

# Enhancing Artistic Style Transfer: Integrating CycleGAN, Diffusion Models, and Neural Painting for Monet-Inspired Image Generation

Yuening Li<sup>1,2,\*</sup>, Runxuan Yu<sup>1</sup>, Sparsh Amarnani<sup>1</sup>, Bhatt Sunny<sup>1</sup>, Yushan Xie<sup>1</sup>

<sup>1</sup>Department of Computer Science, School of Engineering, Vanderbilt University, TN, USA

<sup>2</sup>yuening.li@vanderbilt.edu

\*corresponding author

**Abstract.** This project explores various approaches to style transfer, specifically focusing on transforming natural images into the iconic style of Claude Monet's paintings. We explored three methods: CycleGAN-based, neural painting-based, and diffusion-based. CycleGAN enables style transfer between domains without paired training data, neural painting simulates the physical painting process, and diffusion models leverage a denoising process for high-quality results. To evaluate the effectiveness of these approaches, we conduct a survey assessing the aesthetic appeal, naturalness, and adherence to Monet's style of the generated images. Our analysis provides insights into the strengths and limitations of each method and identifies areas for future improvement in reproducing Monet's iconic style.

**Keywords:** Style Transfer, CycleGAN, Diffusion Models, Neural Painting, Image-to-Image Translation.

## 1. Introduction

Style transfer, a technique that applies the artistic style of one image to the content of another, has garnered significant attention in recent years. By leveraging the power of deep learning, style transfer allows for the creation of novel and visually appealing images that blend the essence of two distinct sources Gatys et al. [1]. This project focuses on exploring various approaches to style transfer, specifically in the context of transferring natural images into the iconic style of Claude Monet's paintings.

Claude Monet, a pioneer of the impressionist movement, is renowned for his distinctive brushwork, vibrant color palettes, and the way he captured the ephemeral qualities of light Heinrich [2]. His paintings, such as the famous "Water Lilies" series, showcase his mastery in depicting the interplay of light and color in natural scenes. By applying Monet's style to contemporary landscape images, we aim to create a unique fusion of past and present, allowing modern scenes to be reinterpreted through the lens of a celebrated artist.

To achieve this goal, we explored three distinct approaches to style transfer: CycleGAN-based, neural painting-based, and diffusion-based methods. CycleGAN Zhu et al. [3] is a powerful image-to-image translation framework that allows for the transfer of style between two domains without the need

for paired training data. Neural painting-based methods, such as the one proposed by Zou et al. [4] Zou et al. [1], aim to simulate the physical process of painting by generating brushstrokes on a canvas. Diffusion-based models, like the one introduced by Ho et al. [5] Ho et al. [6], leverage a denoising process to gradually refine an image, allowing for high-quality and diverse style transfer results.

To evaluate the effectiveness of these approaches, we conducted a survey to gather feedback from a diverse group of participants. The survey aimed to assess the aesthetic appeal, naturalness, and adherence to Monet's style of the generated images. By analyzing the survey results, we gained insights into the strengths and limitations of each approach and identified areas for future improvement.

## 2. Related Work

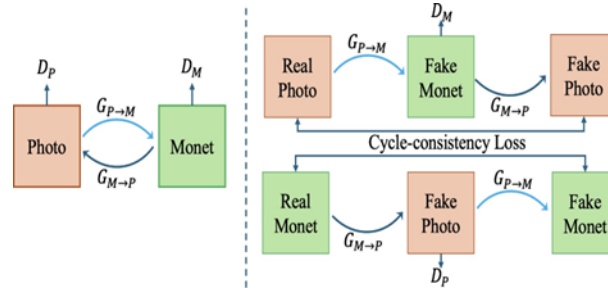
Image translation, designed to translate images from one domain (e.g., an actual photograph) to another (e.g., a work of art), has attracted much attention in recent years. GAN-based image translation, such as CycleGANs Zhu et al. [7], and their Variants Selim et al. [8] found in, e.g., Image synthesis, semantic editing, style transfer, etc. which has been applied to computer-generated artBojanowski et al. [9].

