A Comprehensive Review of Diffusion Models in AI-Generated Content for Image Applications

Yuxin Jia

International Engineering College, Xi'An University of Technology, Xi'An, China

3222241066@stu.xaut.edu.cn

Abstract. Diffusion models in AI-generated content (AIGC) have revolutionized image generation with their ability to transform noise into structured data through a denoising process. Despite their prowess, these models grapple with high computational demands, inconsistent image quality, and challenges in producing high-resolution images. This paper discusses the performance of diffusion models in image generation, emphasizing the need for improved computational efficiency, generation quality, and contextual understanding. Future research is aimed at developing real-time generation technologies, addressing ethical and legal concerns, and enhancing data privacy measures. With these advancements, diffusion models are expected to achieve broader application, enhanced performance, and increased safety in image generation, expanding their utility in fields such as advertising, medical imaging, and art creation. The potential of diffusion models to meet the increasing demands for quality, content, and diversity in image applications makes them a pivotal area of study in visual technology and artificial intelligence.

Keywords: AIGC, diffusion model, image generation.

1. Introduction

The rapid evolution of Artificial Intelligence-Generated Content (AIGC) technologies has ushered in transformative changes across various domains, particularly in image generation. Traditional image generation methods are increasingly unable to satisfy the growing demands for higher quality, diversity, and content authenticity in fields such as digital media, advertising, and medical imaging [1]. This has propelled the development of more sophisticated techniques, among which diffusion models have emerged as a crucial innovation. These models leverage a progressive denoising process to enhance the realism and detail of generated images, thus addressing some of the limitations of earlier generative technologies [2].

Currently, diffusion models stand at the forefront of AIGC research due to their ability to produce high-fidelity images. This class of generative models has gained prominence for its unique approach, which transforms noise into detailed, structured images through a controlled process of adding and removing noise [3]. As technology progresses, the application of diffusion models has expanded beyond simple image synthesis to more complex tasks like text-to-image translation, medical imaging, and interactive media creation. The integration of diffusion models with other AI technologies, such as deep learning and neural networks, has further enhanced their capabilities, making them a powerful tool in the AI toolkit [4].

This paper aims to provide a comprehensive review of the current state of diffusion models within the AIGC landscape, highlighting their operational principles, applications, and the challenges they face. It will explore how these models have been optimized for better performance and detail the advancements that have been made in generating high-resolution images. Additionally, the paper will discuss the potential future developments in this area, focusing on increasing computational efficiency, enhancing the quality of generated images, and expanding their application domains. By delving into these aspects, the study seeks to offer valuable insights to researchers, engineers, and professionals working with AI-driven image generation technologies [5].

2. Overview of AI-Generated content (AIGC)

2.1. Evolution of AIGC in image generation

Based on the research on artificial intelligence, the development of AIGC has also received more attention in daily applications due to the continuous deepening of machine learning and the continuous improvement of generative models [4]. Nowadays, with the development of technology, people have utilized technologies such as deep learning, generative adversarial networks, autoregressive models, and diffusion models, making AIGC an emerging direction in the field of artificial intelligence. These technologies can automatically generate images, music, text, and so on, and their generation quality is relatively high [5]. In order to generate content that better meets people's requirements and has certain innovation and characteristics, the core part of AIGC technology is to train through a large amount of data, so that the model can capture and learn from the data. In the early days, the application areas of AIGC technology mainly focused on generating images, text, and other directions, and then generated more artistic and creative works through simulation and creation. With the continuous advancement of technology and breakthroughs in deep learning and generative adversarial networks, AIGC technology is becoming increasingly mature and has been further developed [6]. With the continuous evolution of technology, AIGC technology is no longer limited to generating static images and other content, but can even generate animation, video, and natural language content. Its application areas and generation quality have significantly improved, and it also has unlimited potential [7]. Based on GAN technology, AIGC has undergone continuous improvement and maturity in its field, from initially only being able to generate text content to later being able to perform style control and large pre trained models, learning patterns from data. The autoregressive model based on Transformer has also provided new ideas and perspectives for AIGC in the field of image generation, and has achieved remarkable results. Finally, diffusion models, with their stable training methods and high-quality generated content, have attracted a large number of researchers' interest as a new research direction and field [8]. Overall, AIGC has made significant progress in various aspects of image generation technology, which not only represents technological advancement but also indicates that AIGC has broader prospects in practical applications.

2.2. Principles of AIGC in image generation

AIGC technology, also known as artificial intelligence generated content technology, is a representative of the integration of artificial intelligence and generative computing [9]. It can not only capture and learn patterns from large-scale data, but also generate high-quality content that better meets people's expectations and requirements. AIGC is a new breakthrough and technology based on digital technology, which requires the comprehensive use of various AI models within it, so that it can be extended to the field of images at the technical level, including more possibilities for image acquisition, generation, processing, etc [10].

