according to the context of the image [3], and the image can be transformed. In addition, GAN can also be applied to Text generation, speech and language generation, video prediction, etc..

Manuscript received November 30, 2019
Gao Zengbin, School of Computer Science and Technology, Tiangong University, Tianjin, 300387, China

II. BASIC THEORY

Generating adversarial neural networks

Generating adversarial neural networks is not the only way to apply neural networks to unsupervised learning. There are also Boltzmann machines [4] and automatic decoders [5]. All three are dedicated to extracting features from data by learning identity functions, and all rely on Markov chains to train or generate samples. The original intention of GAN design is to avoid the use of Markov chains, because the calculation cost of Markov chains is very high. Another advantage over Boltzmann machines is that GANs are much less restrictive.

As a generative method, generative adversarial networks are basically composed of two competing neural networks. One is called a generator network and the other is a discriminator network. The generator tries to capture the data distribution according to the input sample data distribution, and generates a new data distribution. The new data distribution needs to be as close to the real data distribution as possible. Real. The role of the discriminator is to determine whether the input data comes from real data or data generated by the generator, and estimate the probability that the sample comes from the training data instead of the generator. That is, the generator attempts to create the same samples as the training set and discriminator attempts to distinguish what the generator is creating from the original samples in the training set. During training, the generator tries to better fool the discriminator and the discriminator tries to catch fakes generated by the generator. This competitiveness helps them imitate any data distribution. By adversarial training of two deep neural networks and optimization using stochastic gradient descent, it not only avoids the calculation of the distribution function caused by repeated application of the Markov chain learning mechanism, but also does not need the lower limit of variation or approximate inference, which greatly improves Application efficiency has profound implications for the development of generative models.

III. Methods

GAN Optimization Derived Model

With the development of artificial intelligence, researchers are enthusiastic about GAN research, including improvements to the GAN model framework and theory, transformation of the GAN model, or combination of other learning methods, making it applicable to more scenarios.

Radford et al. Proposed a deep convolutional generative adversarial network (DCGAN) [6] and applied a convolutional neural network (CNN) to the generative adversarial network. By changing the architecture of GAN, the stability of GAN training is improved. The principle of DCGAN is the same as that of GAN, but the DCGAN network replaces G and D of the GAN network with two convolutional neural networks.The DCGAN network is made more stable by making some specific restrictions on the two convolutional neural networks. And use the obtained feature representations for image classification, and get better results to verify the expression ability of the generated image feature representations.

In 2017, Arjovsky et al. Proposed a new method for training generative adversarial networks, WGAN [7]. WGAN pointed out that GAN's use of cross entropy is not suitable for measuring the distance between distributions with disjoint parts, which leads to unstable training. Therefore, WGAN uses wassertein distance to measure the distance between the generated data distribution and the real data distribution.

BEGAN [8] proposed a new method for discriminating the authenticity of generated samples. A discriminator is used to determine the error distance between the generated sample distribution and the real sample distribution. When the error distance is small, that is, the generated sample distribution is similar to the real sample distribution, the generation The samples generated by the processor are close to the real samples. Therefore, many types of GAN structures and loss functions can be used for training.

InfoGAN [9] model. The input model of GAN is a continuous noise signal without any constraints. As a result, GAN cannot use this noise signal, and the specific dimensions of the noise signal correspond to the semantic features of the data, so it is not an interpretable expression. InfoGAN decomposes the noise signal into an incompressible noise and an interpretable hidden variable, making the noise data interpretable

In 2017, the SeqGAN [10] model was proposed, and the strategy gradient of reinforcement learning was used to solve the problem that GAN can only be applied to the continuous output of the generator. In this model, the discriminator is a two-classifier and receives a sequence (sentence) generated by a generator. Both pre-training and adversarial training are trained using cross-entropy as the objective function. The generator is a language model based on recurrent recurrent neural networks. In adversarial training, the author of this paper considers it as a sequence decision process, and each sequence state is a generated sequence, which is trained using a strategy gradient. SeqGAN is updated through a policy gradient, and the goal of the generated model is the following formula, where a discriminator is used as a reward guidance generator.

