Research on Artificial Intelligence Graphic Generation Technology

Youyi Zhang[a]
Media Institute, Hankou University. Wuhan, Hubei, 430212, China

Keywords: Generative Adversarial Network, Variational Autoencoder, Diffusion Model, Graphics Technique.

Abstract: Generative artificial intelligence has been applied to many scenarios in society. It is an innovative technology field, and its core principle is based on deep learning algorithms and neural network models. From the initial generation of more blurred and simple images, it is now possible to generate extremely complex and highly detailed images. The topic of this paper is the development and application of artificial intelligence graphics technology. This paper lists the development history of image generation and GAN, VAE and Diffusion image generation technology, but different algorithms have their own shortcomings. With the passage of time, the algorithm is constantly optimized and the processing power is also enhanced. Contemporary generative adversarial networks and variational autoencoders, for example, can be used in many aspects of image generation. In the later stage, this paper also analyzes the possibility of combining image generation with games. The development prospect of generative artificial intelligence is bright, and it will bring many conveniences and changes to people's life and work in the future, and promote the development of intelligence in various industries and different field.

1 INTRODUCTION

Since the birth of artificial intelligence, its functions have experienced leaps and breakthroughs again and again. From symbolic logic reasoning to perceptual cognition, to natural language processing, learning decision-making, and gradually realize autonomous action and creativity, giving personalized service and intelligent decision-making. It can realize the functions of language synthesis, text generation, image repair and generation, data prediction, intelligent recommendation and intelligent painting. In the present day, artificial intelligence has been applied to natural language processing, machine learning, computer vision, robotics, autonomous driving technology, medical care, agriculture, gaming and other fields.

This paper will introduce the development of graphics technology. From the early computer graphics and texture mapping phase (1960s-1980s) to statistical learning and traditional machine learning methods in the mid-2000s (1990s-2000s) to the present generation adversarial networks (GANs) and variational autoencoders (VAE) and Diffusion models Diffusion. The image generation techniques of GAN, VAE and Diffusion are also introduced.

During the 1960s and 1980s, key developments in early computer graphics and texture mapping included photorealistic graphics and texture mapping methods.

There are light reflection models in photorealistic graphics: Bouknight proposed the first light reflection model, Gouraud proposed the diffuse reflection model plus interpolation Gouraud shading treatment, Phong introduced the Phong illumination model. These models make computer graphics more realistic. Solid modeling technology: Since 1973, the University of Cambridge and the University of Rochester have developed solid modeling systems that allow users to build complex models from basic geometry.

Ray tracing and irradiance algorithm: Whitted proposed ray tracing algorithm to simulate the interaction between light and object surface; The radiosity method introduces multiple diffuse reflection effects to enhance the realism of the rendering (Yang, 2012).

However, early computer graphics and texture mapping were limited by limitations in computational

[a] https://orcid.org/0009-0006-2787-4615

Zhang, Y.
Research on Artificial Intelligence Graphic Generation Technology.
DOI: 10.5220/0013233800004558
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 1st International Conference on Modern Logistics and Supply Chain Management (MLSCM 2024)*, pages 79-85
ISBN: 978-989-758-738-2
Proceedings Copyright © 2025 by SCITEPRESS – Science and Technology Publications, Lda.

performance, rendering quality, texture mapping, and user interaction.

Texture mapping methods include projection mapping, which applying 2D textures to map 3D object surfaces through Projector and UV Mapping. Affine transform texture mapping and decal texture mapping are two classical texture mappings. The former scales, rotates, and distorts textures to increase dynamic variation. The latter combines multiple textures to create a rich visual effect.

During the 1990s and 2000s, Statistical learning and traditional machine learning methods are used in graphics techniques. Take the classic support vector machine (SVM) as an example. The development of SVM dates back to the early 1990s, when it was first proposed by Vladimir Vapnik and his colleagues at Bell LABS. The principle of SVM is to find a hyperplane, which is a dividing line that divides the sample data into two categories in two-dimensional space. The goal of SVM is to maximize the interval from the hyperplane to the support vector, and the larger the interval, the better the generalization ability of the model. Classification, regression and anomaly detection are the main applications of SVM (Gaur & Mohrut, 2019).

