Exploring Deep Learning-Based Generative Image Techniques: Methods and Applications

Yi Huang

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China
YiHuang_Essen@njust.edu.cn

Abstract. The pervasive integration of deep learning within the realm of image generation has catalyzed profound advancements and breakthroughs in this technology. The advent of emblematic models such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) has positioned image generation at the forefront of discussions in computer vision and artificial intelligence. This paper delves into these three quintessential deep learning-based image generation models: GANs, VAEs, and Autoregressive Models (ARMs). It offers an in-depth examination of their methodologies, recent enhancements, and the trajectories of their development, aiming to elucidate the current landscape of image generation technologies and the practical challenges they encounter. Furthermore, the paper projects future trends and potential avenues for research in image generation, spotlighting emergent areas of scholarly interest. By presenting a comprehensive review of extant image generation technologies, this manuscript seeks to furnish invaluable insights and resources for researchers in allied domains, thereby fostering the further evolution and utilization of image generation technology.

Keywords: Generative models; Generative Adversarial Networks; Autoregressive Models.

1 Introduction

The exponential growth of digital technologies has paved the way for significant advances in the field of image generation, a domain that is witnessing profound transformations due to the development of sophisticated generative models. Among these, Generative Adversarial Networks, Variational Autoencoders, and Autoregressive Models have emerged as fundamental architectures that enable a myriad of applications ranging from artistic image synthesis to medical imaging and beyond. Each of these models leverages unique mechanisms to produce images, pushing the boundaries of what artificial intelligence can achieve in terms of realism and accuracy. GANs, for instance, employ adversarial processes to refine image quality, while VAEs use a probabilistic approach to generate new data points from learned distributions. ARMs, on the other hand, predict future outputs based on past data, which is invaluable for time-sensitive applications. This paper aims to delve into the operational principles, improvements, and applications of these models, highlighting how they have evolved to address specific challenges such as image

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_52

quality and model stability. Through this exploration, the paper seeks to underscore the profound impact these technologies are having on various industries and to outline potential future directions in the field of image generation.

2 Relevant Theories and Typical Methods

Below is an overview of the basic network architectures adopted by typical methods in the field of image generation, including the architectures of GANs, VAEs, and ARMs [1].

2.1 Basic Principles of Generative Adversarial Networks (GANs)

GANs were first proposed by Ian Goodfellow et al. in the paper titled "Generative Adversarial Nets". This model consists of two sub-networks: the Generator and the Discriminator [2]. The algorithm can be summarized as alternating training between the two networks. In the first step, the Generator is fixed, and the Discriminator is trained. In the second step, the Discriminator is fixed, and the Generator is trained. The two sub-networks are in a continual adversarial state. As depicted in Fig.1, the Generator progressively transforms random variables into images resembling real ones. Meanwhile, the Discriminator distinguishes between the received images to determine whether they are generated or real, driving the Generator to produce increasingly realistic images. The ultimate goal of the GAN model is to continually train and optimize both the Discriminator and the Generator.

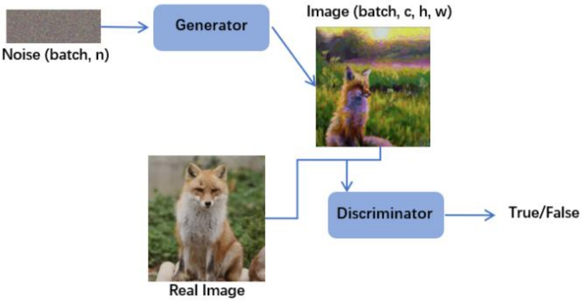

Fig. 1 GAN network structure (Photo credit: Original).

