IE 7615

NEURAL NETWORKS AND DEEP LEARNING

FINAL PROJECT - PAPER

TITLE:

Artistry Mimicking
Painting Image Generation infused with artists styles using GANs

Authors:

Patlannagari, Akshara Reddy, patlannagari.a@northeastern.edu
Gandikota, Ruchirkanth, gandikota.r@northeastern.edu
Mathur, Shubham, mathur.shu@northeastern.edu

Abstract

In this paper we present a CycleGAN model aimed at performing image to image translation between unpaired datasets of Monet paintings and photorealistic images. The goal is to explore the boundary between art and artificial intelligence, evaluating the capability of GANs to emulate the nuanced brushstrokes and color palettes characteristic of Monet's work. The approach leverages Tensorflow and Keras on a TPU infrastructure, focusing on efficient data processing and model training. Our work demonstrates a practical approach to leveraging deep learning for artistic style transfer, providing insights into the model's architecture and training dynamics.

1. Introduction:

This project focuses on using CycleGANs to create images in the style of Monet, which builds on traditional image-to-image translation algorithms. Through this endeavor, we hope to investigate the convergence of machine learning and art, demonstrating the potential of neural networks in creative style transfer challenges.

The motivation for this project arises from a desire to democratize art creation and investigate novel means of artistic expression. Our goal with CycleGANs is to enable artists, regardless of their level of experience, to obtain a greater knowledge of renowned artists' works, like Monet's, by casually discovering and immersing themselves in their individual artistic approaches. This initiative also helps advance the field of computer vision and machine learning by pushing the boundaries of what is possible in automated image generation.

Our approach centers on the utilization of CycleGANs, a type of generative adversarial network specifically designed for unpaired image-to-image translation tasks. Unlike traditional methods that rely on paired training data, CycleGANs can learn to perform style transfers directly from unpaired image sets, offering greater flexibility and scalability. By training the model on a diverse dataset encompassing both Monet paintings and photographic images, we aim to capture the essence of Monet's impressionistic style and transfer it onto real-world photographs.

For our experiments and evaluation, we utilized a dataset comprising a collection of Monet paintings as well as a diverse set of photographic images. This dataset provides the necessary diversity and richness required for training the CycleGAN model to effectively perform style transfers. Through rigorous experimentation and evaluation on this dataset, we aim to assess the model's ability to generate high-quality Monet-style images while maintaining the fidelity of the original content.

2. Background:

The evolution of image style transfer began with the introduction of Convolutional Neural Networks (CNNs), which initially faced challenges related to resolution and noise in the generated images impacting scalability and realism [3]. This was significantly improved with the advent of Generative Adversarial Networks (GANs), which brought about a leap in the quality of generated imagery influencing subsequent advancements in image translation technology [1]. A notable advancement within this domain is Pix2Pix [4], a conditional GAN framework that conditions the generation of output images on given input images, while a discriminator evaluates the realism of the transformation.

Building on these ideas, CycleGAN emerged [5], refining the concept by enabling image-to-image translation without needing paired examples, thus widening the application scope in unsupervised learning scenarios.

Further advancements include AttentionGAN (AttnGAN) [6], which integrates attention mechanisms into the GAN architecture. This modification enhances the model's ability to perform fine-grained text-to-image synthesis, allowing for the creation of detailed and contextually appropriate images from textual descriptions, which is particularly advantageous for artistic and creative tasks.

