

AI 驱动的图片生成技术：美学、技巧与创新

冯海桐

2025 年 1 月 6 日

摘要

随着人工智能技术的迅猛发展，AI 驱动的图片生成技术已成为多个领域的研究热点。本综述综合分析了相关文献，探讨其在美学呈现、创作技巧以及创新性应用方面的进展。从早期基于规则的方法到如今深度学习主导的模型，如生成对抗网络（GANs）等，这些技术不仅变革了图像创作流程，还拓展了艺术表达边界。通过对大量研究成果的梳理，揭示了当前技术面临的挑战，包括生成结果的可控性、审美标准的量化等，为未来进一步研究与应用提供参考。

Abstract: With the rapid development of artificial intelligence technology, AI-driven image generation technology has become a research hotspot in multiple fields. This review comprehensively analyzes relevant literature to explore its progress in aesthetic presentation, creative techniques, and innovative applications. From the early rule-based methods to today's deep learning-dominated models, such as Generative Adversarial Networks (GANs), these technologies have not only transformed the image creation process but also expanded the boundaries of artistic expression. By sorting out a large number of research results, the current challenges faced by the technology are revealed, including the controllability of generation results and the quantification of aesthetic standards, providing a reference for future further research and applications.

1 引言

在当今数字化时代，AI 驱动的图片生成技术正以前所未有的速度改变着我们与视觉内容的交互方式。从艺术创作、广告设计到医疗影像、虚拟现实等众多领域，该技术都展现出巨大潜力。研究这一主题，对于深入理解技术发展脉络、挖掘创新应用以及探

索艺术与科技融合边界至关重要。本文献综述旨在系统梳理相关研究，为该领域研究者、从业者提供全面知识基础，助力技术进一步突破。

2 方法

2.1 文献检索策略

运用学术数据库如 Web of Science、IEEE Xplore、ACM Digital Library 等，结合关键词“AI 图像生成”、“生成对抗网络（GANs）”、“美学与图像生成”、“图像生成技巧”、“创新应用图像生成”等进行检索。检索时间范围设定为近十年（2014 - 2024），确保涵盖最新研究成果。

2.2 文献选择标准

纳入标准为直接围绕 AI 驱动图像生成技术核心主题，探讨技术原理、美学特性、创作技巧、创新实践等方面的原创性研究论文、综述文章及高质量会议报告；排除标准为单纯介绍软件工具使用、无实质技术分析以及偏离主题的非学术资料。

3 主要研究

在过去几年中，由于生成式人工智能的进步，特别是深度生成模型 [2] 和神经渲染 [6]，图像生成取得了非常令人印象深刻的进展。

先看 [2]，它是生成式对抗网络（GANs）领域的奠基之作，由 Ian Goodfellow 等人撰写。在这篇文献里，首次详细阐述了 GANs 的基本原理，通过构建生成器与判别器之间的对抗博弈，使得生成模型能够学习到真实数据的分布，进而生成逼真的图像。这种创新性的思路打破了传统图像生成方法的局限，为后续无数的研究提供了核心理论支撑，诸多基于 GANs 的变体模型皆是在其基础上拓展而来，开启了图像生成技术的新篇章。

[7] 关于 Conditional GANs (CGANs) 的研究，为多模态引导的图像生成奠定了基础。它提出在 GANs 框架中引入条件信息，如类别标签、文本描述等，使得生成模型能够根据特定的需求生成定制化的图像。这一思想的落地，开启了从单一模式图像生成向有条件、多模态驱动图像生成的转变，催生了大量后续的改进模型，广泛应用于广告设计、虚拟场景构建等诸多领域，满足了人们对于个性化、多样化图像的需求。

[5] 的重要性同样不可忽视，其针对多模态图像合成展开深入探索，提出递归变分自编码器在结合图像描述生成新视觉场景方面的独特优势。在当时，多模态信息融合用于图像生成尚处于起步阶段，该文献的出现为后续研究人员指明了方向，让大家意识到利用文本等其他模态信息引导图像生成的巨大潜力，诸多跟进的研究在此基础上不断优化算法，致力于实现更加智能、精准的图像生成效果，满足多样化的实际应用需求。

