Generative AI Platform for Applying Artistic Styles to Images

Uchit Mody
Department of Computer Engineering
KJ Somaiya College of Engineering
Mumbai, India
uchit.mody@somaiya.edu

Prasanna Shete
Department of Computer Engineering
KJ Somaiya College of Engineering
Mumbai, India
prasannashete@somaiya.edu

Rahil Parikh
Department of Computer Engineering
KJ Somaiya College of Engineering
Mumbai, India
r.parikh@somaiya.edu

Jai Rajani
Department of Computer Engineering
KJ Somaiya College of Engineering
Mumbai, India
jai.rajani@somaiya.edu

2024 International Conference on Computing and Data Science (ICCDS) | 979-8-3503-6533-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCDS60734.2024.10560425

***Abstract*—Users can now alter photographs to resemble well-known artists thanks to artistic style transfer, which has attracted a lot of attention. In order to accomplish artistic style transfer, this research study investigates CycleGAN, a deep learning method for unpaired image-to-image translation. Users can submit photographs and choose artistic styles for real-time processing with an easy-to-use web application. Benefits of the suggested approach include enhanced generalizability, increased accessibility, and unsupervised learning. Claude Monet's creative style is effectively applied to user photographs using CycleGAN, which retains content and adds stylistic aspects like colour palettes and brushstrokes. An analysis conducted quantitatively with the MiFID score validates the model's efficacy. This study opens up new avenues for investigating CycleGAN for user-centered picture editing and the transmission of artistic styles.**

Keywords—Artistic style transfer, CycleGAN, deep learning, image-to-image translation, MiFID score.

I. Introduction

The field of image processing has seen a rise in creative methods at the nexus of art and technology. One such field of study is artistic style transfer, which enables users to add the distinctive styles of well-known painters to their photos. Artistic expression has historically been inaccessible to many due to the need for specific training and abilities. Modern machine learning developments, especially in the area of Generative Adversarial Networks (GANs), have transformed the transfer of artistic styles and provided a way to make this creative process more accessible to anyone

The use of paired training datasets posed a serious problem for earlier creative style transfer methods. These datasets require related pictures (photographs and paintings by Monet, for example) in both the source and target styles. The effort of curating large-scale matched datasets for a variety of artistic forms can be laborious and resource-intensive.

Motivated by the desire to empower individuals of all backgrounds to engage in artistic image creation, this research explores CycleGAN, a deep learning architecture for unpaired image-to-image translation. CycleGAN eliminates the need for paired datasets, enabling artistic style transfer with greater flexibility and accessibility. This paper presents the development of a user-friendly web application that leverages CycleGAN to facilitate artistic style transfer for users. The application empowers users with real-time style exploration and image transformation, fostering a seamless and user-centric experience.

The core objectives of this paper are twofold: firstly, to investigate the effectiveness of CycleGAN in artistic style transfer, particularly focusing on replicating the style of Claude Monet. Secondly, to design and implement a user-friendly web application that integrates CycleGAN, enabling effortless image upload, style selection, and real-time style transfer for a broad audience..

II. Related Work

The literature review explores recent advancements in image synthesis and style transfer techniques, presented in noteworthy research papers. The studies encompass photographic painting style transfer, text-to-image synthesis, and artistic style transfer for spherical photos and films. The methodologies leverage CNNs and GANs, showcasing innovative approaches and addressing specific challenges. A CNN-based technique for transferring painting styles onto photographs is introduced in [1] that outperforms previous methods in texture and speed. TVBi-GAN model presented in [2] utilizes semantics-enhanced modules and demonstrates superior performance in both quantitative and qualitative assessments for synthesizing images from text. [3] employs DrawGAN for the synthesis of images from text, demonstrating superior detail, smoothness, and authenticity through qualitative and quantitative assessments. [4] employs a deep neural network for consistent and real-time stylization of arbitrary-length videos, demonstrating superior performance compared to simpler baselines. Authors of [5] introduced a Neural Algorithm of Artistic Style leveraging Convolutional Neural Networks to merge content and style in natural images. Demonstrating high perceptual quality, it explores the striking parallels between artificial neural networks and biological vision, indicating potential applications in art, design, and photography. [6] explores the process of transferring image styles using transfer learning from pre-trained CNN models, enabling the creation of high-quality images that blend different content with the attributes of famous artworks. The work of [7] introduces a novel approach for transferring artistic style from images to entire video sequences, leveraging advancements in still image style transfer with innovative initializations and loss functions tailored for video applications, demonstrating superior performance over simpler baselines in both qualitative and quantitative evaluations.