The Generator loss function L_G and the Discriminator loss function L_D are specifically defined as (Cao et al., 2024): $\mathrm{L}_G = -E_{\hat{x} \sim P_G}[\log_2(D(\hat{x}))]$ and

$$\mathrm{L}_D = -\frac{1}{2} E_{x \sim P_{data}}[\log_2(D(x))] - \frac{1}{2} E_{\hat{x} \sim P_G}[\log_2(D(\hat{x}))].$$

Where ‘ $\hat{\mathrm{x}}$ ’represents generated images, ‘ x ’represents real images, ‘ E ’denotes the expectation operation, ‘ P_G ’represents the probability distribution of features in generated images, and ‘ P_{data} ’stands for the probability distribution of real image features.

Innovations and Applications of Generative Adversarial Networks. The mainstream improvements of these models can be summarized into the following two points: 1. Optimizing the model by combining with other models; 2. Improving methods to address the model's shortcomings. This also applies to the next two typical models that will be introduced. For GAN, its drawbacks are quite prominent. Due to the method of using two independently trained models to engage in adversarial training, the training of GAN models is highly unstable, often leading to issues such as model collapse and vanishing gradients. Additionally, ensuring the balance and synchronization between the two models poses significant challenges, resulting in the poor controllability of GAN models. Therefore, there have been many improvements to GANs targeting the aforementioned shortcomings [3].

For example, to address the instability in GAN training, the paper "Wasserstein GAN" proposed mitigating mode collapse by minimizing the Wasserstein distance instead of the original GAN's Jensen-Shannon divergence [4]. Another example is the challenge of image generation requiring the model to generate images as closely as possible to the description provided, So the Conditional Generative Adversarial Networks (Conditional GANs) emerged as one of the solutions. The earliest CGAN model adopted the GAN-INT-CLS method, introducing text descriptions into the Generator and the Discriminator [5]. Meanwhile, some researchers have also proposed new architectures and training techniques to address the instability and mode collapse issues in GAN training, for instance, optimizing the interaction between the Generator and the Discriminator [6]. Another challenge in image generation technology is how to generate images of the highest possible quality. Progressive Growing GANs is one of the methods specifically designed to tackle this challenge. By gradually increasing the image resolution and the depth of the network, PGGAN excels in generating high-resolution and high-quality images [7].

As research on Generative Adversarial Networks continues to deepen, many application areas have emerged. Because GANs themselves have very powerful distribution learning capabilities, in addition to image generation tasks, GANs are also widely used in tasks such as data generation, image super-resolution, image restoration, image style transfer, and cross-modal image generation. These new application areas provide new opportunities and challenges for the development of GANs, promoting the continuous improvement and innovation of GANs. In addition, there are some typical improvement methods, some of which are listed below but not elaborated on: Deep Convolutional Generative Adversarial Networks (DCGANs), Cycle Generative Adversarial Networks (CycleGANs) and Large-Scale Generative Adversarial Networks (BigGANs), among others [8].

2.2 Operational Mechanisms of Variational Autoencoders (VAEs)

Auto-Encoders (AE) models, as a form of self-supervised learning, are predominantly applied in reducing data dimensionality, classifying images, detecting objects, and removing noise from images. Additionally, AE models possess the capability to generate data samples resembling the training data, rendering them suitable for data augmentation and unsupervised neural network pre-training. An autoencoder model is composed of two primary components:

Encoder: Acquires the latent characteristics of the input data, condensing the data into a representation within a latent space. Decoder: Reconstructs the original input data from the learned low-dimensional features [9].

VAE, a typical deep learning generative model, were proposed by Kingma et al. in 2014, rooted in Variational Bayes inference. As depicted in Fig.2, VAE employ two neural networks to construct two models of probability density distributions: one for conducting variational inference on the original input data, generating the probability distribution from variational inference of latent variables, referred to as the inference network; the other for reconstructing the approximate probability distribution of reconstructing the original data from the variational probability distribution of latent variables, referred to as the generative network [10].