Style transfer methodologies produce novel images by applying the stylistic characteristics of a particular artist's work. This is achieved by rendering a content input image using stylistic elements extracted from an authentic artwork. The first attempt at style transfer relied on parameters and non-parametric methods for texture synthesis and transfer. Wei and Levoy [10] Huang et al. [11] Since the mid-1990s, the artistic theories behind attractive artworks have attracted the attention of artists and many computer science researchers. Much research and technology are exploring how to automatically convert images into synthetic artwork. Recently, inspired by the power of convolutional neural networks (CNNs), Gatys et al. [12] first studied how to use CNN recreates famous painting styles in nature picture. They suggest modeling the content of a photo as Feature responses from pre-trained CNN and further Modeling the style of artwork as summary feature statistics.

## 3. Approach

### 3.1. CycleGAN Model

CycleGAN, developed by Jun-Yan Zhu et al. [13], provides a solution to the challenge of transferring styles across domains without the need for paired training data, offering a way to convert images from one style domain to another with flexibility and ease. This model is composed of two generators,  $G_{P \rightarrow M}$  and  $G_{M \rightarrow P}$ , that work to enable a cyclical translation between the domains of photos and Monet-styled images. Complementing these are discriminators,  $D_P$  and  $D_M$ , which is to ascertain the authenticity of the photos and the Monet-styled images respectively. Illustrated in Figure 1, this setup promotes a cycle-consistency loss, which ensures that each image can be translated back to its original form, thus maintaining content integrity while altering style.



**Figure 1.** CycleGAN Structure

Properties The model's design to function without paired training data stands as its prime advantage, offering robustness for varied style transfer applications. However, there may be a trade-off, as CycleGAN may overemphasize style at the cost of losing some original content details, which can result in a slight departure from the intricate nuances characteristic of Monet's original works.

### 3.2. Style-Transferred outcomes

The style transfer efficacy of CycleGAN is visually demonstrated in Figure 2. The model adeptly captures and applies Monet's painterly style to contemporary landscape photographs, as shown.

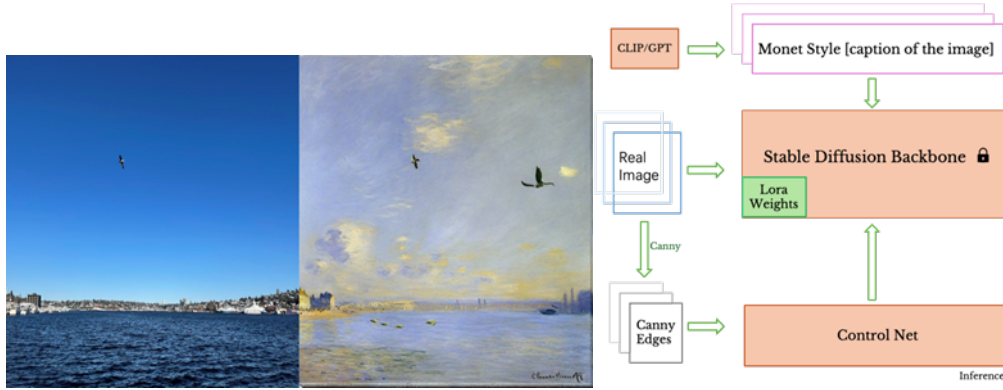
In the first set, the model transforms a rural scene into an impressionistic painting, where the vivid sunset and serene homestead are re-envisioned with Monet's characteristic brushstrokes and rich, pastel hues. The second set depicts a tranquil lakeside, where the transformation brings about a harmonious blend of sky and water, imbued with the luminosity and dynamic textures akin to Monet's approach to light and reflection. These results show that CycleGAN can replicate Monet's style and also add an impressionistic touch to modern photographs.



**Figure 2.** Style transfer using CycleGAN

### 3.3. Diffusion Model

The diffusion model approach in our project leverages several state-of-the-art techniques to achieve high-quality style transfer in the manner of Claude Monet's paintings. We utilize LoRA Hu et al. [14] to efficiently fine-tune the Stable Diffusion model Rombach et al. [15] using a dataset of real Monet images and their corresponding captions. This process allows the model to learn the specific style elements and themes present in Monet's artworks.