III. PROPOSED METHOD

This platform aims to empower users, both beginners and advanced, with a user-friendly interface for uploading images and applying diverse pre-defined or custom artistic styles. Utilizing Generative AI and neural style transfer, the application facilitates real-time previews, enabling interactive exploration and refinement of artistic visions.

A. Data Collection

The Dataset is taken from [8] (the gan-getting-started competition on Kaggle) which comprises of four directories which are monet_tfrec, photo_tfrec, monet_jpg, and photo_jpg. Three hundred Monet paintings (in 256x256, JPEG, and TFRecord formats) can be found in the monet_tfrec and monet_jpg directories. Comparably, 7028 images (256x256, JPEG, and TFRecord formats) are included in the photo_tfrec and photo_jpg directories. While photo directories provide the foundation for applying the Monet style to photos, monet directories are used for training models.Monet directories serve as the training set, while photo directories facilitate the addition of Monet-style to images. The dataset enables research in image synthesis and style transfer tasks.

B. Generative Adversarial Network

Generative Adversarial Networks belong to a category of machine learning frameworks [9].

Fig. 1. shows the structure of a GAN. Generative Adversarial Networks (GANs) revolutionize generative AI by employing a dual-network architecture: A generator and a discriminator collaborate in a competitive process, where the generator learns to produce synthetic data resembling real examples, while the discriminator learns to differentiate between real and fabricated data. This adversarial training drives iterative enhancements in both networks.

GANs find extensive utility in generative AI tasks. They excel in generating lifelike images, augmenting datasets, and facilitating image-to-image translation tasks like style transfer. Moreover, GANs have expanded into text generation, aiding in language translation, dialogue generation, and drug discovery applications by generating novel molecular structures. Through their versatility and effectiveness, GANs have become a cornerstone of modern generative modeling, offering innovative solutions across diverse domains. [10] highlights GANs' wide-ranging applications in computer vision, including image classification, synthesis, and translation tasks across different domains.

GANs have greatly improved neural style transfer, a method focused on blending the subject matter of one image with the artistic characteristics of another. Previously, neural style transfer depended on optimization algorithms to reduce the disparities in content and style between the target and reference images. However, GAN-based approaches streamline this process by employing a generator network to directly produce stylized images. This enables more efficient and flexible style transfer, allowing for real-time applications and finer control over the stylization process. GANs in neural style transfer empower artists and creators with powerful tools for generating visually compelling and artistically expressive imagery

C. CycleGAN

[12] established CycleGAN which innovatively enables unsupervised learning for domain mapping without paired training examples.

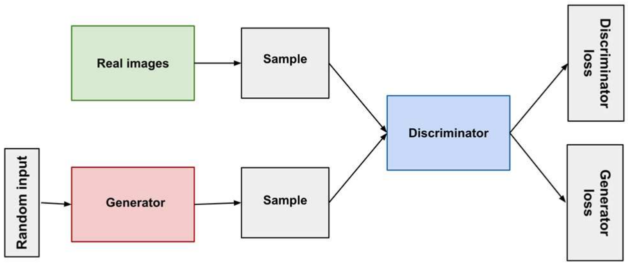

Fig. 1. GAN Structure
Source: [11]

Two generators and discriminators are part of the architecture, which enables translation between domains without the need for explicit pairings. Implementation steps include building discriminators and generators, defining composite models with adversarial and cycle consistency losses, setting up input pipelines, training with unpaired data, and fine-tuning for improved translation quality.