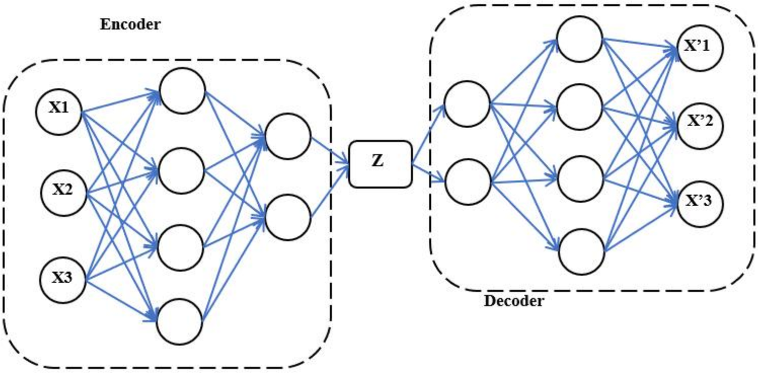

Fig. 2 VAE network structure (Photo credit: Original).

Consider an original dataset where each data sample X_i is randomly drawn from mutually independent continuous or discrete distribution variables, forming the dataset $X' = \{x_i'\}_{i=1}^{N}$. It is posited that this process yields latent variables Z, indicating that some variables are not directly visible. The observable variable X is a random vector within a high-dimensional space, whereas the latent variable Z exists as a random vector in a comparatively lower-dimensional space. This model can be delineated into two distinct processes: (1) the approximate inference of the posterior distribution of the latent variable Z, represented by $q_\phi(z|x)$, which serves as the inference network; the process generating the conditional distribution of the variable X', represented by $P_\theta(z) \cdot P_\theta(x'|z)$, known as the generation network [11].

VAE is named due to its structural resemblance to AE. However, its operational principles diverge significantly from those of AE. In contrast to AE, the Encoder and Decoder outputs in VAE are probability density distributions defined by parameters, instead of fixed encodings.

Novel Research in Variational Autoencoders. VAEs can be seen as a hybrid of neural networks and Bayesian networks. Its greatest advantage lies in its ability to learn a low-dimensional representation of data distribution by maximizing the marginal likelihood of the data, explicitly modeling the probabilistic relationship between observed data and latent variables. This holds true for both continuous and discrete variables, enhancing interpretability. Additionally, VAE avoids the complex Markov chain sampling process through parameter transformation, resulting in more stable training compared to GAN.

However, VAE generates images by sampling from the latent space rather than directly copying input data. This randomness often leads to ambiguous and fuzzy representations, causing unstable output quality and having poor expressive power for complex models. That is the culprit behind VAE's tendency to produce ambiguous images. Moreover, it incurs significant training costs as it requires multiple iterations [12].

To address this issue, many researchers have also improved VAEs. As illustrated in Table 1, various VAEs are capable of producing distinct image samples tailored to specific task requirements, significantly enhancing the quality of the generated outputs. Table 1 lists only a few improvements and variants.

Table 1. Improvements/Variants of VAE

Improvements/Variants of VAE	Abbreviation	Year
Conditional Variational Autoencoders	CVAE	2015
Variational Fair Autoencoder	VFAE	2015
Importance-Weighted Autoencoders	IWAE	2015
Conditional Variational Autoencoders with GAN	CVAE-GAN	2017
Variational Lossy Autoencoders	VLAE	2017
Channel-Recurrent Variational Autoencoders	CRVAE	2017
Least Square Variational Bayesian Autoencoders	LSVAE	2017
Information Maximizing Variational Autoencoders	IMVAE	2017
Multi-Stage Variational Auto-Encoders	MSVAE	2017
Nonparametric Variational Autoencoders	NpVAE	2017
Memory-enhanced Variational Autoencoders	MeVAE	2017
Fisher Autoencoders	FAE	2018

As mentioned earlier in the article, most improvements to models revolve around optimizing the methods within the model to address its shortcomings or combining the advantages of other models to enhance the model.

The conventional VAE is an unsupervised model capable of producing output data resembling the input. But it cannot control the directed generation of specific category sample data. To enable the VAE model, which performs image generation tasks, to generate images that match descriptions as closely as possible, conditional information was first introduced to constrain the model's generation. By adding category information labels to the input of the encoder, Conditional VAE (CVAE) controls the generation of samples for specified categories. Therefore, CVAE has also transformed from a traditional unsupervised mode to a semi-supervised mode.