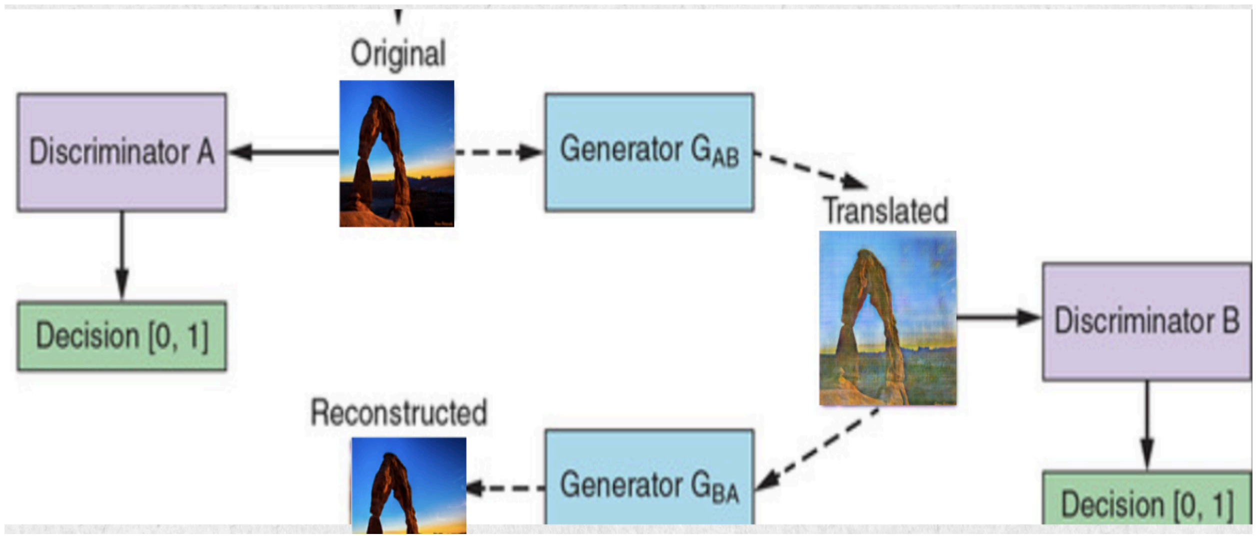
Another innovative development is SinGAN (Single Image GAN) [7], designed for high-quality image generation from a single input image. This model excels in generating a variety of outputs based on one input, proving extremely beneficial for creative and artistic purposes where extensive datasets are not available.

3. Approach:

The core of our methodology revolves around employing CycleGAN, a specialized variant of GANs, to tackle the complex challenge of unpaired image-to-image translation. This approach is particularly well-suited for artistic applications where paired examples (pre- and post-transformation images) are scarce or nonexistent.

CycleGAN architecture:

Our model is composed of two generators and two discriminators. Each generator learns to translate images from one domain to the other, while each discriminator learns to distinguish between translated images and real images from its domain.



Key Loss Functions

- Adversarial Loss: Ensures that the generated images by both generators are indistinguishable from real images in their respective domains. This is defined as:

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x)))]$$

- Cycle Consistency Loss: Compels a source image, when translated to the target domain and back, to resemble the original image. This loss is crucial for learning accurate and meaningful translations without paired examples.

$$\mathcal{L}_{\text{cycle}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1]$$

The model underwent training across 10 epochs, each encompassing a full cycle through the dataset. Our CycleGAN architecture integrates crucial upsampling and downsampling techniques to efficiently transform images. In the discriminator, strided convolution layers perform downsampling, reducing spatial dimensions to concentrate on critical high-level features essential for differentiating between real and generated images. This streamlined feature summarization is vital for effective discrimination.

In contrast, the generator employs upsampling, beginning with a latent noise vector and progressively expanding this into detailed images that replicate the style and content of the target domain. This expansion is facilitated through transposed convolution layers, which gradually enhance resolution and detail, crucial for crafting textural and structural complexity in the image output. This process ensures the generated images are convincingly realistic, matching actual photographs in style and resolution.

Throughout the training, our primary goal was to reduce losses associated with the optimization process. By continuously refining the model parameters, we aimed to improve the quality of the generated images and enhance the overall efficacy of the CycleGAN system.

4. Results:

4.1 Dataset:

The dataset contains four folders: monet_tfrec, photo_tfrec, monet_jpg, and photo_jpg.