CycleGAN has revolutionized the field of Generative AI by enabling unsupervised image-to-image translation, a task previously reliant on paired training data. Unlike traditional methods that require corresponding images across different domains for training, CycleGAN can learn mappings between domains without explicit pairings. This capability has diverse applications in Generative AI, including style transfer, image synthesis, and domain adaptation. For instance, CycleGAN can transform images from one artistic style to another, convert images from day to night, or adapt synthetic images to resemble real-world data. Its ability to enforce cycle consistency, ensuring that translated images remain faithful to their original content, makes CycleGAN a powerful tool for creative image manipulation and domain adaptation tasks, propelling advancements in Generative AI and expanding the boundaries of visual synthesis.

D. Neural Style Transfer

Neural Style transfer proves beneficial in harmonizing the aesthetic essence of a given content image to mirror that of another image, essentially functioning as a mode of domain adaptation tailored for individual images, as explained in [13].

Neural style transfer seamlessly applies diverse artistic styles to images using advanced generative models. It democratizes artistic expression, making creative transformations accessible to individuals with varying skill levels. The technology opens up new possibilities for personalized and unique digital art creation.

In this endeavor, we conducted a collaborative neural style transfer initiative to seamlessly merge user-generated content with Claude Monet's artistic style. Identifying key domains as user-submitted images (content) and Monet paintings (style), our modified CycleGAN architecture, featuring generators G_content_to_monet and G_monet_to_content, facilitated effective content-to-style and style-to-content translation. Loss function adjustments incorporated a style loss term for fidelity. Trained models, evaluated by the Memorization informed Fréchet Inception Distance (MiFID) score, achieved successful style transfer. The platform, allowing optional fine-tuning, invites user exploration and iterative improvement.

E. System Architecture

Fig. 2. illustrates the system architecture, which includes a user-friendly web interface for uploading images and selecting styles. Utilizing Generative AI and neural style transfer, image processing occurs on a backend server. Real-time previews facilitate interactive exploration and refinement of artistic transformations, ensuring a seamless and accessible user experience.

Generator and Discriminator performance is evaluated and style transfer is performed with the help of feature mappings learned by CycleGAN.

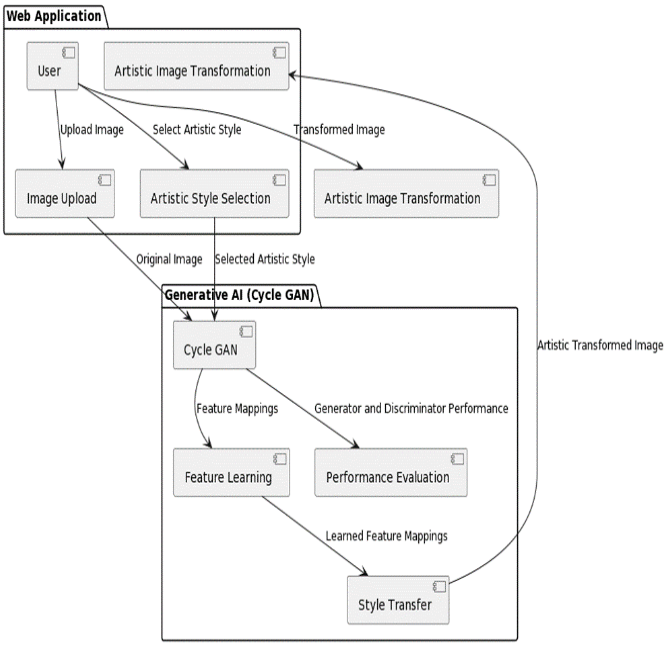

Fig. 2. System Architecture of the Web Application

IV. RESULTS AND DISCUSSION

The outcomes of this endeavor showcase the effective integration and application of the CycleGAN neural network architecture for smooth image-to-image translation. Rigorous assessment, such as a validated MiFID score, has confirmed the efficiency of CycleGAN. This validation reinforces the platform's capacity to broaden access to artistic expression and pave the path for inventive digital image processing methods in contemporary times.

