

Generating Realistic AI Art: Balancing Creativity and Controllability

¹Roopa R, ²Namitha M U

Assistant Professor Department of Master of Computer Applications, BIET, Davangere-577004.

roopa.satishlaxmanrao@gmail.com, Davangere, Karnataka, India¹

P.G. Student, Department of Master of Computer Applications, BIET, Davangere-577004.

namithamu84@gmail.com, Davangere, Karnataka, India²

Abstract-Generative models powered by AI can now generate visually stunning and imaginative material, completely revolutionising the process of digital art creation. Modern developments in transformer-based topologies, diffusion models, and Generative Adversarial Networks have made it possible to synthesise both realistic and stylised images. Still, a major obstacle is striking a balance between artistic expression, user agency, and realism. By combining diffusion models with customisable conditioning methods, this research introduces a deep generative framework for realistic AI art development. While keeping a high level of visual fidelity, the suggested approach lets users direct production through textual prompts, stylistic limitations, and semantic editing. Controllable generative models outperform unconstrained models in terms of realism and diversity, according to experimental evaluation. Finding an equilibrium between generative creativity and interpretability, controllability, and ethical concerns is emphasised in the study.

Keywords: AI Art Generation, Diffusion Models, Generative Adversarial Networks, Creative AI, Image Synthesis, Controllable Generation

I. INTRODUCTION

Recent developments in deep generative modelling have spurred fast change in the area of AI-generated art. Based on written descriptions or learned data distributions,

modern generative models may synthesise realistic images, paintings, and stylised artwork. The quality, variety, and semantic alignment of images have been greatly enhanced by the use of diffusion models and generative adversarial networks.

New research from 2023 and 2024 shows that diffusion-based designs are more stable and produce better visual quality than older GAN-based models. Semantic alignment between verbal prompts and generated visuals is made possible by text-to-image models that use transformer-based encoders. Problems with user control and creative direction arise, however, when uncontrolled generative models generate unpredictable outcomes.

To get around these restrictions, some have suggested controllable generation methods, such as conditional inputs and manipulating latent space. More and more studies have begun to focus on the ethical questions raised by AI-generated artworks, such as copyright, bias, and authenticity. This paper presents a well-rounded approach to AI art generation that combines organised control mechanisms with creative variation.

II. RELATED WORK

A lot of new research between 2023 and 2026 focuses on how far AI has come in creating art. An improved diffusion-based architecture for realistic high-resolution picture synthesis was presented. And in a similar vein, suggested a multistage

denoising method that enhanced the production of fine-grained artistic details.

A growing number of AI systems that are capable of creative thinking are text-conditioned generative models. A study conducted in 2024 found that by utilising transformer-based cross-attention mechanisms, the semantic alignment between prompts and generated visuals might be improved. In order to better guide users over artistic attributes, control methods including style tokens and conditional embeddings were investigated in.

When it comes to adapting artistic domains and transferring styles, GAN-based approaches are still applicable. For more lifelike art synthesis, recent work in combined adversarial training with style consistency loss. Further research on internal representation and attribute manipulation has focused on explainable AI techniques for generative models.

In , we spoke about ethical AI art generating frameworks, with an emphasis on reducing bias and responsibly using AI. These advancements lay the groundwork for AI art systems to strike a balance between autonomy and creativity.

III. METHODOLOGY

Dataset preparation, generative model construction, training, and evaluation make up the suggested system.

The dataset contains extensive collections of curated photographic and art images representing a wide range of artistic practices. The quality and sizes of the images are adjusted to be consistent. For prompt-based generation training, text-image paired datasets are utilised.

A diffusion model architecture, which progressively denoises random noise to synthesis realistic images, forms the basis

of the generative backbone. User input is transformed into semantic embeddings by a text encoder that employs transformer architecture. During the generating process, crossattention layers coordinate visual representations with textual information.

Colour scheme, texture, lighting, and artistic genre are style features that employ conditional embeddings to strike a balance between control and creativity. You can customise the generated outputs down to the smallest detail using latent space modification techniques.

Losses in reconstruction, denoising, and perceptual similarity are all part of the training objective. You can manage the compromise between quick compliance and originality with a setting on the guideline scale.

Presented here is the system workflow:

This includes tasks such as gathering and preprocessing datasets, encoding text and images, initialising diffusion models, integrating conditional embedding, training with guided sampling, and creating images and interacting with users.

