International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 12 Dec 2023, pp: 39-45 www.ijaem.net ISSN: 2395-5252

Review of Intelligent Image Generation Technology and Its Application Prospects

QIAO Jiaxin

School of Management Science and Engineering, Anhui University of Finance & Economics, Bengbu City, Anhui Province, China.
Corresponding Author: QIAO Jiaxin

Date of Submission: 06-12-2023 Date of Acceptance: 16-12-2023

ABSTRACT:With the rapid development of artificial intelligence and computer vision technology, intelligent image generation technology has become one of the hot topics in the field of deep learning research, and has made many key breakthroughs. From Variational AutoEncoder (VAE), Generative Adversarial Networks (GAN), to the current mainstream Diffusion Model (DM), this article comprehensively analyzes the working principles and derivative models of three image generation technologies, and focuses on the characteristics of the three image generation technologies, We have delved into their advantages, disadvantages, and improvement directions, studied the challenges faced by intelligent mapping technology, and looked forward to its future development trends..

KEYWORDS: image generation technology, variational auto-encoder(VAE), generative adversarial network(GAN), diffusion model(DM).

I. INTRODUCTION

With the rapid development of artificial intelligence and computer vision technology, intelligent image generation technology has become one of the hot topics in the field of deep learning research, and has made many key breakthroughs. From Variational Auto-Encoder (VAE)[1], Generative Adversarial Networks (GAN)[2], to the current mainstream Diffusion Model (DM)[3], this article comprehensively analyzes the working principles and derivative models of three image generation technologies, and focuses on the characteristics of the three image generation technologies, We have delved into their advantages, disadvantages, and improvement directions, studied the challenges faced by intelligent mapping technology, and looked forward to its future development trends.

II. INTELLIGENT MAPPING TECHNOLOGY

2.1 Variational Auto-Encoding (VAE)

2.1.1 Working principle of VAE model

In 2014, a new generative model, variational auto-encoder VAE, was proposed. Its structure is shown in Figure 1 [4] [5],VAE (Variational Auto-Encoder) is an unsupervised deep learning algorithm primarily used for data dimensionality reduction, generation, and reconstruction. and it also consists of an encoder (also known as the recognition or inference model) and a decoder (also known as the generative model). In VAE, the focus is on variational reasoning for encoding. Therefore, VAE provides a suitable framework for learning latent variables and effective Bayesian inference. Structurally, VAE is similar to auto-encoders, and the variational auto-encoder model attempts to model the input distribution in a decoupled continuous latent space. It is a directed model that uses approximate inference and can be trained purely using gradient based methods. The basic principle of an auto-encoder is achieved through encoding and decoding. In a convolutional auto-encoder, encoding is dimensionality reduced through convolutional pooling, while decoding is dimensionality increased through deconvolution pooling. The goal is to minimize the difference between the input and output images.

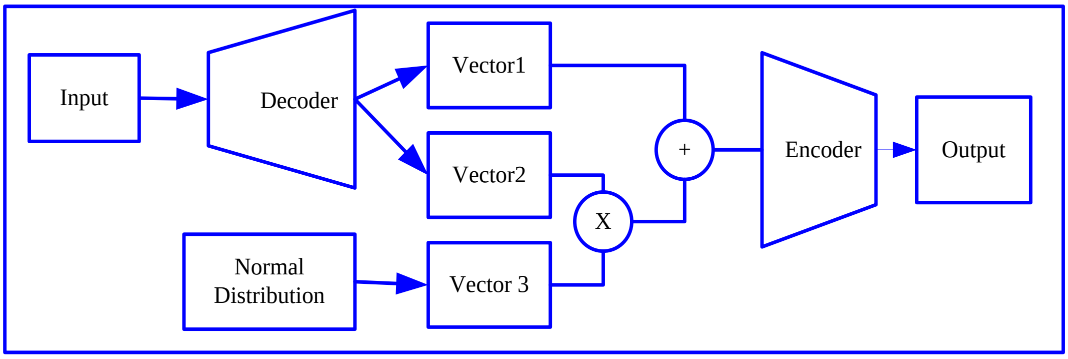

Fig1. Variational Auto-Encoding (VAE) Model Structure

2.1.2 Derived models of VAE

Variational Auto-Encoding is suitable for learning latent spatial images with good structures. In order to effectively maximize the lower boundary of the likelihood function, resulting in blurry and low-quality images, researchers at home and abroad have proposed some new methods to improve them. Below are two types of generative models.

