Monet Sytle Transfer

Songhee Beck, Kevin Cai, Satyabrat Srikumar, Sifan Tao

September 21, 2025

1 Introduction

The fusion of art and machine learning has opened new avenues for creative expression, enabling computers to replicate artistic styles with remarkable fidelity. Generative Adversarial Networks (GANs) have been instrumental in this advancement, particularly in the area of image style transfer. The project focuses on emulating the distinctive style of Claude Monet's paintings using GANs, inspired by the Kaggle competition "I'm Something of a Painter Myself." Our objective is to develop models that can transform landscape photographs into images that reflect Monet's impressionistic style.

To achieve this goal, we explore 4 models: **Deep Convolutional GAN (DCGAN)** as a baseline, followed by **Neural Style Transfer (NST)**, **CycleGAN**, and **CycleGAN with Least Squares GAN (LSGAN) Loss and ResNet Architecture**. The subsequent sections provide a detailed explanation of our methodology.

2 Dataset

The dataset is published in a Kaggle competition. The dataset contains four folders in total, including a folder consisting of 300 Monet paintings in JPEG format, a folder consisting of 7,028 landscape photographs in JPEG format, a folder containing 5 TFRecord files storing the 300 Monet paintings, and a folder containing 20 TFRecord files storing the 7038 landscape photographs.

3 Data Preprocessing through Data Augmentation

For training the **DCGAN**, we faced the challenge of having a limited number of Monet paintings (only 300 images) to work with. Since DCGANs require a large amount of data to effectively learn and generate high-quality images, we implemented an extensive data augmentation strategy to expand our dataset. We generated 100,000 images by applying various transformations to the original Monet paintings. All augmented images were resized to 64×64 pixels, due to computational constraints.

In contrast, for other models we chose not to use the image augmentation which we had proposed in the last deliverable 2. The decision was based on the observation that models trained with augmented images performed worse than those trained on the original dataset, which is reasonable since the FID score between augmented images and Monet paintings is around 100.

4 Methods

4.0.1 Deep Convolutional Generative Adversarial Networks (DCGAN)

DCGAN [Ink24] [RMC15] is implemented as our baseline where in the generator takes random noise vectors as input and generates images attempting to mimic Monet's painting style. The discriminator evaluates these images against real Monet paintings to classify them as real or fake. Our DCGAN architecture comprises convolutional layers with batch normalization and ReLU activations. To address the numerical instability often encountered during GAN training, we incorporated **spectral normalization** in the discriminator network. Spectral normalization stabilizes the training by constraining the spectral norms of the weight matrices, which helps prevent exploding gradients and improves convergence. By using this architecture and extensive data augmentation, the DCGAN learned to generate images that exhibit some characteristics of Monet's style. However, due to the complexity of Monet's artwork and limitations inherent in DCGAN architecture, the generated images were less detailed compared to those produced by the CycleGAN and NST models.

4.0.2 NST

Neural Style Transfer (NST) blends the content of a landscape photo with the artistic style of a Monet painting, leveraging a pre-trained VGG19 model to extract content and style features. Content features capture the high-level structure of the input photo, while style features reflect Monet's textures and patterns. Our NST process optimizes a target image by minimizing a loss function comprising Content Loss, which measures the difference in content features between the target and input images; Style Loss, which quantifies the difference in style features using Gram matrices; and Total Loss, a weighted sum of the content and style losses to balance structure and artistic transformation. To reduce runtime due to computational constraints, we processed 500 content images (out of 7,028), transforming each over 1,000 optimization steps. This process produced stylized images that reflect Monet's artistic style while preserving the structural integrity of the content images.

4.0.3 CycleGAN

CycleGAN is a kind of GAN supporting unpaired image-to-image translation with the help of its special architecture [ZPIE17][Nai20]. CycleGAN maps images between two domains, which are Monet paintings and landscape photos for this project, using two generators and two discriminators. One of the generators translates images from landscape photos to Monet paintings and another generator translates images from Monet paintings to landscape photos. The discriminators try to distinguish real images in each domain from generated images. By combining these four models with a cycle consistency loss, it ensures that when an image is translated to another domain and translated back to its original domain, the translated image is similar to the original image, making paired data unnecessary. Since it is good at unpaired image-to-image translation, we take advantage of this model to generate Monet-style images.