The Application of GAN

A. *Image generation*

The original purpose of GAN was to generate images with input vectors. The pixels of the generated images were not very high at first. Later, with the continuous research of GAN, the quality of generated images is also continuously improved. Ledig et al. [11] proposed an increase in image resolution. GAN-SRGAN, which can convert low-resolution images into super-resolution images. The SRGAN generator uses a residual network (ResNet) [12], and the discriminator uses VGG [13]. Karras et al. [14] Progressive GAN—a progressively trained GAN. During the training process, a network layer capable of processing high resolution is added to generate false and real images. In addition, ProGANSR [15] can also generate high resolution images. It The multi-scale progressive principle is used, and the upsampling reconstruction quality is also increased. GAN can generate human faces, and non-existent human faces can be generated based on real samples; animals and other objects can also be generated. In addition to generating human, animal and other objects , GAN can also generate real scenes, which can be applied to scenarios such as driverless. In these applications, the requirements for image resolution will also become higher and higher with the development of technology.

B. *Image Style Conversion*

Style transfer refers to applying the style learned from the target image to the source image, so that the source image retains the content while having the style of the target image, such as turning an image into an oil painting style image. Li et al. [16] proposed a method based on Markov chain generation adversarial network (MGAN) to achieve image style transfer. Azadi et al. [17] use GAN to perform style transfer learning on fonts, which can generate a variety of different style fonts. In addition, style transfer also Can be applied to other fields such as music, but the effect is not as good as the style transfer of the image.

C. *Video generation and prediction*

On the basis of generating images, GAN can also generate videos. Tukyakov et al. [18] proposed a method of generating videos by decomposing actions and content—MoCoGAN, which maps a sequence of random vectors to a sequence of video frames, each The random vector contains an action and a content. Video prediction is to predict the next content of the video based on the current frame or frames of video, such as when a person picks up a ball to predict what the person will do next. Xiong et al [19] proposed a multi-stage dynamic generation adversarial network to generate time-lapse video. First, a realistic content video is generated for each frame, and then motion modeling is performed to make the object movement between adjacent frames more vivid, while maintaining the realistic content. Another video prediction method is the method proposed by Jang et al. [20]-specifying the future by appearance and motion, reducing uncertainty, and solving the problem of how to choose the model when there are multiple correct and equal possible futures. problem.

D. *The text generated*

In addition to generating pictures, GAN can also generate text. Zhang et al. [21] used GAN to generate text, and its discriminator is a convolutional neural network (CNN). GAN is generally based on continuous space, but it can also be applied to discrete data. However, there are two problems with GAN for generating sequences. GAN generates continuous data, and it is difficult to directly generate discrete sequences. Another problem is that GAN can only score the entire generated sequence, and it is difficult to judge a part of the generated sequence. It now generates the quality of the sequence and the quality of the entire sequence afterwards. SeqGAN [22] can solve the above two problems, it combines reinforcement learning, can generate discrete sequences, and uses the policy gradient training generator G to solve the gradient when outputting discrete data The problem of generating models cannot be returned, and reward signals are obtained by Monte Carlo search. WGAN can also be popularized for discrete data space. SeqGAN can be applied to speech Into, poetry generation, machine translation, and the like to generate a dialogue. MaskGAN [23] The text content to fill the missing context, and introducing actor-critic architecture.

In addition, GAN can also generate image-text matching images based on text descriptions. StackGAN [24] can generate refined high-resolution images based on input text descriptions. AttnGAN [25] can generate more detailed images by focusing on natural language descriptions Relevant vocabulary can synthesize fine details in different sub-regions of the image.In addition, a deep attention multimodal similarity model is also proposed to calculate the fine-grained image-text matching loss used to train the generator.

Existing problems and development trends

A. *Existing problems*

The GAN model is a framework that can be combined with other methods. The GAN model includes a generator G and a discriminator D. During the training process, the synchronization between the generator G and the discriminator D needs to be maintained, so it is easy to cause training instability. Although WGAN uses EM distance instead of JS divergence, it is not necessary to maintain the synchronization between generator G and discriminator D, which solves the problem of unstable training. However, the training stability problem of models using JS divergence has not been obtained. Very good solution. In addition, GAN also has a mode collapse problem. WGAN can solve the mode collapse problem, but if the parameter of the loss function is not set properly, it will cause the gradient to disappear or the gradient to explode. Therefore, the mode collapse problem has not been completely solved .