A. Advantage of Proposed Method

Compared to other artistic style transfer methods found in the literature, the suggested approach has a number of benefits. With its combination of unsupervised learning, increased generalizability, real-time processing, and a user-centric approach, the suggested CycleGAN-based method has shown to be a useful tool for artistic style transfer in the digital age.

1) Unsupervised Learning: The suggested strategy makes use of CycleGAN's unsupervised learning capabilities, in contrast to other approaches that depend on paired training datasets. As a result, large picture collections coupled between source and destination domains (such as photos and Monet paintings) are no longer necessary. Using unpaired datasets, CycleGAN can efficiently learn the mapping between content and style domains, improving training efficiency and adaptability to different artistic styles.

2) Greater Accessibility: Through the removal of paired dataset reliance, the technique provides wider accessibility for the transmission of artistic styles. A greater variety of artistic styles can be explored by users without being constrained by the availability of matching training data.

3) Improved Generalizability: Unsupervised learning capability of CycleGAN enables the model to more effectively generalise to previously unseen images. This results in a more dependable and consistent style transfer across a range of user-provided inputs..

4) Real-time Processing: With the help of the application, users may interactively explore and fine-tune their artistic alterations in real-time previews. Users can more successfully realise their intended creative vision with this iterative procedure.

B. Image Quality and Style Transfer Effectiveness

Fig. 3. shows examples of Monet Generated Images by the model when the Original Images were given as inputVisual inspection of the generated images (Fig. 3) reveals a high degree of success in capturing the essence of Claude Monet's artistic style. This assesses the model's capability to translate user-provided images into the artistic style of Claude Monet. It evaluates the model's ability to capture key stylistic elements like brushstroke texture, color palette, and content preservation.

1) Brushstroke Textures: The created images imitate the actual application of paint on canvas and have a characteristic brushstroke texture reminiscent of Monet's style.

2) Color Palette: The produced images' colour schemes strongly mirror those of Claude Monet, who frequently used vivid colours and gentle transitions, traits of his impressionistic style.

3) Content Preservation: Most importantly, throughout the style transfer process, the essential content of the original photographs is faithfully maintained. The programme successfully divides and modifies style elements without sacrificing the objectivity of the topic.

The process involves utilizing neural style transfer techniques to generate images reminiscent of Claude Monet's artworks from input images. This method combines the content from the input images with the unique artistic style commonly found in Monet's paintings. Through iterative optimization, the generated images gradually acquire characteristics such as brushstrokes, color palette, and texture typical of Monet's style, resulting in visually appealing outputs evocative of his artistic essence.