Improvements like CVAE-GAN combine other models. Considering that images produced by CVAE tend to appear blurry, GAN can use its adversarial nature to maintain the fidelity of the generated images. By adding the GAN's Discriminator after the CVAE's Decoder, it ensures that the images generated by CVAE are of high quality [13].

The Cyclic Channel Variational Autoencoder (CRVAE) is a VAE variant proposed by Shang et al. in 2017. CRVAE integrates Convolutional VAE (cVAE), Long-Short Term Memory (LSTM), and Generative Adversarial Networks (GAN), achieving a multi-channel cyclic interconnected network structure. It overcomes many drawbacks of traditional VAEs, such as generating blurry images, poor representation of complex structures, and unsuitability for sequential model applications, and shows good performance in image generation and data reconstruction.

It is worth mentioning that, although the methods listed in Table 1 are not from recent years, they are typical improvement methods. The cutting-edge improvements of the three typical models introduced in this paper are all aimed at their respective task goals, combining one or more variants or improvements to achieve better results. For example, Sina et al. proposed a method for synthesizing cardiac MR images for the medical field, which also combines GAN and VAE. Most new research improvements and variants are inseparable from these typical models and typical improvement methods [14]. Therefore, this paper lists some of the new improvements but does not provide a detailed introduction.

2.3 Application Framework of Autoregressive Models (ARMs)

Autoregressive (AR) models are statistical and time series models used for analyzing and forecasting based on the previous values of data points. These models are widely employed in various fields, including economics, finance, signal processing, natural language processing, and the image generation field being discussed in this paper, among others.

At the core of autoregressive modeling is the AR(p) model, with "p" denoting the model's order. And the present value of a variable is depicted as a linear sum of its preceding "p" values along with a white noise error term. The general formula for the AR(p) model can be expressed as follows:

$$X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + \varepsilon_t \tag{1}$$

X_t represents the value of a time series at time t, which is the value to be predicted or understood.

c is the constant term, sometimes included to account for nonzero mean.

$\varphi_1, \varphi_2, \varphi_3, \ldots, \varphi_p$ are autoregressive coefficients representing the weights assigned to previous values. These coefficients determine the strength and direction of influence of past values on the current value.

ε_t is the error term, typically assumed to be white noise, representing the variance or randomness that cannot be explained at time t. Autoregressive models require the selection of an appropriate order (p) and estimation of suitable autoregressive coefficients (phi) to assess the goodness of fit of the model. Different orders and estimation methods may yield different results. In the field of image processing and generation, autoregressive models can also be utilized. Due to the correlations between image pixels, autoregressive models can be employed for pointwise prediction and computation by considering a two-dimensional image as one-dimensional sequential data. In image generation, autoregressive models are often combined with other models to model the dependencies between image pixels. One of the most typical models is PixelRNN and PixelCNN [15]. Fig.3 illustrates a simple schematic example, using PixelCNN proposed by Van den Oord et al.