However, in the processing of rich information in complex images, large amounts of data, high label requirements, and in the processing of nonlinear relations, timing information and large-scale data. The effectiveness of statistical learning and traditional machine learning methods is limited.

In the second section, the principles of GAN, VAE and diffusion models are introduced. The third section describes the application of GAN, VAE and diffusion models in graphics technology and discusses the possibility of combining the graphics technology of artificial intelligence with medicine, games and other fields. The fourth section is the summary and future outlook.

2 OVERVIEW OF ARTIFICIAL INTELLIGENCE GRAPHICS TECHNOLOGY

2.1 GAN

In 2014, it was proposed by Goodfellow, and its full name is Generative Adversarial Network. GAN stands for GANs, and it is a type of deep learning model. Its basic principle involves at least two modules: a generator and a discriminator. Through the mutual game and learning between these two modules, the output is generated. The task of the generator is to generate as realistic fake data as possible in order to deceive the discriminator, it receives a random noise vector as input, and through a series of operations and transformations, it outputs a newly generated data sample. The task of the discriminator is to distinguish whether the input data is real or fake data generated by the generator. It receives both real data and data generated by the generator, and finally outputs its own decision, represented as the probability that the input data is real. In this process, the generator and discriminator engage in an adversarial training and learning process (Chai & Zhu, 2019). Figure 1 is a schematic diagram of GAN.

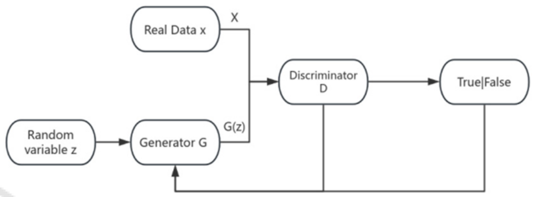

Figure 1: Schematic Diagram of GAN （Photo/Picture credit: Original).

The discriminator needs to continuously learn from both real data and fake data generated by the generator to improve its ability to distinguish between true and false data. The generator adjusts its parameters based on the feedback given by the discriminator to generate more realistic fake data, thereby increasing the probability of deceiving the discriminator. As the training and adversarial process progresses, the generator produces increasingly realistic data, and the discriminator's ability to identify true and false data also becomes stronger. When a certain level is reached, the data generated by the generator becomes so realistic that it is difficult for the discriminator to distinguish it from the real data. Through this adversarial learning process, GAN is able to learn the underlying distribution of the data and generate new data with similar characteristics. The optimization process of the generator (G) and discriminator (D) can be defined as a two-player game with a minimax problem.

$$min_D max_D \ \mathrm{V(D, G)} = E_{x\sim p\ data(x)}\ [lgD(x)] + E_{z\sim p\ z(z)}\ [\lg\ (D((G(z)))] \quad (1)$$

The advantages of GANs in generating images are evident, as they can produce realistic and creative images. However, they also have drawbacks. The training process is unstable, making it difficult to control the quality. There is a risk of mode collapse, limiting the diversity of the generated images. Additionally, GANs can potentially be misused for malicious purposes. Overall, while GANs possess powerful capabilities, their shortcomings must be addressed and improved upon in applications.

2.2 VAE

VAE were proposed by Kingma and Welling. VAE is a generative model which is mainly used to learn the potential variable representation of data and realize the generation and reconstruction of data (Nakai & Shibuya, 2022). VAE are shown in Figure 2.

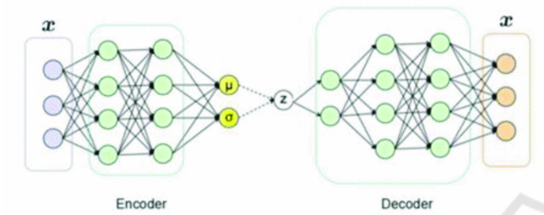

Figure2: VAE (Nakai & Shibuya, 2022).