3. Diffusion models

3.1. Fundamental principles of diffusion models

Diffusion model is a new type of generative model. Mainly through a series of steps, noise is gradually converted into structured data to synthesize images, and these models are also known as parameterized

Markov chains. The diffusion process starts from the data distribution $x_0 \sim q(x_0)$. Simultaneously add Gaussian noise in an increasing manner over T time steps. The data xt will be transformed and denoised at each step t:

$$q(x_{1:T}|x_0) := \prod_{t=1}^{T} q(x_t|x_{t-1}) \qquad (1)$$

$$q(x_t|x_{t-1}) = N\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \qquad (2)$$

In the formula, βt refers to the variance hyperparameter of the noise.

In the inverse process of the model, the goal is to gradually denoise these data, this is similar to the inverse process of a Markov chain. This inverse process starts from the noise vector x_T and transitions to the original data distribution $q(x_0)$ simultaneously. The generative model will parameterize pθ $(x_t-1|x_t)$ to a normal distribution through inverse transformation:

$$p\theta(x_{t-1}|x_t) = N(x_{t-1}; \mu\theta(x_t,t), \textstyle\sum_{\theta}(x_t,t)) \qquad (3)$$

Deep neural networks are typically instantiated by architecture. Among them, deep neural networks (usually instantiated by architectures such as UNet) parameterize the mean μθ (x_t ,t) and variance Σθ(x_t, t). Input noise data$x_t$and time step t, output parameters of normal distribution, thus predicting the required noise ϵθ for the model to reverse the diffusion process. Firstly, sample the noise vector $x_T \sim$ $p(x_T)$, then continuously sample from the learned transformation kernel $x_t-1 \sim$ pθ $(x_t-1|x_t)$, when t=1, it proves that the reverse diffusion process is complete, thus synthesizing a new data instance x_0.

3.2. Text-to-Image diffusion models

The application of diffusion models is gradually expanding, from initial text generation to image generation, marking an important advancement in generative models. GLIDE is the first system to formalize text to image conversion into a diffusion model, achieved by training the model in a noisy image space. This process involves gradually adding noise and reverse denoising, allowing the model to understand how to generate corresponding visual content from textual descriptions. Imagen has adopted a more efficient strategy by using pre trained large language models (LLMs) as text encoders. This method not only reduces computational requirements, but also improves the quality of generated images. Imagen encodes textual information into feature vectors and then uses a diffusion model to generate high-resolution images from these features. The innovation of both lies in optimizing the efficiency of model training and generation, promoting further integration between computer vision and natural language processing fields.

4. Application domains

Image generation: When generating high-quality images, the diffusion model can start with random noise and generate detailed images through gradual iteration and adjustment. For example, when generating real faces, the model can capture skin texture, hair details, and lighting effects, making the final generated face look natural and highly realistic. In landscape image generation, models can create complex natural scenes such as mountains, lakes, and plants, preserving tiny details such as the texture of leaves and the refraction of light, thereby achieving vivid and realistic visual effects. The core of this technology lies in its gradual "denoising" process, which enables each step to more accurately reflect the details and features of the image.

Image restoration: The specific process is as follows: Identify missing areas: Firstly, identify the damaged or missing parts in the image. These areas may have been caused by physical damage, deletion, or other reasons. Establish model input: Input the damaged area along with the surrounding normal area into the diffusion model. The model understands the context by analyzing the information of surrounding pixels. Generating content: The diffusion model starts with random noise and gradually adjusts and generates pixels to fill the region. Through training, the model can generate textures, colors, and structures that are consistent with the surrounding content, making the repaired area look natural and seamless. Detail enhancement: After generating preliminary repair results, post-processing may be

performed to enhance details and realism, such as adjusting details, color balance, and lighting effects to perfectly blend the repaired area with the original image.

This method can handle complex repair tasks, such as filling missing objects, repairing damaged photos, or reconstructing damaged artworks, generating content that is natural and highly consistent with the original image. Image super-resolution: Elevate low resolution images to high resolution, enhancing image details and clarity.

Image super-resolution. The use of diffusion models for image super-resolution involves the following detailed steps: Firstly, take a low resolution image as input. This image may have blurry details and insufficient clarity due to limitations in the shooting equipment, compression, or other reasons. By preprocessing, adjust low resolution images to the format required by the model. The preprocessing steps may include standardization, resizing the image, or converting to a specific color space. Then generate high-resolution images, and visually inspect the generated high-resolution images to ensure that the details and clarity meet the requirements. Through these steps, diffusion models can effectively elevate low resolution images to high resolution, enhance image details and clarity, and are widely used in fields such as photography, medical imaging, and satellite imagery.

Image editing. To remove, add, or modify objects in an image while maintaining its naturalness, the following detailed steps are involved: Firstly, object recognition and segmentation are performed, followed by object removal, addition, modification, and other operations. Finally, processing and optimization are carried out. Through these steps, the diffusion model can efficiently remove, add, or modify objects in the image, ensuring that the processed image maintains naturalness and consistency.

5. Challenges and future directions

5.1. Current technical challenges

High computational resource requirements: Diffusion models require a large amount of computational resources for training and inference, especially when generating high-resolution images, which may result in high computational costs and time delays.

Unstable generated image quality: Although diffusion models perform well in many scenarios, in some complex or detail rich image generation tasks, the model may generate inconsistent or unnatural details, affecting the quality of the final image.

