Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD27180

From Probabilistic Models to Transformers: The Technological Trajectory of Generative AI in Vision Tasks

Chengcheng Dong

University of Wollongong, Wollongong, Australia
127dcc@gmail.com

Abstract. With the rapid advancements in generative artificial intelligence (GAI), visual computing has witnessed transformative changes across a range of applications such as image synthesis, restoration, super-resolution, 3D reconstruction, and medical imaging. This review systematically examines the evolution of generative models, from early statistical approaches to state-of-the-art transformer-based architectures. Key models including variational autoencoders (VAE), generative adversarial networks (GAN), and diffusion models are compared in terms of their structure, training stability, image quality, and suitability for various visual tasks. In addition to technical progress, the review highlights the ethical, explainability, and safety challenges associated with GAI deployment, especially in high-stakes fields like healthcare and manufacturing. While GAI enables highly realistic and semantically meaningful image generation, challenges remain in balancing innovation with interpretability, computational efficiency, and social responsibility. The paper also acknowledges the limitations of static literature reviews in a rapidly evolving domain and calls for ongoing comparative studies and interdisciplinary collaboration to shape a responsible and sustainable future for generative AI in visual computing.

Keywords: Generative AI, visual computing, diffusion models, image synthesis

1. Introduction

With the rapid development of generative AI (GAI) in the field of visual computing, it has shown great potential in many cutting-edge applications such as image synthesis, editing, super-resolution, 3D reconstruction, and medical imaging [1]. GAI breaks through the limitations of traditional discriminative models by simulating and generating realistic new data, and promotes the technological innovation and application diversification of machine vision [2]. However, the complexity of related technologies, training stability, computing resource requirements, and ethical and security issues have also brought many challenges to the continued development of this field. Given that current research mainly focuses on a single model or specific application scenario, lacks systematic sorting and comparison, this paper aims to review the technological evolution of generative AI in visual tasks, mainstream model architectures [3], and their application performance in different fields through a comprehensive literature review.

This paper will mainly adopt systematic literature analysis and inductive summary, covering early statistical generative models, variational autoencoders (VAE), generative adversarial networks

(GAN), diffusion models, and the latest architectures based on transformers, focusing on the performance advantages and disadvantages of each model in key visual tasks such as image synthesis, restoration, enhancement, and 3D reconstruction. In addition, this article also focuses on the ethical, explainable and security issues of generative AI technology, and looks forward to future technology development directions and research hotspots based on the current application status of high-risk fields such as medical imaging and industrial detection. By comprehensively integrating existing results, this article aims to provide researchers and application developers with systematic theoretical support and practical guidance to promote the healthy and sustainable development of generative AI in visual computing.

2. Technological evolution of generative AI

2.1. Early generative models

Generative modeling originated from statistical and probabilistic methods, aiming to achieve data generation by characterizing the joint distribution of observed data and latent variables. Early classic models such as Hidden Markov Model (HMM) and Boltzmann Machine attempted to simulate and generate samples that conform to statistical characteristics by capturing the transition rules between states and the inherent probabilistic structure of data. These models have shown certain effects in fields such as speech recognition and natural language processing. However, when faced with complex and high-dimensional visual data, this type of traditional generative model exposes obvious limitations. On the one hand, with the rapid growth of data dimensions, the model parameter space expands sharply, resulting in a significant increase in computational complexity, and the training process is time-consuming and difficult to converge; on the other hand, the model's estimation of parameters depends on precise probabilistic assumptions and approximate methods, and these assumptions are often difficult to meet the complexity and nonlinear characteristics of actual data, limiting its ability to express generative effects.A substantial shift came with the introduction of the Variational Autoencoder (VAE). VAEs have used neural networks to combine deep learning with probabilistic modeling. Given the complexity of these theoretical relationships, they introduced an encoder-decoder structure that tends to suggest a mapping of input data into a smooth, continuous latent space. This made it easier to sample and move between different points in the data space. The evidence reveals that VAEs were mathematically elegant and relatively stable during training. But they presumably had a downside. Their images often looked blurry. This was predominantly due to their use of Gaussian assumptions and the type of loss function typically employed in training [4].