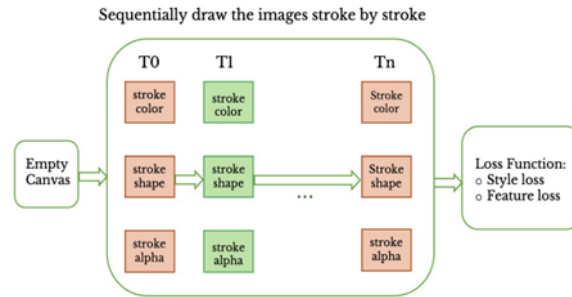


**Figure 3.** Style Diffusion Model (Photo → Monet) **Figure 4.** Overall inference pipeline

By leveraging these advanced techniques, the diffusion-based approach aims to generate high-fidelity style transfers that carefully balance the preservation of the original image's content with the artistic characteristics of Monet's paintings. The combination of LoRA-fine-tuned Stable Diffusion, ControlNet for structure guidance, and CLIP for content understanding allows for a comprehensive and nuanced style transfer process that accurately captures Claude Monet's artistic style. Figure 3 is an example transformation from a real photo into a synthesized Monet-style painting using Diffusion Model.

### 3.4. Stylized Neural Painting

Stylized Neural Painting is an advanced computational technique that employs artificial neural networks to replicate the painting process. Stylized Neural Painting Structure depicts the onset of a neural network's style transfer, commencing with an empty canvas. Iteratively, the network applies modifications to the stroke color, shape, and transparency, layering these elements to progressively infuse the canvas with the desired artistic style, transforming the initial void into a richly stylized image.



**Figure 5.** Stylized Neural Painting Structure

## 4. Experiments

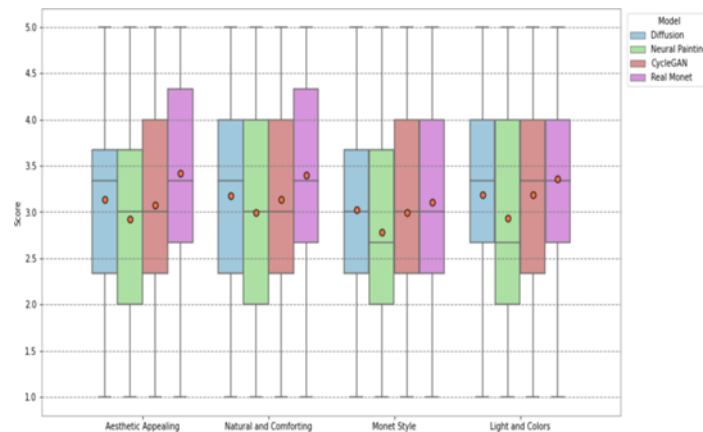
### 4.1. Survey

The primary objective of our survey was to gauge public perception regarding the effectiveness of deep learning models in artistically reinterpreting landscape photographs into the distinctive style of Claude Monet's paintings. We aimed to gather insights into the public's ability to differentiate between original Monet paintings and our AI-generated art, as well as to evaluate the perceived quality of the generated artworks. The survey was designed to be both online and in-person to maximize participant reach and diversity. The structure of the survey was as follows:

#### 4.1.1. Image Evaluation

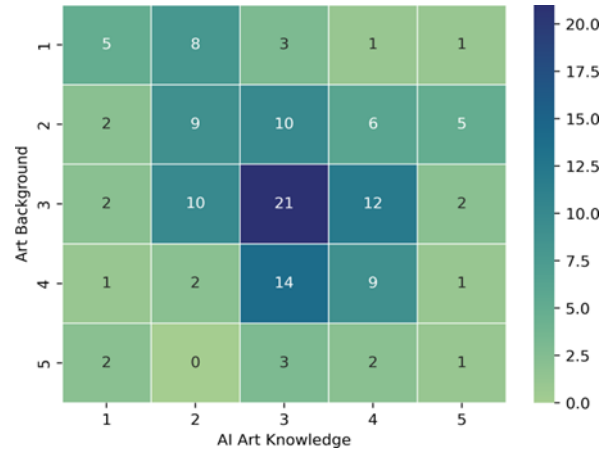
Participants were shown twelve images, a mix of authentic Claude Monet paintings and AI-generated artworks. Each image was evaluated based on the following criteria: Aesthetic Appeal: How aesthetically appealing do you find this image? Naturalness and Comfort: Do the colors in the image work together in a way that feels natural and comforting? Monet Style Adherence: Do you think the image is in Monet's style? Color and Light Interpretation: How well does the image capture a moment with its light and colors?