B. *The Development Trend*

Since the introduction of GAN, it has received widespread attention. In the past two years, research interest in GAN has only increased. Although GAN can achieve huge functions, some problems in this model have not been completely solved, affecting its generation effect. In terms of model training stability and evaluation indicators, there is still room for improvement. GAN has excellent application effects in the image field, and has good applications in other fields. GAN's application of text is based on discrete data, and the effect is not ideal. There is a lot of room for improvement in this area. With the continuous development of artificial intelligence, the research on unmanned driving is also maturing. GAN can be combined with its generating effect on the image field and applied to unmanned driving. It will also be a very popular research direction. Video and voice are necessary in daily life. As people's needs increase, the application of GAN in video and voice also needs to be improved.

IV. CONCLUSION

Generative adversarial networks include two competitive neural network generative models. They have strong generative capabilities. They have unlimited modeling capabilities without restrictions. They are mainly used in image vision. Today GANs can generate super-resolution Images, which can be combined with semi-supervised, reinforcement learning, and feature learning, have been widely used in images, vision, and text. GAN has a significant application effect in the image field. Although it is also widely used in other fields, the effect achieved is not as good as that of images. Good. How can GAN be better applied to more fields, and constantly optimize performance while applying, and exert greater advantages, it will be interesting research. With the development of artificial intelligence, people are pursuing more intelligent, How to combine different technologies such as deep learning to improve the application effect of GAN and make GAN more intelligent is the future development of GAN. GAN is a relatively important model in the field of deep learning and provides a powerful tool for unsupervised learning models. Computing framework is also an important tool for artificial intelligence research. An important feature of GAN is its ability to understand the complex world around it like a human. Continue to explore in this direction of GAN, it is possible to successfully create a machine learning model that can understand the world at a higher level than recognition.

REFERENCES

[1] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.

[2] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.

[3] Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: Feature learning by inpainting[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2536-2544.

[4] Ackley D H, Hinton G E, Sejnowski T J. A learning algorithm for Boltzmann machines[J]. Cognitive science, 1985, 9(1): 147-169.

[5] Ballard D H. Modular Learning in Neural Networks[C]//AAAI. 1987: 279-284.

[6] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.

[7] Arjovsky M, Chintala S, Bottou L. Wasserstein gan[J]. arXiv preprint arXiv:1701.07875, 2017.

[8] Berthelot D, Schumm T, Metz L. Began: Boundary equilibrium generative adversarial networks[J]. arXiv preprint arXiv:1703.10717, 2017.

[9] Chen X, Duan Y, Houthooft R, et al. Infogan: Interpretable representation learning by information maximizing generative

adversarial nets[C]//Advances in neural information processing systems. 2016: 2172-2180.
[10] Yu L, Zhang W, Wang J, et al. Seqgan: Sequence generative adversarial nets with policy gradient[C]//Thirty-First AAAI Conference on Artificial Intelligence. 2017.
[11] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 105–114.
[12] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.
[13] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556, 2014.
[14] Karras T, Aila T, Laine S, et al. Progressive growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv: 1710.10196v3, 2018.
[15] Wang YF, Perazzi F, McWilliams B, et al. A fully progressive approach to single-image super-resolution. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, UT, USA. 2018. 977–986.
[16] Li C, Wand M. Precomputed real-time texture synthesis with Markovian generative adversarial networks. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands. 2016. 702–716.
[17] Azadi S, Fisher M, Kim V, et al. Multi-content GAN for few-shot font style transfer. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 7564–7573.
[18] Tulyakov S, Liu MY, Yang XD, et al. MoCoGAN: Decomposing motion and content for video generation. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 1526–1535.
[19] Xiong W, Luo WH, Ma L, et al. Learning to generate time-lapse videos using multi-stage dynamic generative adversarial networks. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 2364–2373.
[20] Jang Y, Kim G, Song YL. Video prediction with appearance and motion conditions. Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden. 2018. 3496–3510.
[21] Zhang YZ, Gan Z, Carin L. Generating text via adversarial training. Proceedings of the 2016 Conference on Neural Information Processing Systems. Barcelona, Spain. 2016. 1–6.
[22] Yu LT, Zhang WN, Wang J, et al. SeqGAN: Sequence generative adversarial nets with policy gradient. Proceedings of the 31st AAAI Conference on Artificial Intelligence. Phoenix, AZ, USA. 2016. 2852–2858.
[23] Fedus W, Goodfellow I, Dai A M. MaskGAN: better text generation via filling in the_[J]. arXiv preprint arXiv:1801.07736, 2018.
[24] Zhang H, Xu T, Li H, et al. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 5907-5915.
[25] Xu T, Zhang PC, Huang QY, et al. AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 1316–1324.