- monet_jpg - 300 Monet paintings sized 256x256 in JPEG format
- monet_tfrec - 300 Monet paintings sized 256x256 in TFRecord format
- photo_jpg - 7028 photos sized 256x256 in JPEG format
- photo_tfrec - 7028 photos sized 256x256 in TFRecord format

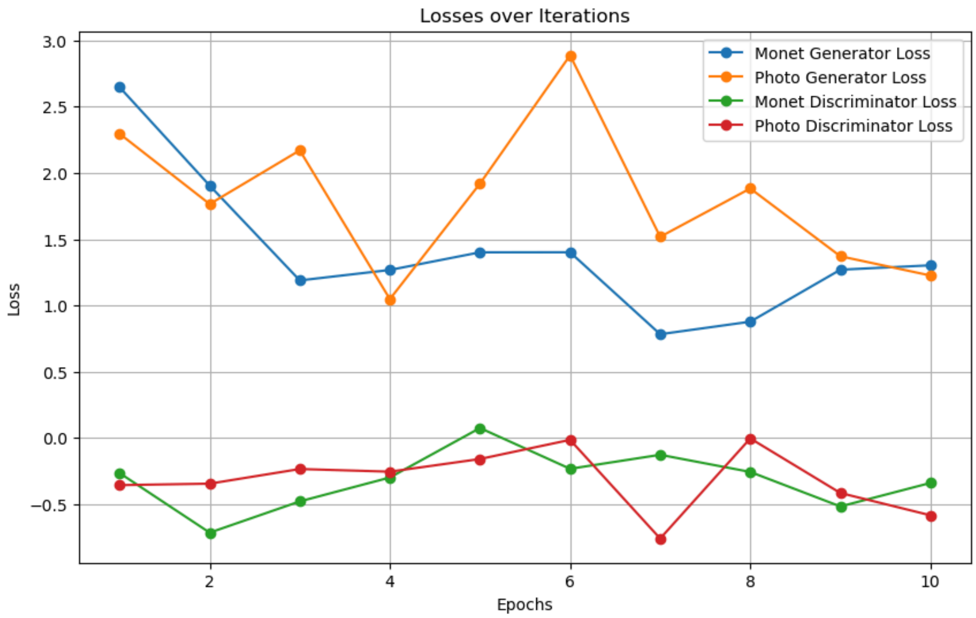
4.2 Experiments and performance evaluation:

During our experimental phase, we meticulously monitored the evolution of loss functions to understand the training dynamics of our CycleGAN model. Our focus centered on two primary loss metrics: adversarial loss and cycle consistency loss, as tracked across different components of the model.

- Adversarial Loss Analysis: Our analysis of adversarial losses provided deep insights into the learning progression of our model's generators. Specifically, we observed a consistent downward trend in the adversarial losses for both generators, with the Monet generator loss recorded at 2.3780 and the photo generator loss at 2.5529. This decline in adversarial loss across both generators indicated their enhanced ability to produce authentic-looking images that closely resemble the style and attributes of the target domains.
- Cycle Consistency Loss Examination: The cycle consistency loss also provided critical data on the model’s performance. With the adversarial losses for the Monet discriminator at 0.6335 and the photo discriminator at 0.5892, there was a clear indication of the discriminators' improving efficiency in distinguishing between real and generated images. The reduction in these values over time demonstrated the model's growing proficiency in maintaining the core content of images throughout the transformation cycle—from Domain A to Domain B and back. This capability ensured that the generated images not

only adopted the stylistic characteristics of the target domain but also retained the essential features of the original images.

These measurements have been instrumental in assessing the model's ability to generate high-quality, diverse images while preserving the integrity of the original content through each cycle of transformation.