(1) Nouveau VAE (NVAE) [6]. NVAE has designed multi-scale encoders and decoders. Firstly, the encoder undergoes layer by layer encoding to obtain the top-level encoding vector, and then slowly moves down from the top layer to gradually obtain the low-level features; As for the decoder, it is also a process of utilizing it from top to bottom, and this part happens to have similarities with the process generated by the encoder. All NVAE directly share the corresponding parameters, which not only saves the number of parameters but also improves generalization performance through mutual constraints between the two.

(2) The combination of VAE and GAN. A classic direction for improving VAE is to combine VAE with GAN, such as CVAE-GAN, AGE, etc[7-11]. Currently, the most advanced result of this direction is IntroVAE. In theory, this type of work is equivalent to implicitly abandoning the assumption of Gaussian distribution and replacing it with a more general distribution, which can improve the generation effect. Introducing GAN into VAE improves the performance of VAE in generating images.

2.2 Generative Adversarial Network

2.2.1 Working principle of GAN model

The structure of the Generative Adversarial Networks (GAN) model is shown in Figure 2[12-15], which is a widely studied image generation model. In 2014, Generative Adversarial Networks (GANs) were first proposed by Goodfellow et al., and this model became one of the hot research directions in the field of computer vision. It is an extended deep learning model based on convolutional neural networks, which uses two basic models, a generator and a discriminator, to achieve forward propagation and backward discrimination in adversarial games. Practice training is used to continuously improve the generation and discrimination abilities, Make the data of generative networks closer to real data, thereby generating realistic images.

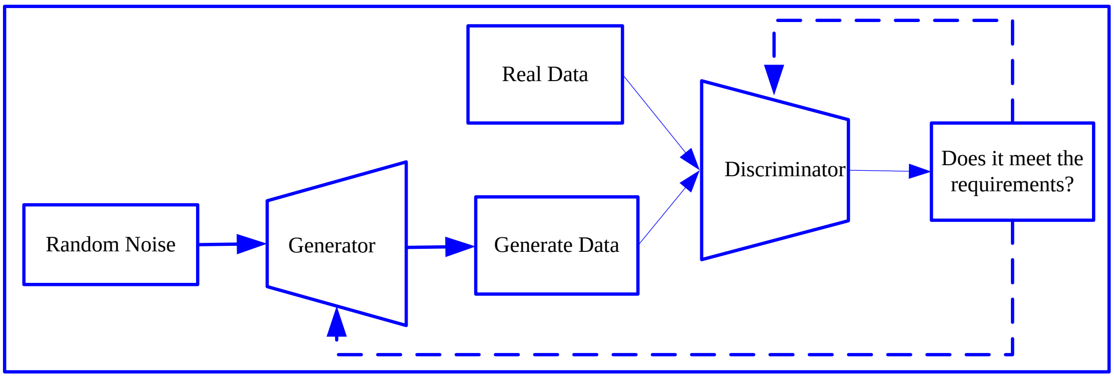

Fig2.Generative Adversarial Network (GAN) Model Structure

2.2.2 Derived models of GAN

The GAN generation process is too free, and the stability and convergence of the training process are difficult to ensure, which can easily lead to pattern collapse and the inability to continue training. The original GAN has problems such as vanishing gradients, difficulty in training, loss of generators and discriminators that cannot indicate the training process, lack of diversity in generated samples, and easy overfitting. Due to the limitations of GAN itself, it is difficult to learn to generate discrete distributions. Many new GAN models or improvements in training techniques are proposed to increase model stability and improve the quality of generated results. In response to the problems existing in the original GAN, researchers at home and abroad have proposed many new methods for improvement. Below are several milestone derived models introduced.