VAE is a model that combines variational Bayes methods and deep learning. VAE represents image x through latent variable z and plays a role in image generation and feature learning. The aim of the model is to maximize the log-likelihood log $p_\theta(x)$ of the image x and compute the approximate posterior distribution $\mathcal{L}(q_\phi(z|x))$ by variational inference.

Specifically, variational inference is achieved by optimizing the lower bound $\mathcal{L}(q_\phi(z|x))$. Including constraint on the KL divergence between $\mathcal{L}(q_\phi(z|x))$ and the prior distribution $p(z)$. The lower bound consists of two parts. On the one hand, it penalizes the difference from the prior distribution, and on the other, it measures the quality of the image reconstruction for a given potential variable z by expecting $E_{q\phi(z|x)}[log p_\theta(x|z)]$.

In the model, the potential variable z is assumed to follow a normal distribution.

$$q_\phi(z_j|x^i) \approx N\left(\mu_j^i, \sigma_j^i\right) \quad (2)$$

$$p(z) \approx N\,(0, I) \quad (3)$$

μ_j^i and σ_j^i are the mean and standard deviation of the j potential variable given the graph x^i. Through deep learning techniques, the model optimizes the parameters θ and ϕ, ultimately enabling efficient representation of latent variables and high-quality image generation.

VAE can generate new samples similar to training data and extract high-dimensional data features. It is easy to manipulate latent space, but the resulting images are often fuzzy and lacking in detail, can suffer from pattern crashes, are sensitive to hyperparameters, and are expensive to train.

2.3 Diffusion

Traditional models GANs learn image features by creating two neural networks and pitting them against each other. Gans also need to train generators and discriminators. While tuning the loss function for Gans is simple, the learning dynamics (including trade-offs between generators and discriminators) are difficult to follow, as are the problems of gradient disappearance and pattern collapse (when there is no diversity in the generated samples) (Liang, Wei, & Jiang, 2020). As a result, training GAN models becomes very difficult. Diffusion models are easier to train than Gans to produce diverse and complex images. The diffusion model can solve the problem of GAN training convergence because it is based on the same training data set. This is because it is equivalent to presupposing the result of neural network convergence, thus ensuring the convergence of training (DHARIWAL, 2021), and can predict the images generated in the middle during training and give labels in order to classify the images. Next, the gradient is calculated using the cross entropy loss between the classification score and the target class by using the gradient to guide the generation of samples. In addition, after the introduction of conditional control diffusion module, the diffusion model can use heavy control to output a specific style of image.However, the current diffusion model algorithm also lacks certain theoretical guarantee, high computation cost and large memory occupation.

3 GAN, VAE, DIFFUSION IMAGE TECHNIQUES AND VARIATIONS

3.1 GAN

Based on the principles of GAN, it can learn the features and patterns of images as well as the underlying data distributions. Based on the initial version of GAN, many variations and improvements have emerged, such as DCGAN, BiGAN, CycleGAN, and many others. The following is an explanation of these different GANs

3.1.1 DCGAN

Deep Convolutional Generative Adversarial Networks (DCGAN) incorporates convolutional neural networks into the GAN model. It modifies the convolutional neural network architecture to improve the quality of the generated samples and the speed of convergence. DCGAN possesses better capabilities for generating images. DCGAN is able to generate higher-quality images and related models, which to some extent addresses the previous issue of instability in GAN training (Liu & Zhao & Ye,2023). This is because convolutional neural networks have a powerful ability to process images. Today, DCGAN has become a fundamental model in the field of image generation. The generation effect of DCGAN is demonstrated in Figure 3.

Figure 3: Generation Effect of DCGAN (Liu & Zhao &Ye, 2023).