Difficulty in generating high-resolution images: During the process of generating images from low resolution to high-resolution, the model may have difficulty maintaining consistency in details, especially when generating very high-resolution images, where blurring or artifacts may occur.

Challenges in handling large-scale content: When removing, adding, or modifying objects, the model needs to handle a wide range of contextual information to ensure that the generated content naturally integrates into the image, which places high demands on the model's contextual understanding ability.

Ethical and legal issues: The generated images may be used for the dissemination of false information or improper purposes, and ensuring that the application of the model complies with ethical and legal norms is an important challenge.

Data privacy: When using real-world image data for training, protecting data privacy and avoiding data abuse are key issues that need to be addressed.

5.2. Future research directions and trends

Improved computational efficiency: Research more efficient algorithm and model architectures to reduce computational resource requirements and generation time, such as through model compression and optimization of inference processes.

High resolution generation: Developing technologies that can generate images with higher resolution and rich details, improving the quality of generation and visual effects.

Context understanding enhancement: Improve the context understanding ability of the model to better handle object removal, addition, or modification tasks in complex scenes, making the generated content more naturally integrated into the image.

Ethics and Security: Addressing ethical and legal issues in generating content, ensuring that model applications comply with regulations, and preventing misuse and dissemination of false information.

Data privacy protection: Develop data privacy protection technologies to securely use and process sensitive data while complying with relevant regulations.

These directions and trends will drive the application of diffusion models in the field of images towards higher quality, wider range, and safer directions.

6. Conclusion

AIGC based on diffusion models is rapidly developing in the field of images, demonstrating its strong potential in generating images. These models are capable of generating highly realistic images, but face a series of technical challenges. Firstly, these models typically require a large amount of computing resources, which leads to high computational costs and time delays, especially when processing high-resolution images. Secondly, the quality of the generated images may sometimes be unstable, especially when dealing with detailed or complex scenes, where the model may experience missing or unnatural details.

Future research directions will focus on several key areas. Firstly, improving computational efficiency is an important goal, including optimizing model architecture and inference processes to reduce resource requirements and accelerate generation speed. Secondly, in response to the challenges of high-resolution generation, researchers will strive to improve techniques to ensure the clarity and consistency of details when generating higher resolution images. In addition, enhancing the contextual understanding ability of the model to better handle tasks such as object removal, addition, or modification in images will be another important direction, which will help generate content that blends more naturally into the original image.

In addition, expanding the application scope of the model to other fields such as medical imaging, artistic creation, and virtual reality will help promote the diversified application of technology. Developing adaptive and real-time generation technologies to instantly generate or modify images based on user input will also significantly enhance the user experience. At the same time, it is essential to address ethical and legal issues related to content generation, ensure that technology applications comply with regulations, and prevent the spread of false information. Finally, protecting data privacy and ensuring compliance with relevant regulations when using and processing sensitive data will further enhance the credibility and security of the technology. Through these comprehensive advancements, diffusion models will achieve higher quality, wider applications, and safer use in the field of image generation.

References

[1] Zhang C, Zhang C, Zhang M and Kweon I S 2023 Text-to-image diffusion models in generative AI: a survey arXiv preprint arXiv:2303.07909

[2] Xing Z, Feng Q, Chen H, Dai Q, Hu H, Xu H, Wu Z and Jiang Y-G 2023 A survey on video diffusion models ACM Comput. Surv.

[3] Cao Y, Li S, Liu Y, Yan Z, Dai Y, Yu P S and Sun L 2023 A comprehensive survey of AI-generated content (AIGC): a history of generative AI from GAN to ChatGPT arXiv preprint arXiv:2303.04226

[4] Zhu X, Xu H and Zhao Z 2021 An environmental intrusion detection technology based on WiFi Wireless Personal Commun. 119(2) 1425–1436

[5] Du H, Zhang R, Niyato D, Kang J, Xiong Z, Kim D I, Shen X and Poor H V 2023 Exploring collaborative distributed diffusion-based AI-generated content (AIGC) in wireless networks IEEE Network 38(3) 178–186

[6] Lin L, Gupta N, Zhang Y, Ren H, Liu C-H, Ding F, Wang X, Li X, Verdoliva L and Hu S 2024 Detecting multimedia generated by large AI models: a survey arXiv preprint arXiv:2402.00045

[7]   Foo L G, Rahmani H and Liu J 2023 AI-generated content (AIGC) for various data modalities: a survey arXiv preprint arXiv:2308.14177

[8]   Zhu M, Chen H, Yan Q, Huang X, Lin G, Li W, Tu Z, Hu H, Hu J and Wang Y 2024 GenImage: a million-scale benchmark for detecting AI-generated image Advances Neural Inf. Process. Syst. 36

[9]   Jamal S 2024 Applications of predictive and generative AI algorithms: regression modeling, customized large language models, and text-to-image generative diffusion models

[10] Du H, Zhang R, Niyato D, Kang J, Xiong Z, Cui S, Shen X and Kim D I 2023 User-centric interactive AI for distributed diffusion model-based AI-generated content arXiv preprint arXiv: 2311.11094