(1) Deep Convolutional Generative Adversarial Network (DCGAN)[16]. A milestone in the development of GAN is the deep convolutional generative adversarial network DCGAN proposed by Radford et al. It combines the convolutional neural network CNN and GAN, which perform well in the field of computer vision. It sets a series of limitations on the network topology of CNN to enable stable training, uses learned feature representations for image classification, and obtains good results to verify the model's feature expression ability.

(2) Conditional Generative Adversarial Network (CGAN)[17]. GAN, as an unsupervised learning method, learns probability distribution patterns from unlabeled datasets and represents them, which is a slow and free process. A natural idea is to add constraints, set goals for the generator, and add the proposed condition GAN or CGAN. When inputting random variables and real data, the GAN model also includes conditional variables. The added information is used to add constraints to the model, guiding the data generation process. In this way, CGAN transforms the unsupervised GAN into a supervised model.

(3) Cascade Generative Adversarial Network (LAPGAN)[18]. The early GAN networks represented by DCGAN generated images with low resolution and poor quality, all of which did not exceed 100 × 100, this is because it is difficult to learn how to generate high-resolution samples at once, and the convergence process is prone to instability. Based on this, LAPGAN hierarchical linkage structures have been proposed and widely used. They refer to pyramid structures in the image field to generate images step by step from coarse to fine, and perform residual learning to gradually improve and ultimately generate higher resolution images.

(4) Multi discriminator and generator generative adversarial network (GMAN)[19]. When facing complex image generation problems, the ability of a single discriminator and generator is insufficient. Therefore, the structure of multiple discriminators and generators has been widely studied by researchers. The benefits of using multiple discriminators bring advantages similar to model integration, and they can independently complete different tasks. The design of multiple generators can achieve division of labor and cooperation, improving pattern richness.

2.3 Diffusion Model (DM)

2.3.1 Working principle of diffusion (DM) model

The structure of the Diffusion Model is shown in Figure 3[20][21]. The working principle of the diffusion model is to define a Markov chain with a diffusion step, continuously add random noise to the data until a pure Gaussian noise data is obtained, and then learn the process of inverse diffusion to generate an image through reverse denoising inference. The diffusion model systematically perturbs the distribution in the data and then restores the data distribution, presenting a gradually optimized nature throughout the process, ensuring the stability and controllability of the model. The diffusion model has a simple definition, high training efficiency, and can generate high-quality samples, sometimes resulting in better results than other types of generative models.

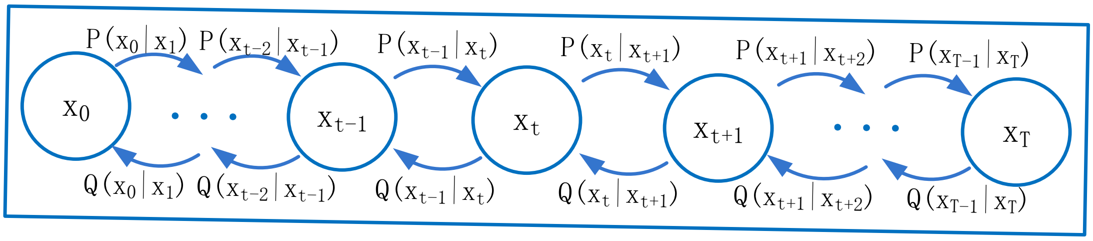

Fig 3.Diffusion (DM) Model Structure

2.3.2 Derivative models of DM

Due to the complexity of calculation steps, diffusion models also suffer from slow sampling speed and weak generalization ability for data types. Therefore, domestic and foreign researchers have proposed some new methods for improvement, and two derivative models are introduced below.

(1) Combination of DM and CLIP[22][23]. The advantage of the CLIP model is that it is based on multimodal contrastive learning and pre training, which can align text and image features, so there is no need to annotate data in advance, making it perform well in zero sample image text classification tasks; At the same time, the grasp of text description and image style is more accurate, and the ability to change unnecessary details of the image without changing accuracy, resulting in better performance in generating image diversity.

(2) Stable Diffusion Model (SDM)[24][25]. It is an AI drawing model that uses Gaussian noise for image generation. The Stable Diffusion model mainly uses three techniques: Gaussian noise generation, iterative diffusion algorithm, and deep learning model. Through interaction and collaboration, it generates images with realistic, detailed, and dynamic effects.