[4] 围绕图像到图像的转换任务，运用 GANs 架构取得了重大突破。它展示了如何利用成对的数据训练模型，实现诸如将草图转换为逼真图像、将低分辨率图像提升为高分辨率图像等神奇效果。这不仅拓展了 GANs 的应用边界，也为图像处理相关行业，如设计、影视后期制作等提供了强大的技术工具，使得图像编辑与创作变得更加高效、便捷。

[8] 引入的 Transformer 模型，虽然最初并非专为图像生成设计，但它所蕴含的自注意力机制革新了模型对序列数据的处理方式。在图像生成领域，这种机制被借鉴用于捕捉图像不同区域之间的长距离依赖关系，突破了传统 CNN 架构在感受野上的限制，使得模型能够更好地理解图像的整体结构与语义信息，进而在生成图像时实现更精细的细节控制与更高的质量保证，为图像生成技术的进一步发展开辟了新路径。

[6] 聚焦于神经辐射场 (NeRF) 技术。这一研究创新性地提出利用多层感知机 (MLP) 来表示连续的场景函数，通过对不同视角光线的采样与渲染，实现了高质量的三维场景重建与新视角合成。以往的三维重建方法往往面临精度不足、数据处理复杂等问题，而 NeRF 的出现改变了这一局面，它以简洁高效的方式为三维感知图像合成带来了全新的思路，使得研究者们能够更加精准地捕捉和再现现实世界中的三维结构与光影效果，推动了从二维向三维图像生成的关键跨越。

[3] 所介绍的扩散模型，从一个全新的视角切入图像生成问题。通过模拟噪声逐渐去除的逆过程，让模型学习到数据的生成分布。与传统方法相比，扩散模型在生成图像的多样性与质量上展现出独特优势，尤其是在处理复杂场景、高细节要求的图像生成任务时表现出色，为图像生成领域注入了新的活力，引发了学术界和工业界的广泛关注与深入研究。

4 研究空白和争议

4.1 当前研究局限性

尽管生成技术进步显著，但生成结果精准控制仍困难，如指定生成含复杂情感、特定文化隐喻图像难达标；模型训练依赖海量数据，存在隐私、版权隐患；对小众艺术风

格、特殊审美理解不足，难以传承与创新小众文化视觉表达。

4.2 存在争议或不同观点

关于 AI 生成图像艺术地位争议不断，部分艺术家认为其缺乏人类创作灵魂，仅是算法拼凑；在技术发展路径上，有人主张强化端对端生成简化流程，也有人强调模块化精细控制；审美标准界定，不同文化、群体差异大，统一量化指标可行性存疑。

除此之外，一些学者还注意到 AI 图像生成存在偏见问题，如生成结果中存在性别、种族歧视等 [1]，这也是当前研究亟待解决的问题。

4.3 未来研究建议

加强多模态信息融合，结合文本、音频等引导生成，提升语义理解与表达；探索可解释性 AI 图像生成，揭开“黑箱”让创作者更好掌控；深化跨文化研究，构建包容多元审美模型，拓展全球适用创意边界。

5 总结

本综述系统回顾 AI 驱动图像生成技术文献，呈现从基础模型构建到跨领域蓬勃发展历程。当前技术在多方面成果斐然，但诸多挑战待攻克。未来研究应聚焦填补空白、化解争议，沿着多模态融合、可解释性、跨文化适应方向，推动图像生成技术迈向新阶段，为各行业持续赋能，实现科技与人文艺术更深度交织共荣。

参考文献

- [1] Aadi Chauhan, Taran Anand, Tanisha Jauhari, Arjav Shah, Rudransh Singh, Arjun Rajaram, and Rithvik Vanga. Identifying race and gender bias in stable diffusion ai image generation. *Journal of AI Research*, 2023.
- [2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27:2672–2680, 2014.
- [3] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

- [4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1125–1134, 2017.
- [5] Elman Mansimov, Emilio Parisotto, Jimmy Lei Ba, and Ruslan Salakhutdinov. Generating images from captions with attention. *arXiv preprint arXiv:1511.02793*, 2015.
- [6] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, 2020.
- [7] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- [8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 5998–6008, 2017.