3.1.2 BigGAN

BigGAN is the first to generate images with high fidelity and low intra-class diversity gap. It incorporates the concept of batch size into its training process. Unlike traditional GANs, BigGAN enhances the number of convolutional channels and grid parameters. Additionally, it incorporates truncation tricks and functions to control model stability. In addition, there are also variations like BiGAN and BigBiGAN with their own unique generation effects. Figure 4 shows the generation effect of BigGAN.

Figure 4: Generation Effect of BigGAN (Liu & Zhao & Ye,2023).

3.1.3 CycleGAN

CycleGAN is primarily applied in the field of domain transfer. This refers to the process of transferring data from one domain to another domain (DHARIWAL, P, 2021). Its core idea is: Assuming there are domains X and Y, map Y domain to X domain, and vice versa, creating a cyclic process. Its core principle is illustrated in Figure 5.

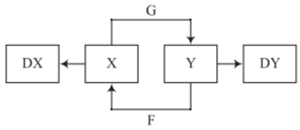

Figure 5: Core Principle of CycleGAN (Liu & Zhao & Ye,2023)

CycleGAN can learn the latent distribution characteristics of similar things and create a mutual transformation between two domains. The premise is that there is some commonality between the things in these two domains, as demonstrated in Figure 6 with the transformation from a zebra to a horse, and from summer to winter.

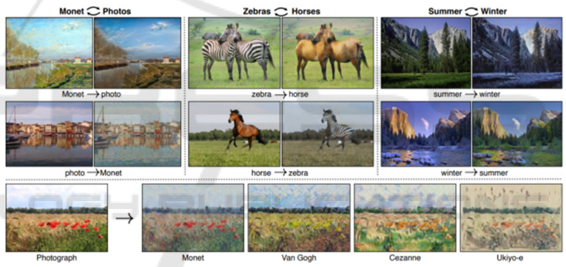

Figure 6: Case Demonstration of CycleGAN (Liu&Zhao &Ye, 2023).

3.2 VAE

The production of high-quality samples, the potential space for interpretation, the ability to process missing data, the smooth potential space, and the wide range of adaptability and model flexibility all make VAE an excellent performance in graphics technology.

3.2.1 Model of VAE-GAN

Xu and abdelouahed explored the reconstruction of multi-spectral images (MSI) from RGB images in their research (Yang et al., 2019). MSI is used in a wide range of applications, including satellite remote sensing, medical imaging, weather forecasting and the interpretation of artworks. In this study, VAE is combined with GAN. Specifically, the method replaces the traditional autoencoder for VAE and adds an L1 regulator.

Two classical datasets CAVE and ICVL were used in the study. Four quantitative indicators: root mean square. Square error (RMSE), normal or relative root mean square. Square error (nRMSE or rRMSE), peak signal-to-noise ratio. Ratio (PSNR) and Structural Similarity Index (SSIM). These four quantitative indicators are applied to the comparison of VAE-GANs with CNNs and cGANs. Table1 shows the comparison of VAE-GANs with CNNs and cGANs.

As can be seen from Table 1, when rebuilding RGB from MSI, RMSE is reduced by 66%. This shows that the method is more effective in capturing and reconstructing image features, and can recover image information more accurately.

Table 1: Comparison of VAE-GANs with CNNs and cGANs.

Metrics	**Berk**	**Kin**	**Ours**
Approach	CNNs	cGANs	VAE-GAN
Ratio of training and testing	N/A	50%:50%	50%:50%
RMSE~ (0-255)	2.55	5.649	1.943
RMSE~ (0-1)	0.038	N/A	0.0076
PSNR	28.78	N/A	42.96
SSIM	0.94	N/A	0.99

3.2.2 VAE with Priori Segmentation

Nakagawa and Haseyama et al. combined VAE with image segmentation prior in their research (Nakagawa et al., 2021). VAE is an independent potential variable z_{FG} and z_{BG} for foreground and background region learning respectively. Model is shown in the figure7. Table 2 shows the Estimation Error under different potential variables.