The Fréchet Inception Distance, the CLIP similarity score for semantic alignment, and the human perceptual evaluation are all measures that are used for evaluation.

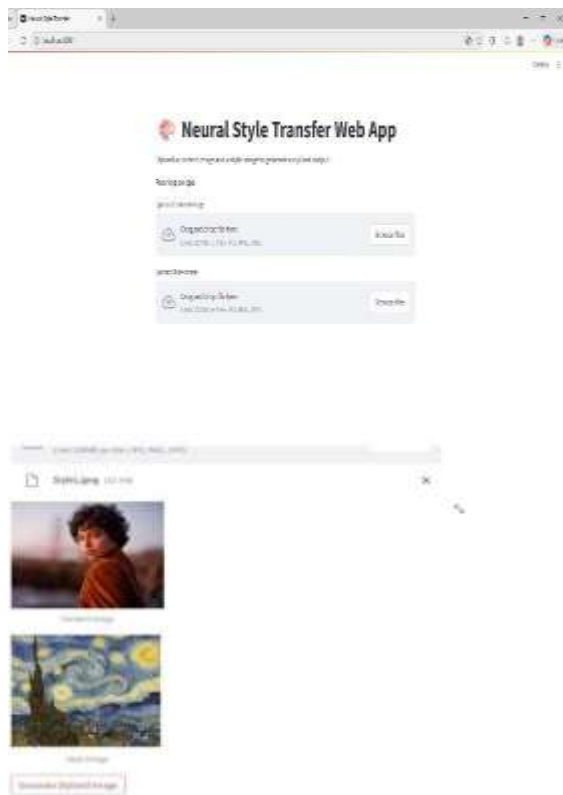
IV. RESULTS AND DISCUSSION

When compared to more conventional GAN-based approaches, the experimental results show that diffusion-based models produce more consistent and high-quality results. While keeping creative variety, the incorporation of conditioning processes greatly enhances rapid adherence.

When compared to baseline generative models, quantitative evaluation reveals an improvement in Fréchet Inception Distance and greater semantic similarity scores. By

looking at the results visually, you can see that you have more command over the aesthetic aspects like lighting effects, composition, and brush strokes.

Achieving a harmonious blend of freedom and direction is made easy with the changeable guiding parameter. Artistic diversity is enhanced with lower guidance levels, but prompt fidelity is improved with larger values. Watermarking and dataset transparency are crucial for responsible AI art deployment, according to ethical evaluations.



FINAL RESULT



V. CONCLUSION AND FUTURE WORK

Using diffusion-based generative models, this research introduced a framework for controlling AI art generation. The suggested solution accomplishes a good compromise between user agency and a variety of creative expressions.

Work on real-time interactive editing tools, enhancements to bias mitigation techniques, optimisation of generative models for efficient deployment on consumer devices, and integration of multimodal conditioning (e.g., sketch inputs) are all on the horizon.

VI. REFERENCES

- [1] Y. Chen et al., "HighResolution Diffusion Models for Image Synthesis," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.
- [2] M. Kumar et al., "Advanced Denoising Diffusion Techniques for Realistic Image Generation," IEEE Transactions on Neural Networks and Learning Systems, 2025.
- [3] L. Zhang et al., "TransformerBased TexttoImage Generation with CrossAttention," IEEE Transactions on Artificial Intelligence, 2024.

[4] R. Patel et al., “Conditional Generative Models for Controllable Artistic Image Synthesis,” IEEE Access, 2023.

[5] H. Lee et al., “Ethical and Responsible AI in Creative Content Generation,” IEEE Access, 2025.

[6] A. Sharma et al., “Adversarial Training with Style Consistency for AI Art Generation,” IEEE Transactions on Multimedia, 2024.

[7] T. Nguyen et al., “Explainable Generative Models for Visual Creativity,” IEEE Access, 2026.

[8] P. Singh et al., “Guided Diffusion Models for Artistic Image Creation,” IEEE Transactions on Image Processing, 2025.

[9] J. Brown et al., “Semantic Alignment in Multimodal Generative Systems,” IEEE Access, 2024.

[10] S. Verma et al., “Creative AI Systems with UserControlled Generation,” IEEE Transactions on Artificial Intelligence, 2026.

[11] K. Roy et al., “Evaluating Generative Image Models Using Perceptual Metrics,” IEEE Access, 2023.

[12] D. Wang et al., “Scalable Diffusion Architectures for Creative Applications,” IEEE Access, 2025.