4.0.4 CycleGAN with LSGAN Loss

We also implemented a variant of CycleGAN [Lig] that replaces the traditional binary cross-entropy loss with Least Squares GAN (LSGAN) loss and adds identity loss to further improve output quality. LSGAN was chosen based on its demonstrated ability to provide more stable training by mitigating vanishing gradient problems and generate higher-quality images than regular GANs [MLX+16]. This implementation maintains two generator and two discriminator networks, using Mean Squared Error for adversarial loss and L1 loss for both identity preservation and cycle consistency. This combined approach enables more stable training while preserving input image characteristics.

4.1 Method for Model Evaluation

To find the best model for this problem, we use Frechet Inception Distance (FID) and Structural Similarity Index(SSIM) as our performance metric to compare the performance of each model. FID is used to check if the style of generated images is similar to Monet paintings while SSIM is used to check if generated images preserve the structural details in the original landscape photos.

5 Results

5.1 Performance of DCGAN

We evaluated the DCGAN by generating 5,000 images and computing the FID score between these images and the 300 original Monet paintings, obtaining an FID of 162. This high FID indicates that the DCGAN-generated images are less similar to real Monet paintings compared to those produced by the other models. The DCGAN struggled to capture the intricate details of Monet's style, possibly

	DCGAN	NST	CycleGAN	CycleGAN with LSGAN Loss
FID	162	127.98	114.89	101.97
SSIM	N/A	0.2796	0.7647	0.3966

Table 1: FID and Average SSIM score of Images Generated by Each Model

due to the model's architectural limitations and the complexity of Monet's art. Additionally, we did not compute the SSIM for DCGAN-generated images because SSIM requires a direct correspondence between generated images and content images, which is not applicable in DCGAN as it generates images from random noise without specific content reference.

5.2 Performance of NST

For the NST process, we calculated the FID between the 500 stylized images and the real Monet paintings. The FID score was 127.98, suggesting that while NST effectively captured Monet's artistic style, the limited diversity in style references and the optimization-based approach introduced some structural inconsistencies. Additionally, we applied NST to 1,000 content images to analyze the effect of dataset size on performance. While the FID score improved slightly from 127.98 to 127.63, the improvement was marginal compared to the significantly increased runtime. The results indicate that increasing the dataset size might improve FID, but the marginal gain in performance may not justify the added computational cost, especially given NST's image-by-image optimization process.

5.3 Performance of CycleGAN

After generating all 7038 Monet-style images corresponding to the 7038 content images using CycleGAN, we calculated the FID between the generated images and the real Monet paintings. The value of FID is around 120. The SSIM value between content images and generated images is 0.7647.

5.4 Performance of CycleGAN with LSGAN Loss

The FID value between generated images and real Monet paintings is 101.97, with an SSIM value of 0.3966 between input photos and generated images. These results show that the LSGAN modification achieves lower FID compared to the original CycleGAN (114.89), suggesting better quality and closer distribution to real Monet paintings. However, the SSIM value is lower (0.3966 vs 0.7647), indicating less structural similarity preservation compared to the original implementation. The lower FID likely results from LSGAN's smoother gradients and more stable training, while the lower SSIM suggests this implementation may prioritize style transfer over structural preservation despite the addition of identity loss. Further investigation seems necessary to understand this trade-off.

5.5 Performance Comparison

Shown in Table 1, our CycleGAN with LSGAN Loss achieved the best results, with the lowest FID score. Standard CycleGAN also performed well, with a high FID and the highest SSIM, demonstrating strong structural preservation. NST showed moderate performance but was limited by high computational costs and less effective style diversity. DCGAN, as the baseline, struggled with the complexity of Monet's art, resulting in the highest FID.

6 Conclusion

Our findings underscore the importance of model architecture and loss functions in achieving high-quality transfer of artistic style. Future work could focus on integrating the strengths of the approaches, using larger style datasets and incorporating advanced loss mechanisms to further enhance style fidelity and computational efficiency. This study highlights the transformative potential of machine learning in creative expression, paving the way for more refined and versatile applications in art and technology.

References

[Ink24] Nathan Inkawhich. Pytorch dcgan tutorial. 2024. Accessed: 2024-12-07.

[Lig] Lightning.ai. Pytorch lightning cyclegan tutorial.

[MLX+16] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. *arXiv preprint arXiv:1611.04076*, 2016.

[Nai20] Aakash Kumar Nain. Cyclegan. 2020.

[RMC15] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015. Under review as a conference paper at ICLR 2016.

[ZPIE17] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017.