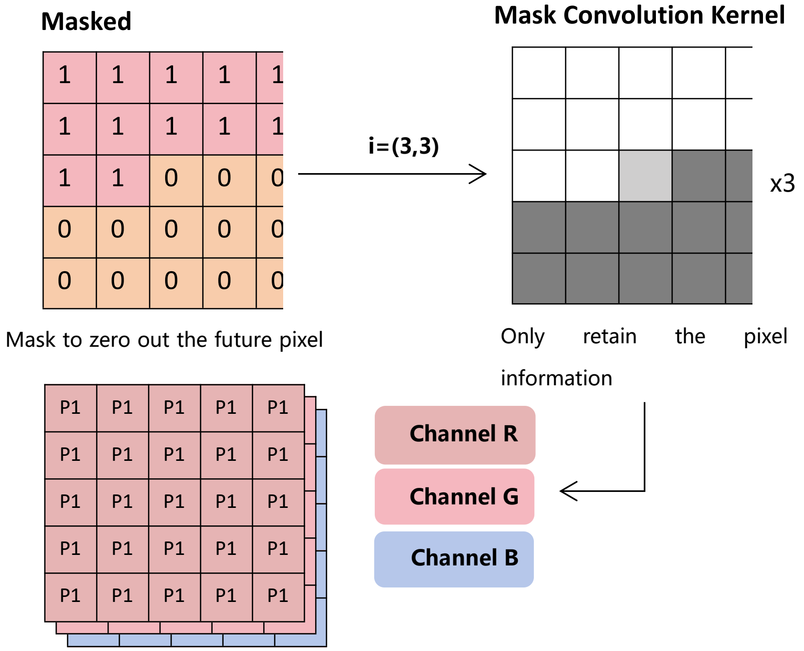

Fig. 3 PixelCNN network structure (Photo credit: Original).

Standard convolutional layers extract information from all pixels simultaneously. Therefore, in PixelCNN, the authors relied on traditional CNN techniques and introduced Masked Convolutions while eliminating pooling layers to model the data. They stacked multiple convolutional layers to predict the value of each pixel in the image pixel-wise. During the generation process, the prediction of each pixel considers the previously generated pixels [16].

Recent Developments in Autoregressive Models. The strength of Autoregressive Models (ARMs) lies in their versatility with input data, making them ideal for a range of sequence generation tasks such as text, speech synthesis, time series forecasting, and image generation. ARMs can adapt to generate videos with appropriate modifications due to their sequential data handling capabilities. Additionally, the incremental nature of their generation process, where each output depends on the preceding one, allows for greater interpretability and controllability [17]. This structured model approach also facilitates specific enhancements and the integration of constraints to remedy shortcomings.

However, these benefits come with limitations. The dependency on previous outputs can lead to information loss or error accumulation over long sequences, and the focus on local information might affect the overall coherence and consistency of the generated images. Moreover, the sequential computation increases the model’s computational complexity exponentially with the sequence length, which can be resource-intensive, especially for large-scale and high-resolution tasks. To mitigate these drawbacks, ARMs have been combined with other models such as GANs and VAEs, and enhanced with optimization techniques like multi-channel methods, attention mechanisms, and regularization. For example, AR-GAN integrates ARM with GAN, enhancing the generative capabilities, while AR-VAE merges ARM with VAE to balance generative efficiency and variability [18].

Additionally, Sparse Vector Autoregressive Models have been developed to address the computational challenges associated with long sequences. By introducing sparsity, these models reduce the number of parameters and limit the scope of attention mechanisms, decreasing computational demands while preserving generation quality and improving generalization and interpretability. Recent advancements also highlight the use of autoregressive models in video generation, showcasing their potential as leading tools in this area due to their inherent flexibility and capacity for handling dynamic, sequential data.

3 Analysis of Technological Applications

3.1 Applications in Artistic Creation

The advancement of image generation techniques, empowered by deep learning, has revolutionized artistic creation and found widespread applications in various domains in recent years. Among these applications, image-related tasks stand out, including but not limited to image super-resolution, restoration, completion, and style transfer.

With the capability to produce highly realistic images, GANs and VAEs have emerged as essential tools and sources of inspiration for artists. GANs, by learning from a vast amount of artistic styles and contents, can generate new artworks with distinct styles, enabling artists to explore novel creative avenues. Moreover, they contribute to art restoration and repair by reconstructing damaged artworks based on known pixel information or removing redundant details.

Prominent AI painting software in recent years includes MidJourney, OpenAI's DALL-E, and DeepArt. Furthermore, image generation technologies have found applications in producing animations and special effects for movies.

In the realm of music, image generation techniques also shine, particularly in music visualization. Models can generate corresponding visual effects based on the rhythm and emotional tones of music, enhancing immersive experiences. Popular software applications include Magic Music Visuals, Spotify, and Suno.