III. DISCUSSION AND OUTLOOK

3.1 Characteristics and improvement directions of various image generation technologies

3.1.1 Variational Autocoding (VAE)

Characteristics. The generation of images by VAE requires a decoupled continuous latent space, which is continuous and uniformly distributed. However, this may not necessarily be true in reality. The assumption of continuity in latent space may limit the model's ability to accurately model the distribution of real data, especially in the presence of discrete or complex data. At the same time, VAE trains the model by maximizing the probability lower bound, which leads to information loss in latent space. The lower bound of maximizing probability includes reconstruction error and divergence term. When the divergence term is too large, it will force the latent variables to follow a prior distribution and may also lose some useful information; VAE assumes that the latent variables follow a prior distribution, usually choosing Gaussian distribution or uniform distribution. However, when such assumptions may not be appropriate, leading to inaccurate modeling of data distribution and severe distortion of generated images, these are all shortcomings of VAE. But VAE also has its advantages: VAE can learn the potential distribution of input data, generate new sample images with a certain degree of continuity, provide a potential space based on VAE, interpolate and reconstruct samples, have strong interpretability, VAE training is stable, use variational inference and reparameterization techniques, and effectively learn model parameters by maximizing the lower bound.

Improvement direction. 1、 Introduce a priori of non-uniform distribution. In addition to traditional uniform or Gaussian prior distributions, some methods attempt to introduce the prior of non-uniform distributions, such as using more complex distributions (such as mixed distributions) or by learning the parameters of prior distributions. This can better model the distribution of data and provide more flexibility in the latent space; 2、 Improve potential space representation. In order to improve the continuity and representation ability of latent spaces, some methods introduce more complex latent distributions, such as using flow models or non parametric methods. These methods can better capture the complexity of data distribution and provide richer potential representations; 3 、 Improve the training process. Some improvement methods focus on improving the training process of VAE to alleviate the problem of information loss. This includes using more complex optimization algorithms or designing loss functions that are more suitable for data distribution. In addition, some methods combine the ideas of Adversarial Generative Networks (GANs) and VAEs to better balance reconstruction and sampling capabilities.

3.1.2 Generative Adversarial Network (GAN)

Characteristics. GAN uses two basic models, generator and discriminator, to achieve forward propagation and backward discrimination in adversarial games. However, the training process of GAN is relatively unstable, and the balance between generator and discriminator is easily disrupted, which may lead to pattern collapse or pattern collapse. GAN training requires a long time and a large amount of data sets, and is very sensitive to the selection and adjustment of hyperparameters, The images generated by GAN have some unreal or blurred details, especially in the generation of images with rich details, there is still room for improvement. These shortcomings of GAN. The advantage of GAN is that it can generate realistic images through the game process between the training generator and discriminator. The generator can learn to generate samples similar to real images. The generation process of GAN is unsupervised learning, which does not require labeling or category information on input data. Therefore, it is suitable for unsupervised learning tasks. GAN generates diverse image samples, which not only can generate real images, but also, It can also generate novel works of art.

Improvement direction. The stability and convergence of GAN are still hot topics for improvement. In terms of stability, researchers have proposed various improved GAN variants, such as WGAN, LSGAN, etc., to solve the problem of training instability; In terms of convergence, Conditional GAN (CGAN) is proposed to introduce conditional information, enabling the generator to control the generation of images of specific categories; There are still issues with blurring and insufficient details in the generated results of GAN. Improvement directions include increasing the number of layers in the GAN network or combining it with other image generation models, such as VAE.