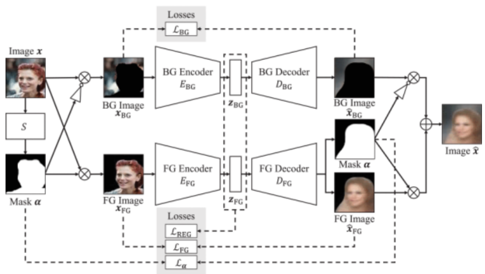

Figure 7: VAE-based model by splitting an image into several disjoint regions (Nakagawa et al., 2021).

Table 2: Estimation Error of latent variable (The direction of the arrow indicates whether the estimate error is better to increase or decrease).

Model: Ours ($\beta = 100$)	
Input	Estimation Error
None	19.48%
Z	17.40%$\pm$0.037%
z_{FG}	17.40%$\pm$0.040%
z_{BG}	19.48%$\pm$0.031%

As can be seen from the data of z_{BG} and z_{FG} in Table 2, the study of Nakagawa and Haseyama verified the disentanglement and transferability of VAE.

The above two studies demonstrate the effectiveness, effectiveness, separability and transferability of VAE.

3.3 Diffusion Model

With the rapid development of diffusion models, their potential in text image generation has increased significantly. According to the architecture, the researchers propose a text image generation method based on diffusion model. These methods include cascaded based diffusion models, unCLIP priori based diffusion models, discrete space based diffusion models and potential space based diffusion models.

The cascaded based diffusion model mainly uses guiding strategies to create high-resolution images. First, text conditions are entered into the diffusion model to capture the overall structure of the image content.

After that, a high-resolution image is generated by an upsampled diffusion model to improve detail and maintain authenticity and variety. For example, the GLIDE model proposed by Nichol et al. (NICHOL et al., 2022) adopts the cascade diffusion method. First, Transformer is used to encode the text, and the encoded text embedding replaces the class embedding in the ADM model to transform the rough image with 64×64 resolution. Then, an upsampling diffusion model is trained to improve the image to 256 ×256 high resolution, and the image details are refined. In addition, the implicit classifier guidance strategy used in the training process can ensure the diversity and fidelity of images while supporting flexible text prompt generation. However, the generation of complex prompts still faces challenges (GAO, Du, & Son, 2024). The GLIDE model proposed by Nichol et al. adopts the cascade diffusion method (NICHOL et al., 2022). Firstly, Transformer is used to encode the text and the encoded text embedding replaces the class embedding in the ADM model. The upsampled diffusion model is then trained to 256 × 256 high resolution to refine the image details. In the training process, the implicit classifier guidance strategy can ensure the diversity and fidelity of images, and support flexible text prompt generation. However, generating complex text prompts remains a challenge. As shown in Figure 8.

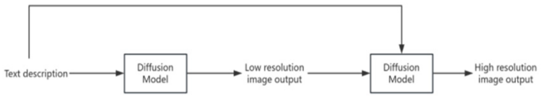

Figure 8: Cascade based diffusion model （Photo/Picture credit: Original).

3.4 Discussion

GAN, VAE and diffusion models have their own advantages. In terms of generation quality, GAN and diffusion models perform better. In terms of stability, VAE and diffusion models are better. VAE is simpler in terms of complexity. In terms of speed, VAE and GAN are generated more quickly. According to the different characteristics of GAN, VAE and diffusion models, they have different effects when combined with other algorithmic models. As mentioned in the third section of the graphics technology, DCGAN is mainly used to generate high-quality images. BigGAN focuses on generating high-resolution and diverse images. CycleGAN is used for unsupervised image style transformation. VAE-GAN works well in situations where learning is required. The combination of VAE and image segmentation prior has better disentanglement. The cascade based diffusion model can ensure the diversity and fidelity of images and support flexible text prompt generation.