Regarding video creation, generative models introduce an additional temporal dimension, allowing the generation of a series of continuous and correlated image frames, thereby producing coherent video sequences.

That above has sufficiently highlighted the multifunctionality and immense potential of image generation technology across various creative domains.

3.2 Applications in Virtual and Augmented Reality

Virtual Reality (VR) and Augmented Reality (AR) are also crucial application domains, but they cannot be solely achieved through image generation technology. Instead, they require integration with fields such as digital image processing and computer vision to achieve their full potential. Image generation models can assist in creating high-quality virtual environments and AR content, offering users a more realistic and immersive experience.

In virtual reality, generation models can be used to create realistic scenes and characters, such as scene expansion similar to image completion, scene generation, and generating scenes in 360-degree panoramic videos or images. For instance, due to the autonomous unsupervised learning and stochastic sampling advantages of generative adversarial networks, coupled with their powerful distribution learning capabilities, they can create highly complex virtual environmental scenes, even extending to three dimensions.

In augmented reality, image generation techniques can be used for generating and optimizing AR effects. By generating and superimposing virtual objects onto the real environment in real-time, AR generation can appear more realistic and natural.

3.3 Prospects in Other Fields

Image generation technology also has wide-ranging applications in fields such as healthcare and education. For instance, in medical training, it can generate realistic images of medical cases for more intuitive training, without requiring a large number of real case samples. For example, generating CT scans for interns or trainee doctors to practice diagnosing conditions. In the field of early childhood education, images are often more easily accepted by young children than text. However, high-quality image resources require significant manpower for creation, leading to issues such as a limited variety of learning resources and difficulty in creating personalized resources. Image generation technology can effectively address these challenges.

Furthermore, image generation models can provide a considerable quantity of samples for any object that requires images. For instance, one image generation model can generate a large number of images to provide samples for another image generation model. This can be applied to scenarios where a large number of samples are needed or samples are difficult to obtain, such as training models for facial recognition that require a vast amount of data, or providing interns with generated images of brain tumors and cardiac MRIs in medical training scenarios. By synthesizing a series of pseudo-pathological synthetic subjects with the desired corresponding features, this technology can effectively help train new medical personnel.

4 Challenges and Solutions

The development and application of artificial intelligence-related technologies inevitably raise concerns about privacy, security, and ethics among the public. In future development and application, these are unavoidable challenges.

4.1 Social Acceptance and Costs

Although image generation technology has many advantages in applications such as art, medicine, and education, it also poses significant challenges and triggers new conflicts in some areas. One of the most notable areas is its application in the field of art. With technological advancements, the use of artificial intelligence in the field of painting has become more common and prominent, reaching a peak in 2020 and 2021. AI-created artworks have garnered widespread attention, with some pieces even being sold at high prices at auctions, causing a sensation. On one hand, many young people admire the skill and creativity of AI-generated artworks, seeing it as a manifestation of technological progress and bringing new possibilities to the art world. On the other hand, practitioners in these fields are very concerned that the rise of AI will replace traditional professions, especially in the art field, fearing that the value of human creation will be diminished and will affect the ecological balance of the art industry. Additionally, with many individuals and groups using AI painting to generate images they need, questions have been raised about copyright issues and abuse of AI works. Who owns AI works—the AI itself, the operator, or the company or studio to which the AI belongs? Is the widespread appearance of AI works encroaching too much on the space for human works?

On the other hand, developing and deploying practical image generation technology models requires a large amount of resources, including hardware devices, human resources, and data storage. Even with abundant resources, there needs to be a balance between cost and performance. Currently, mainstream AI painting software belongs to some well-capitalized large companies, and monopolies and technological blockades are also potential challenges.

4.2 Issues of Data Privacy Protection and Security

The application of image generation technology relies on the collection and processing of data, so the second challenge is the public's privacy and security concerns. Firstly, while not all image generation models collect personal information data, it is inevitable in some generative tasks. Particularly, for targeted and personalized generation, such as when generating facial data, many real personal facial photos are needed. This data may include personal identity information or other sensitive information, so strict compliance with privacy regulations and best practices is necessary during data collection, storage, and processing. If this data is leaked, it may affect the privacy of users, such as identity theft if facial information is leaked.