2.2. Key advancements and evolutionary milestones

As limitations of the earlier system became obvious, the expedient focus of the research community seemed to have transitioned to the latter applications. Conditional GANs are one of the initial significant improvements in this broader evaluation setup. These models have extended the input with extra information, such as class labels or text descriptions, to help steer the generation process. This means that generating more specialized outputs became feasible. This is extremely relevant for controlling tasks such as semantic image synthesis when the model should produce detailed images conditioned on complex inputs [5]. A breakthrough with autoregressive models and diffusion-based models emerges. Autoregressive models, such as PixelCNN, generate images one pixel at a time—a process that is slow but preserves fine details. Diffusion models adopt a different strategy, given the subtle nature of these findings. They start with random noise and progressively refine the noise into a sharp image via a learned denoising process [6]. These models excel in terms of image quality and stability during training, especially when generating high-resolution images. In addition, the rise of the Transformer architecture marks a new stage in the development of generative models. It has shown strong advantages in multimodal data fusion and complex pattern capture, opening up broader prospects for generative vision tasks[7]. In summary, these technological advances have not only greatly promoted the performance improvement of generative artificial intelligence, but also laid a solid foundation for building more efficient, stable and controllable visual generative models in the future.

2.3. Comparison of mainstream model architectures

The variational autoencoder (VAE) effectively learns the latent representation of data and models uncertainty through a structured encoding-decoding mechanism. However, due to its Gaussian distribution assumption and loss function design, the images generated by VAE are usually blurry and lack details. The generative adversarial network (GAN) can generate clearer and more realistic visual content through an adversarial training mechanism, but its training process is extremely challenging and prone to problems such as mode collapse and training instability. It requires sophisticated hyperparameter adjustment and training techniques to maintain model stability [8]. In contrast, diffusion models have attracted much attention in recent years due to their significant advantages in image generation quality and training stability. Although diffusion models usually require higher computing resources and longer inference time, they can gradually transform random noise into high-fidelity images, significantly improving the generation effect. As an emerging force in the field of vision, models based on the Transformer architecture have demonstrated strong scalability and excellent processing capabilities for multimodal data, and have gradually become an important direction for promoting the development of generative models.

3. Applications of generative AI in vision tasks

3.1. Image synthesis and restoration

One of the earliest and most common uses of generative AI is image synthesis and restoration. Within this broader analytical framework, this includes tasks like image inpainting and super-resolution [9]. Inpainting fills in missing or blocked parts of an image. The model uses surrounding visual information to guess what should be there. Generative models are well-suited for this task. They ostensibly learn the structure of images and can restore missing areas in a natural way. Super-resolution represents another key application, which involves transforming low-resolution images into high-resolution ones . Given the complexity of these theoretical relationships, GANs and diffusion models are typically used here. These methods can add fine details that are not present in the original image, offering perceptually sharper results [10].

3.2. Image enhancement and style transfer

Generative AI plays a significant role in image enhancement. This includes fundamental tasks like deblurring, denoising, low-light enhancement, and contrast adjustment. These techniques suggest particular utility when image quality seems compromised due to camera limitations or challenging environmental conditions. In low-light scenes, the generative models brighten images and adjust colors, without introducing noise or visual artifacts. These findings show that diffusion models and transformer-based architectures demonstrate substantial strength in this domain. These models adjust lighting while preserving image details by learning complex patterns of light and shadow [11]. This method can achieve style transfer while basically preserving the image structure. Related models usually separate the content and style information in the image through disentangling technology, and then integrate new style elements (such as painting style, lighting changes, or sketch effects) into the original content. Since this process involves complex image feature modeling and mapping, this type of technology has been widely used in creative design, mobile applications, and digital media production. Research further pointed out that the application of generative models in the image field is continuing to expand. For example, they can automatically convert X-ray images into MRI-style medical images, or convert satellite remote sensing images into map-like visual expressions. These examples fully demonstrate the powerful capabilities of generative models in visual content re-rendering and cross-domain reconstruction and their future application potential..