All three models also have advantages in areas such as gaming and healthcare. GAN generates high-quality images that enhance the gaming experience and improve the accuracy of medical diagnoses. VAE has the flexibility to provide games with different styles of design and to generate medical images in different pathological states. The outstanding performance of diffusion model in generating quality and detail recovery is conducive to the de-noising and recovery of realistic game publicity images and medical images. Artificial intelligence graphics technology is not only used in the field of games and medicine, but also has good applications in agriculture, meteorology, architecture and other fields. With the continuous progress of modern social science and technology, the fields related to images, graphics and vision will provide sufficient space for the development of artificial intelligence graphics technology. It is not only limited to the improvement of the quality of production content, but also to the addition of advanced functions such as artificial intelligence-related automation and interaction.

4 CONCLUSIONS

This article introduces the history of artificial intelligence in the development of graphics technology, it introduces the meanings of GAN, VAE, and DIFFUSION technologies, along with some application scenarios and extended discussions and reflections. The future demand for image accuracy continues to increase, and generation is a rapidly developing area in artificial intelligence. Based on deep learning algorithms and neural network models, people can be assisted in generating higher-quality images, significantly enhancing the overall experience. Furthermore, through continuous learning, AI and large models can produce a diverse range of image types. Although there were certain issues with early generative models, over time, VAE and GAN have leveraged the concept of game theory to drive the development of the entire field of adversarial artificial intelligence. Currently, they are mainly applied in areas such as technology integration, innovation model optimization and improvement, application domain expansion, personalized and customized services, as well as security and compliance. While they offer numerous benefits, they also spark additional considerations and reflections, for instance, addressing resource issues can be constrained by limitations in datasets, a more comprehensive system is needed to enrich the models, reducing instability and confrontation are some of the securities or deepfake-related social issues that can arise from AI's involvement in image generation. Appropriate laws and regulations are also needed to impose constraints. In conclusion, generative artificial intelligence has a bright development prospect. In the future, it will bring many conveniences and changes to people's lives and work, driving the intelligent development of various industries and different fields.

REFERENCES

Yang, S., 2012. Research and Implementation of Illumination Model in Photorealistic Graphics Technology. Xidian University.

Gaur, K. & Mohrut, P. 2019. A review on Hyperspectral Image Classification using SVM combined with Guided, Filter 2019 International Conference on

Intelligent Sustainable Systems (ICISS), Palladam, India, pp. 291-294
Chai, M. T., & Zhu Y. P., 2019. Research and Application Progress of Generative Adversarial Networks
Wang M. Q., Yuan W. W., & Zhang, J. 2021 A Review of Research on Generative Adversarial Networks (GANs). Computer Engineering and Design, 42(12): 3389-3395.
Nakai, M. & Shibuya, T. 2022. Efficiency of Reinforcement Learning using Polarized Regime by Variational Autoencoder, 2022 61st Annual Conference of the Society of Instrument and Control Engineers (SICE), Kumamoto, Japan, pp. 128-134.
Liang, J. J., Wei, J. J., & Jiang, Z. F., 2020. A review of generative adversarial networks Exploration of Computer Science and Technology.
DHARIWAL, P., 2021. NICHOL A.Diffusion models beat GANs on image synthesis.
Liu H. D., Zhao X. L., & Ye H. P., 2023. A Review of GAN Model Research. Internet of Things Technologies, 13(01): 91-94.
Yang, G. Lu, Z. Yang, J. & Wang, Y. An Adaptive Contourlet HMM–PCNN Model of Sparse Representation for Image Denoising, in IEEE Access, vol. 7, pp. 88243-88253, 2019.
Nakagawa, N. Togo, R. Ogawa, T. & Haseyama, M., 2021. Disentangled Representation Learning in Real-World Image Datasets via Image Segmentation Prior, in IEEE Access, vol. 9, pp. 110880-110888.
Nichol, A. Q., Dhariwal, P., Ramesh, A., et al., 2022. GLIDE: towards photorealistic image generation and editing with text-guided diffusion models, Proceedings of the International Conference on Machine Learning.
Gao, X., Y., Du, F., Song, L., & J., 2024. A review of comparative research on text image generation based on diffusion model. Computer Engineering and Applications, 1-23.

SCIENCE AND TECHNOLOGY PUBLICATIONS