4.3 Proposed Solutions

Firstly, it is essential to ensure that technology aligns with human values. Relevant departments and institutions should strengthen supervision, clarify the scope of data use and protection, and enhance measures to combat and penalize data privacy breaches. They should also formulate relevant policies and regulations to ensure safety and reliability. Developers and decision-makers need to increase transparency and actively promote public understanding and acceptance of these technologies, remembering the essence and purpose of technological development.

Secondly, awareness and understanding of data privacy and security should be enhanced. For models aimed at public use, especially those that require the collection of user personal information, technical means for data privacy protection should be strengthened. Many practical methods and technologies exist for data transmission, encryption, and protection. For example, decentralized data training methods allow model training without exposing raw data. Federated learning technology allows multiple participants to train models locally and only share model parameter updates, not raw data. Techniques like homomorphic encryption allow computations on encrypted data, thus protecting data privacy. Developers and decision-makers need to remain vigilant to ensure that users' privacy is adequately protected.

In summary, issues related to data privacy, security, social costs, copyrights, and ethics are all important considerations in the development and application of image generation technology models. It requires effective measures and cooperation from all sectors of society to ensure that these technologies bring more benefits to society while minimizing potential risks and negative impacts as much as possible.

5 Future Prospects

In the future, image generation technology is expected to advance in several key areas:

- Authenticity, Diversity, and Controllability of Outputs: Future developments aim to enhance user control over the generation processes, leading to outputs that are more diverse, realistic, and tailored to specific needs. This will expand the potential applications of generated content.
- Real-time Generation: Improvements in algorithm efficiency will enable real-time image generation, essential for interactive applications such as virtual reality games, where images and animations must adapt instantly to user actions and environmental changes.
- High-resolution Generation: There will be a focus on producing images with higher resolution and finer details, which is particularly critical in fields like medical imaging, where precision and clarity are paramount.

Additionally, as model improvements and variations evolve, new combinations of these models are likely to be explored.

A primary future direction for image generation technology will be cross-modal development, extending beyond traditional formats to include transformations like image-to-3D modeling, image-to-video, and image-to-scene generation. This approach will not only enhance virtual reality experiences but also advance technologies in autonomous driving by providing more intuitive interfaces and richer visual information. In summary, future research will likely concentrate on multimodal integration and cross-modal generation, merging information from various sources to enhance multimodal data generation and processing. Ongoing optimization of models and algorithms will improve the robustness and security of generative models, broadening their application across different fields and industries, and fostering cross-domain innovations.

6 Conclusion

This paper surveys deep learning-driven image generation techniques, focusing on three pivotal models:GANs, VAEs, and ARMs. It explores their foundational principles, recent advancements, and diverse applications. The rapid evolution in image generation has ushered in notable enhancements in content quality and diversity, propelling forward the domains of computer vision and artificial intelligence.

Image generation technologies demonstrate vast potential across various sectors including artistic creation, virtual and augmented realities, medical imaging, and autonomous driving. These applications are not only revolutionizing artistic endeavors and digital content creation but are also pivotal in technical fields requiring high fidelity visual simulations. Despite these advances, the deployment of image generation technologies encounters significant challenges including social acceptance, cost efficiency, and data privacy. To address these issues, researchers are refining algorithms, bolstering data security measures, and advocating for robust legal frameworks. Future research directions likely to shape the landscape of image generation include multimodal, real-time, and high-resolution image generation. Key focal areas include enhancing model controllability, boosting image realism and variety, and integrating novel architectural frameworks.

In conclusion, deep learning-based image generation has achieved significant milestones in both theoretical and practical aspects. As algorithmic and hardware developments continue to mature, broader and more effective applications of image generation are anticipated, enhancing user experiences and service quality across industries. This review aims to offer critical insights and serve as a reference for researchers, further catalyzing the progress of image generation technology.