Performance Evaluation Using FID

We used the Fréchet Inception Distance (FID) as our main metric to thoroughly assess the caliber of the images produced by our CycleGAN model. FID, which measures the separation between feature vectors derived by the Inception network from both generated and actual images, is a commonly accepted benchmark for assessing GAN performance. To calculate the FID score, features are first extracted from an intermediate layer of both the generated (g) and real (r) pictures using the Inception network. The distribution of each set is then modeled using these attributes as a multivariate Gaussian with means (μ) and covariances (Σ). The following formula yields the FID score:

$$\mathrm{FID} = ||\mu_r - \mu_g||^2 + \mathrm{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2})$$

where Tr represents the trace, summing the diagonal elements of a matrix. This analysis effectively measures the likeness between the generated and real image distributions, where a lower FID signifies fewer disparities, implying superior image quality. Utilizing FID as our metric enabled us to evaluate both visual fidelity and statistical similarity, providing a holistic assessment of our model's capacity to generate authentic images. We achieved a remarkably low

FID score of 8.355 for Monet-style paintings reflecting the model's exceptional ability to capture artistic nuances. Conversely, the higher FID score of 20.435 for photorealistic images indicates a slightly lower but still acceptable resemblance to the original photos.

4.3 Discussion:

Our project successfully showcased CycleGANs' capacity to produce Monet-style images, yielding a dataset of five excellent images. This accomplishment not only demonstrates how well CycleGANs can perform intricate stylistic modifications without paired training data, but it also opens up new avenues for machine learning-based artistic picture creation. The application of CycleGANs to style transfer problems facilitated a great deal of creativity and successfully captured and replicated Monet's unique impressionistic style.

Broader Context

The results of this project contribute to the growing body of knowledge in neural style transfer, a discipline that has significant ramifications for both computer vision and the art world. By automating the artistic style transformation process, we open up new avenues for artists and designers to experiment with digital media. Moreover, these technologies can be utilized in educational settings to provide students and enthusiasts with a deeper understanding of art styles and history through interactive experiences.

Recommended Future Directions

To build upon the success of this project, we propose two main directions for future research:

- Integration of Text-to-Image Capabilities: We can extend this work by developing a system that can generate Monet-style paintings directly from textual descriptions. This advancement would involve integrating natural language processing with image generation, allowing users to create custom artworks by simply describing scenes in text.
- Expansion to Other Artistic Styles: Another direction is to apply our methodology to a broader range of artistic styles. By training CycleGANs on different sets of artworks, we could facilitate a versatile tool capable of mimicking various historical and contemporary styles.

These enhancements will not only advance the technology but also broaden its applicability and impact, potentially transforming how we interact with and understand art in the digital age.

5. Conclusion:

The project successfully demonstrated the application of CycleGANs to create 5 images in the style of Monet, underscoring the model's ability to perform complex artistic style transfers. By utilizing CycleGANs, we sidestepped the need for paired training data, directly learning from unpaired images to effectively mimic Monet’s impressionist technique.

The outcomes of this study affirm the potential of CycleGANs not just in replicating artistic styles but also in broadening the scope of digital art creation. Our work illustrates the convergence of art and artificial intelligence, presenting new opportunities for creative expression and the exploration of historical art styles through a modern lens.

The key takeaway from our project is the transformative power of machine learning in art, highlighting its role in both preserving traditional artistic heritage and in making the art-making process more accessible to non-artists.

6. References:

[1] A. Singh, V. Jaiswal, G. Joshi, A. Sanjeeve, S. Gite and K. Kotecha, "Neural Style Transfer: A Critical Review," in IEEE Access, vol. 9, pp. 131583-131613, 2021, doi: 10.1109/ACCESS.2021.3112996.

[2] Goodfellow, Ian, et al. "Generative adversarial networks." *Communications of the ACM* 63.11 (2020): 139-144.

[3] L. A. Gatys, A. S. Ecker and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2414-2423, doi: 10.1109/CVPR.2016.265.

[4] P. Isola, J. -Y. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 5967-5976, doi: 10.1109/CVPR.2017.632.

[5] J. -Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2242-2251, doi: 10.1109/ICCV.2017.244.

[6] T. Xu et al., "AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 1316-1324, doi: 10.1109/CVPR.2018.00143.

[7] Q. Xu, J. Chen, J. Tang, Q. Kang and M. Zhou, "Performing Effective Generative Learning from a Single Image Only," 2023 32nd Wireless and Optical Communications Conference (WOCC), Newark, NJ, USA, 2023, pp. 1-5, doi: 10.1109/WOCC58016.2023.10139746.