3.1.3 Diffusion Model (DM)

Characteristics. DM generates images by adding random noise in the forward direction and then inferring noise reduction in the reverse direction. Therefore, its shortcomings mainly include the following: high computational complexity, the training and inference process of diffusion models is usually complex and time-consuming, especially when dealing with large-scale datasets; It is difficult to tune parameters, and tuning the parameters of diffusion models is relatively difficult. It is necessary to carefully select appropriate parameters such as learning rate, step size, and iteration number; Affected by the accumulation of noise, the diffusion model generates samples by iteratively diffusing noise, and the accumulation of noise may lead to a decrease in the quality of generated samples. However, the advantages of DM are also very prominent, with high image quality generated and excellent performance of diffusion models in generating samples, which can generate realistic and high-quality images or data samples; The generation process of images is flexible, and the generation process of diffusion models is controllable. The quality and clarity of generated images can be controlled by adjusting the diffusion steps of noise; Preserve global consistency. The diffusion model maintains global consistency during the generation process, that is, the generated samples remain consistent as a whole, and there will be no fragmented or locally inconsistent situations; Can model potential image distributions. The diffusion model can model the potential structure of data by inferring the potential distribution from the data sample through reverse inference.

Improvement direction. 1、 Accelerate and optimize the training process. Regarding the computational complexity of diffusion models, some acceleration and optimization methods can be adopted, such as parallel computing, initialization strategies for model parameters, and improvements to optimization algorithms; 2、 Enhance the stability of the model. To improve the stability of the model, regularization techniques such as weight decay or batch normalization can be used to reduce instability and overfitting during the training process; 3、 Combine with other generative models. Combining diffusion models with other generative models, such as GAN, VAE, etc., to achieve better generative effects and model performance; 4、 Network architecture improvement. Improve the network architecture of the diffusion model, such as increasing the number of layers, introducing attention mechanisms, or adding residual connections, to enhance the model's representation ability and generation quality; 5、 Introduce a prior distribution. By introducing appropriate prior distributions, the modeling effect of diffusion models on data distribution can be improved, thereby enhancing the diversity and quality of generated samples.

3.2 Future Trends and Prospects of Image Generation Technology

The research on image generation technology is at its peak, with various image generation models emerging. Meanwhile, its shortcomings are also evident. In the future, potential breakthroughs in researching image generation technology include the following aspects:

3.2.1 Highly intelligent

Image generation technology will become more intelligent. Through learning and iteration, the image generation program will be able to generate more accurate and high-quality images, and automatically optimize various details and features of the images. This will enable image generation technology to be more widely applied in various fields.

3.2.2 Diversification and personalization

Image generation technology will become more diverse and personalized. With the continuous development of artificial intelligence technology, image generation programs will be able to generate various types of images, including people, landscapes, animals, buildings, etc. Meanwhile, the image generation program will be able to generate images with personalized features based on the user's personalized needs.

3.2.3 Real time and interactivity

Image generation technology will become more real-time and interactive. With the continuous development of internet and computer technology, people can access and use image generation technology through networks and mobile devices. Meanwhile, the image generation program will be able to generate corresponding images based on real-time input and interaction from users to meet their needs.

3.2.4 Integration with other technologies

Image generation technology will be more closely integrated with other technologies. For example, image generation technology can be combined with virtual reality technology to generate highly realistic virtual scenes and characters; Image generation can be combined with speech technology to generate corresponding images based on user voice commands; Image generation can be combined with blockchain technology to ensure the copyright and data security of images.

IV. CONCLUSION

With the rapid development of artificial intelligence, various deep learning based algorithms have been implemented in daily life and industry. Image generation technology, from Generative Adversarial Networks (GAN), Variational Auto-encoding (VAE) to the mainstream Diffusion Model (DM), has been extensively studied by researchers. The working principles and derivative models of the three image generation technologies have been analyzed, and the advantages, disadvantages, and improvement directions of the three image generation technologies have been discussed in depth, Further outlook on the future development trends of image generation technology.

REFERENCES

[1]. Pu Y C, Gan Z, Henao R, et al. Variational auto-encoder for deep learning of images labels and captions[EB/OL]. 2016-09-28[2017-11_13].https://arxiv.org/pdf/1609.08976.pdf .

[2]. Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[C]. Proc of the 27th international conference on Neural Information Processing Systems. Cambridge. MA: MIT Press. 2014:2672-2680.

[3]. YANG Guang-kai. Fingerprint Image Generation Method Based on Diffusion Models[J]. Journal of the Hebei Academy of Sciences.2023.02(40):13-19.

[4]. Zhou JiaLuo. Research on Image Generation Algorithm Based on Variational Autoencoder [D]. Hangzhou Dianzi University. 2021.