References

1. Cao, Y., Qin, J., Ma, Q., Sun, H., Yan, K., Wang, L., Ren, J.: Survey of Text-to-Image Synthesis. Journal of Zhejiang University (Engineering Science) 58(2), 219–238 (2024). doi:10.3785/j.issn.1008-973X.2024.02.001.
2. Chen, S. A., Li, C. L., Lin, H. T.: A Unified View of cGANs with and without Classifiers. Advances in Neural Information Processing Systems 34, 27566–27579 (2021).
3. Chen, T. Q., Rubanova, Y., Bettencourt, J., Duvenaud, D.: Neural Ordinary Differential Equations. In: Advances in Neural Information Processing Systems, pp. 6571–6583. Springer (2018).
4. Davis, R. A., Zang, P., Zheng, T.: Sparse Vector Autoregressive Modeling. Journal of Computational and Graphical Statistics 25(4), 1077–1096 (2016).
5. Gab Allah, A. M., Sarhan, A. M., Elshennawy, N. M.: Classification of Brain MRI Tumor Images Based on Deep Learning PGGAN Augmentation. Diagnostics 11(12), 2343 (2021).
6. Zhu, X., Huang, Y., Wang, X., Wang, R.: Emotion recognition based on brain-like multimodal hierarchical perception. Multimedia Tools and Applications 1–19 (2023).

7. Jiang, L., Zhang, C., Huang, M., Liu, C., Shi, J., Loy, C. C.: Tsit: A Simple and Versatile Framework for Image-to-Image Translation. In: Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III, pp. 206–222. Springer International Publishing (2020).
8. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive Growing of GANs for Improved Quality, Stability, and Variation. In: Proceedings of the 6th International Conference on Learning Representations (ICLR) (2018).
9. Kingma, D. P., Welling, M.: Auto-Encoding Variational Bayes. arXiv preprint arXiv:1312.6114 (2014).
10. Kolotouros, N., Pavlakos, G., Black, M. J., Daniilidis, K.: Learning to Reconstruct 3D Human Pose and Shape via Model-Fitting in the Loop. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2252–2261. IEEE (2019).
11. Zhu, X., Guo, C., Feng, H., Huang, Y., Feng, Y., Wang, X., Wang, R.: A Review of Key Technologies for Emotion Analysis Using Multimodal Information. Cognitive Computation 1–27 (2024).
12. Papadeas, I., Tsochatzidis, L., Amanatiadis, A., Pratikakis, I.: Real-Time Semantic Image Segmentation with Deep Learning for Autonomous Driving: A Survey. Applied Sciences 11(19), 8802 (2021).
13. Seo, Y., Lee, K., Liu, F., James, S., Abbeel, P.: Harp: Autoregressive Latent Video Prediction with High-Fidelity Image Generator. In: 2022 IEEE International Conference on Image Processing (ICIP), pp. 3943–3947. IEEE (2022).
14. Sina, A., Lorenz, C., Weese, J., et al.: Pathology Synthesis of 3D Consistent Cardiac MR Images Using 2D VAEs and GANs. In: Simulation and Synthesis in Medical Imaging: 7th International Workshop, SASHIMI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings, pp. 34–42. Springer (2022).
15. Van Den Oord, A., Kalchbrenner, N., Kavukcuoglu, K.: Pixel Recurrent Neural Networks. In: Proceedings of the 33rd International Conference on Machine Learning (ICML2016), Vol. 4, pp. 2611–2620. Curran Associates, Inc. (2016).
16. Vahdat, A., Kautz, J.: NVAE: A Deep Hierarchical Variational Autoencoder. Advances in Neural Information Processing Systems 33, 19667–19679 (2020).
17. Weissenborn, D., Täckström, O., Uszkoreit, J.: Scaling Autoregressive Video Models. arXiv preprint arXiv:1906.02634 (2019).
18. Yazici, Y., Yap, K. H., Winkler, S.: Autoregressive Generative Adversarial Networks (2018).