3.3. 3D model generation and virtual reality

Generative AI has also gradually expanded to the field of 3D vision, providing a new paradigm for the acquisition and reconstruction of three-dimensional content. Compared with traditional 3D modeling methods that usually rely on geometric prior knowledge, explicit rules and a large amount of manual intervention, generative methods adopt data-driven strategies, which can learn complex spatial structures and generate realistic three-dimensional content in a more automated way. In recent years, the emergence of technologies such as Neural Radiance Fields (NeRF) represents an important breakthrough in this field [12]. NeRF learns from images of multiple two-dimensional perspectives to build a continuous volume representation, thereby achieving fine reconstruction of three-dimensional scenes and rendering from new perspectives. This technology does not require mesh modeling or point cloud priors and has strong versatility and realism. From a functional point of view, NeRF and its extended models have demonstrated superior performance in processing complex textures, lighting changes and occluded objects. They support image generation conditioned on viewpoints, can render high-quality view conversion results, and greatly enhance the user's immersive experience in virtual reality (VR) and augmented reality (AR) environments. In addition, generative 3D methods have also brought new possibilities to the fields of digital cultural relics restoration, game development, film and television production, etc. With the optimization of training efficiency and the improvement of model generalization ability, 3D modeling based on generative AI is expected to become a mainstream solution in the future [12].

3.4. Applications in medical and industrial inspection

In high-risk fields such as medicine and manufacturing, generative AI has shown unique application value, especially in generating synthetic data, enhancing the quality of diagnostic images, and identifying subtle structures that are difficult to detect with conventional manual inspection. In medical image analysis, generative models are widely used to generate synthetic CT, MRI, and X-ray images to expand training datasets to alleviate the problem of insufficient labeled samples [13]. These synthetic images are usually highly anatomically realistic and can effectively support the training and generalization capabilities of diagnostic classifiers and image segmentation models. In addition, generative AI is also used for cross-modal image conversion tasks, such as converting MRI images to CT images or vice versa without requiring patients to undergo double scanning. This type of method not only improves the accessibility of image data, but also reduces patient radiation exposure and examination costs to a certain extent. With the continuous improvement of model accuracy and stability, the application of generative AI in medical imaging is gradually moving from research to clinical deployment, promoting the development of personalized medicine and intelligent diagnostic systems.

4. Discussion

The field of generative artificial intelligence has undergone a profound transformation, evolving from early probabilistic models to sophisticated deep learning architectures that produce highly realistic and semantically meaningful visual content. Within this broader analytical framework, generative AI in computer vision has not only expanded the capabilities of machine perception and generation, but also redefined the nature of visual tasks themselves. From reconstructing missing image regions to synthesizing three-dimensional virtual environments, generative models enable machines to understand and reproduce the visual world with unprecedented precision and creativity.

Looking ahead, the success or failure of generative AI in the field of visual tasks will depend on its ability to achieve an effective balance between technological innovation and social responsibility. As model architectures become increasingly complex and application scenarios become increasingly diverse, it is particularly important to build systems that are interpretable, transparent, and fair. In particular, innovation in architectural design, optimization of efficient training strategies, and continuous improvement of human-centered evaluation standards will become the core driving force for the practical application and sustainable development of generative models. For example, in medical imaging, generative AI can be used to synthesize scarce pathological images to enhance diagnostic capabilities, but if the model training data is biased, it may lead to misdiagnosis or even ethical issues. Therefore, while improving performance, mechanisms must be introduced to ensure the reliability and traceability of model output. For example, in the creative industry, AI-generated images are widely used in advertising, games, and digital art design, which greatly improves creative efficiency, but also triggers new controversies in originality and copyright protection. Therefore, future development requires not only breakthroughs at the algorithm level, but also interdisciplinary collaboration between researchers, developers, and policymakers to jointly promote the formulation of technical standards and ethical frameworks to ensure that generative AI serves society in a safe, inclusive, fair, and beneficial way for the majority of people.