#### 4.1.2. Boxplot Analysis of Ratings Across Criteria



**Figure 7.** Boxplots of participant ratings for aesthetic appeal, color, style, and light capture

Regarding aesthetic appeal, both Monet’s originals and the Diffusion model closely align in median scores, indicating high visual satisfaction among viewers. The Diffusion model excels in matching the aesthetic quality of Monet’s paintings, likely due to its probabilistic approach, which enables diverse and appealing visual outputs. Especially, Monet maintains a higher average score, suggesting that the real artworks consistently resonate more with the viewers’ aesthetic standards. While CycleGAN and Neural Painting demonstrate competitive median scores, their lower mean scores suggest that further refinement of their algorithms could enhance their aesthetic appeal.



**Figure 8.** AI vs Art familiarity

As shown in the figure 8, a large concentration of our survey takers are located in the middle or slightly away from the center towards the left or up: this implies that people who took our survey were at least somewhat familiar with AI Art and had somewhat familiar with AI art and had some Art Background.

Our survey asked our participants on how likely they think of painting was Monet with the result and we saw that these questions revealed the most about our model particularly against people from different art background.

## 5. Conclusion

**Goals for this paper** Our goal with this paper to explore different style transfer models and analyze their artistic rendition capabilities. For the former we trained three models and for the later we conducted a survey asking users to rate their images produced by our model to the ones produced by new models.

**Results Evaluation** Our models exhibited impressive performance within our survey group, notably with Diffusion and CycleGAN receiving higher average ratings compared to genuine Monet paintings. Despite Neural Painter emerging as the least successful among our models, its performance remained competitive with the others. However, the survey results prompt important inquiries, as some outcomes couldn’t be attributed solely to model superiority. Notably, the Monet Style similarity performance across different models wasn’t predictable based on varying levels of art knowledge among respondents, as depicted in Figure 10. Additionally, we observed a broader range of scores, contrary to our expectations, as illustrated in our KDE graph in Figure 8. We anticipated similar ratings for images generated by comparable models, making this variance unexpected.

**Future Directions** We identify three promising avenues for future research: First, the integration of different model components to enhance performance. Our mentor, Dr. Daniel Moyers, proposed combining the color space from CycleGAN or Diffusion Model with the Differentiable Rendering Pipeline of Neural Painter. This approach could be beneficial as one pipeline appeared to take significantly longer to render an image but offered valuable insights into the painting process. Another suggestion from our mentor was to explore incorporating k-means clusters of CycleGAN images into

our Neural Painter as a prior to achieve more stylized colors. Additionally, integrating cyclic loss into our Neural Painter represents another viable direction.

While our survey primarily assessed the comparative performance of our models against Monet paintings, further investigation into model design could yield valuable insights. One potential approach is to disclose the model used to generate each image and solicit feedback on which aspects could be enhanced. This approach could provide valuable guidance for future research endeavors.

## References

- [1] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.
- [2] Christoph Heinrich. *Claude Monet: A Biography*. Prestel, 2022.
- [3] Piotr Bojanowski, Armand Joulin, David Lopez-Paz, and Arthur Szlam. Optimizing the latent space of generative networks. *arXiv preprint arXiv:1707.05776*, 2017.
- [4] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.
- [5] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12873–12883, 2021.
- [6] Teng Hu, Ran Yi, Haokun Zhu, Liang Liu, Jinlong Peng, Yabiao Wang, Chengjie Wang, and Lizhuang Ma. Stroke-based neural painting and stylization with dynamically predicted painting region. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 7470–7480, 2023.
- [7] Boyan Liu, Jiayi Ji, Liuqing Gu, and Ziheng Jiang. An integrated cyclegan-diffusion approach for realistic image generation. In *Fourth International Conference on Signal Processing and Machine Learning (CONF-SPML 2024)*, volume 13077, pages 92–101. SPIE, 2024.
- [8] Xu Ma. A comparison of art style transfer in cycle-gan based on different generators. In *Journal of Physics: Conference Series*, volume 2711, page 012006. IOP Publishing, 2024.
- [9] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*, 2015.
- [10] Haozhi Huang, Hao Wang, Wenhan Luo, Lin Ma, Wenhao Jiang, Xiaolong Zhu, Zhifeng Li, and Wei Liu. Real-time neural style transfer for videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 783–791, 2017.
- [11] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.
- [12] Ming-Yu Liu, Xun Huang, Arun Mallya, Tero Karras, Timo Aila, Jaakko Lehtinen, and Jan Kautz. Few-shot unsupervised image-to-image translation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10551–10560, 2019.
- [13] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR, 2021.
- [14] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.
- [15] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.