[5]. ZHAI Zhengli, LIANG Zhenming, ZHOU Wei, SUN Xia. Research Overview of Variational Auto-Encoders Models [J]. Computer Engineering and Applications,2019,55(3):1-9.

[6]. GENG Yaogang, MEI Hongyan, ZHANG Xing, LI Xiaohui. Review of Image Captioning Methods Based on Encoding-Decoding Technology [J].Journal of Frontiers of Computer Science and Technology. 2022.16(10):2234-2248.

[7]. YAN Jiayu , BEI Shizhi , ZHANG Le. Credit Card Fraud Detection Model Based on VAE-GAN Algorithm [J]. Journal of Beijing Electronic Science and Technology Institute.2022.12(20):70-81.

[8]. Wang Mingming. Clustering Algorithm Analysis Based on GAN and VAE [D].XIDIAN UNIVERSITY,2021

[9]. LI Dan, ZHANG Yu-An, HE Jie, CHEN Zhan-Qi1, SONG Wei-Fang, SONG Ren-De. Grade Evaluation Algorithm of Yak Based on VAE-CGAN [J]. Computer Systems & Applications. 2023,32(01):249-256.

[10]. Gao R, Hou XS, Qin J, et al. Zero-VAE-GAN: Generating Unseen Features for Generalized and Transductive Zero-shot Learning[C]. IEEE Transactions on Image Processing, 2020, 29:3665–3680.

[11]. Jin Lina, Yu Jiong, Du XuSheng, Wang Song. Generative adversarial network and variational auto-encoder based outlier detection [J/OL].Application Research of Computers, 2021, 39 (3): 774-779.

[12]. Cao Yangjie, Jia Lili, Chen Yongxia, Lin Nan , Li Xuexiang.Review of computer vision based on generative adversarial networks[J].Journal of Image and Graphics. 2018,23(10): 1433-1449.

[13]. GAN Yan , YE Mao , ZENG Fan-yu.Review of Research on Generative Adversarial Networks and its Application [J].Journal of Chinese Computer Systems2020.06(41):1133-1138.

[14]. ZOU Xiu-Fang, ZHU Ding-Ju. Review on Generative Adversarial Network[J]. Computer Systems & Applications, 2019,28(11): 1–9.

[15]. WEI Fuqiang, Gulanbaier Tuerhong, Mairidan Wushouer. Review of Research on Generative Adversarial Networks and Its Application[J].Computer Engineering and Applications.2021.57(19):18-31.

[16]. Radford A, Metz L, Chintala S. Unsupervised representa-tion learning with deep convolutional generative adversarial networks[EB/OL]. 2016-12-20. https://arxiv.org/pdf/1511. 06434.pdf.

[17]. Mirza M, Osindero S. Conditional generative adversarial nets [J]. https:// arXiv preprint

[18]. Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[EB/OL]. 2017-12-06. https: //arxiv.org /pdf /1701.07875.pdf.

[19]. Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of Wasserstein GANs ［EB/OL］.2017-12-25. https: / /arxiv.org /pdf /1704.00028.pdf.

[20]. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[C]. Advances in Neural Information Processing Systems. 2020,33:6840-6851.

[21]. Song J M, Meng C L, Ermon S. Denoising diffusion implicit models[J]. International Conference on Learning Repre-sentations, 2021:1-22.

[22]. Nichol A Q, Dhariwal P. Improved Denoising Diffusion Probabilistic Models [C]. Proceedings of Machine Learning Research. PMLR.2021:1-7.

[23]. Yang L, Zhang Z L, Song Y, et al. Diffusion Models: A Comprehensive Survey of Methods and Applications[J]. arXiv: 2209.00796v9,2022.

[24]. Zheng Kai, Wang Di. The Application of Artificial Intelligence in Image Generation——Taking Stable Diffusion and ERNIE ViLG as Examples[J]. Science & Technology Vision. 2022(35):50-54.

[25]. Zhenkang Lin, Yuyan Ma, Wei WangYu He, et al. A ho-mogeneous and mechanically stable artificial diffusion layer using rigid-flexible hybrid polymer for high-performance lithium metal batteries[J]. Journal of Energy Chemistry. 2023,76(01): 631–638.