5. Conclusion

This review systematically traces the technical evolution, mainstream model architectures, and practical applications of generative artificial intelligence (GAI) in visual computing. From early probabilistic models to more advanced architectures. The field has undergone rapid and transformative development, with each generation of models contributing new capabilities.

In practice, GAI has achieved remarkable success in a variety of visual tasks, including image restoration, enhancement, style transfer, and medical or industrial detection. These models not only improve technical efficiency, but also introduce novel workflows in areas such as medical diagnosis, digital content creation, and virtual reality.

This review has some limitations. Given the rapid development of GAI, some of the reviewed techniques or research results may soon become outdated, highlighting the need for continuous updates in this field. Future research should prioritize the development of more interpretable, resource-efficient, and ethical generative models. More comparative studies are also needed to understand the trade-offs of different architectures in practical applications.

References

[1] Generative AI accelerates homologation: FEV simplifies country-specific type approval processes. (2025). M2 Presswire.

[2] Tasdelen, O., & Bodemer, D. (2025). Generative AI in the classroom: Effects of context-personalized learning material and tasks on motivation and performance. International Journal of Artificial Intelligence in Education, prepublish, 1–22.

[3] Shabeeb, Z., Goyal, N., Nantogmah, A. P., & Lin, S. (2025). Learning the diffusion of nanoparticles in liquid phase TEM via physics-informed generative AI. Nature Communications, 16(1), 6298.

[4] Wu, Z., Cao, L., & Qi, L. (2024). EVAE: Evolutionary variational autoencoder. IEEE Transactions on Neural Networks and Learning Systems, 36(2), 3288–3299.

[5] BrightEdge Survey: Brands adapting to rise of AI-search and shift to generative engine optimization. (2025). Manufacturing Close-Up.

[6] Yang, X., Liu, X., & Gao, Y. (2025). The impact of generative AI on students' learning: A study of learning satisfaction, self-efficacy and learning outcomes. Educational Technology Research and Development, prepublish, 1–14.

[7] Lee, C., Kim, J., Lim, S. J., & Zhang, Y. (2025). Generative AI risks and resilience: How users adapt to hallucination and privacy challenges. Telematics and Informatics Reports, 19, 100221.

[8] Guha, P., Chen, S., Georges, A., & Dutta, A. (2025). Turning to Gen-AI as an empowerment tool for parents of teenage girls for conversations on online sexual harassment. AI & Society, prepublish, 1–13.

[9] Rodger, D., Mann, P. S., Earp, B., & Zhu, X. (2025). Generative AI in healthcare education: How AI literacy gaps could compromise learning and patient safety. Nurse Education in Practice, 87, 104461.

[10] Granić, A. (2025). Emerging drivers of adoption of generative AI technology in education: A review. Applied Sciences, 15(13), 6968.

[11] Vinothkumar, S., Varadhaganapathy, S., Shanthakumari, R., Dhanushya, S., Guhan, S., & Krisvanth, P. (2024, June). Utilizing generative AI for text-to-image generation. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1–6). IEEE.

[12] Croce, V., Caroti, G., De Luca, L., Piemonte, A., & Véron, P. (2023). Neural radiance fields (NeRF): Review and potential applications to digital cultural heritage. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, 453–460.

[13] Koetzier, L. R., Wu, J., Mastrodicasa, D., Lutz, A., Chung, M., Koszek, W. A., ... & Willemink, M. J. (2024). Generating synthetic data for medical imaging. Radiology, 312(3), e232471.