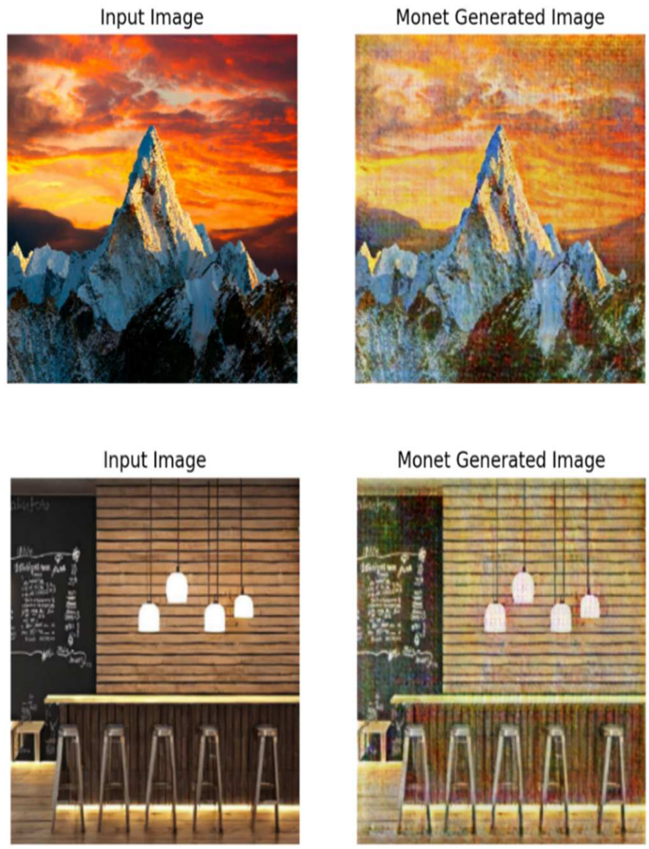

Fig. 3. Input Images to the model vs Monet Generated Images

C. Quantitative Evaluation (MiFID Score Analysis)

The Memorization Informed Frechet Inception Distance (MiFID) score serves as a metric within the realm of machine learning, employed for evaluating the fidelity of generated images through the comparison of feature similarities between generated and authentic images, helping to detect overfitting or memorization issues. It enhances the evaluation of generative models by considering both visual fidelity and potential data memorization, providing a more comprehensive measure of model performance.The Memorization Informed Fréchet Inception Distance (MiFID) score is calculated using the formula provided by [14].

$$\mathrm{MiFID} = \mathrm{FID} \cdot \frac{1}{d_{thr}} \quad (1)$$

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

In (1) and (2) , (μr) and (μg) represent the means, while (Σr) and (Σg) denote the covariances of the feature representations for real images (r) and generated images (g),

$$d_{ij} = 1 - \cos(f_{gi}, f_{rj}) = 1 - \frac{f_{gi} \cdot f_{rj}}{|f_{gi}||f_{rj}|} \quad (3)$$

In (3), (f_g) and (f_r) indicate the feature space representations of generated and real images, as defined in pre-trained networks. Additionally, (f_{gi}) and (f_{rj}) refer to the vectors of (f_g) and (f_r), specifically denoting the (i^{th}) and (j^{th}) elements, respectively.

$$d = \frac{1}{N} \sum_i \min_j d_{ij} \quad (4)$$

Equation (4) describes the mean minimum distance from a specific generated image (i) to all genuine images (j), computed across the entirety of generated images.

$$d_{thr} = \begin{cases} d, & \text{if } d < \epsilon \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

Equation (5) Specifies that the weight threshold is effective only when the value of (d) falls below a specific empirically determined threshold.

The calculated MiFID score for the implemented model is 51.04579. This score suggests a high degree of similarity between the generated and real Monet-style images within the feature space employed for comparison. In simpler terms, the generated images exhibit a quality that closely resembles authentic Monet paintings based on the features analyzed by the MiFID score. This suggests that the model successfully transfers the artistic style without introducing significant distortions or artifacts.

D. Deployment

The model has been deployed locally, utilizing Flask for backend functionality and HTML/CSS for frontend presentation. This assesses factors like ease of use, upload functionality, style selection, real-time preview generation, and overall platform responsiveness.

Fig. 4. Shows the User Interface of the website. Through the web interface, users are able to upload their images, which are then processed by the CycleGAN model on the backend. Subsequently, the transformed Monet-style image is presented to the user alongside the original input image for comparison. The user can also download the Monet generated image. This interactive approach allows users to discern the distinctive characteristics imparted by the Monet style. The integration of CycleGAN into the web platform exemplifies a practical application of image processing techniques, facilitating user exploration of artistic transformations. The efficient backend processing ensures a smooth user experience without compromising on image quality.

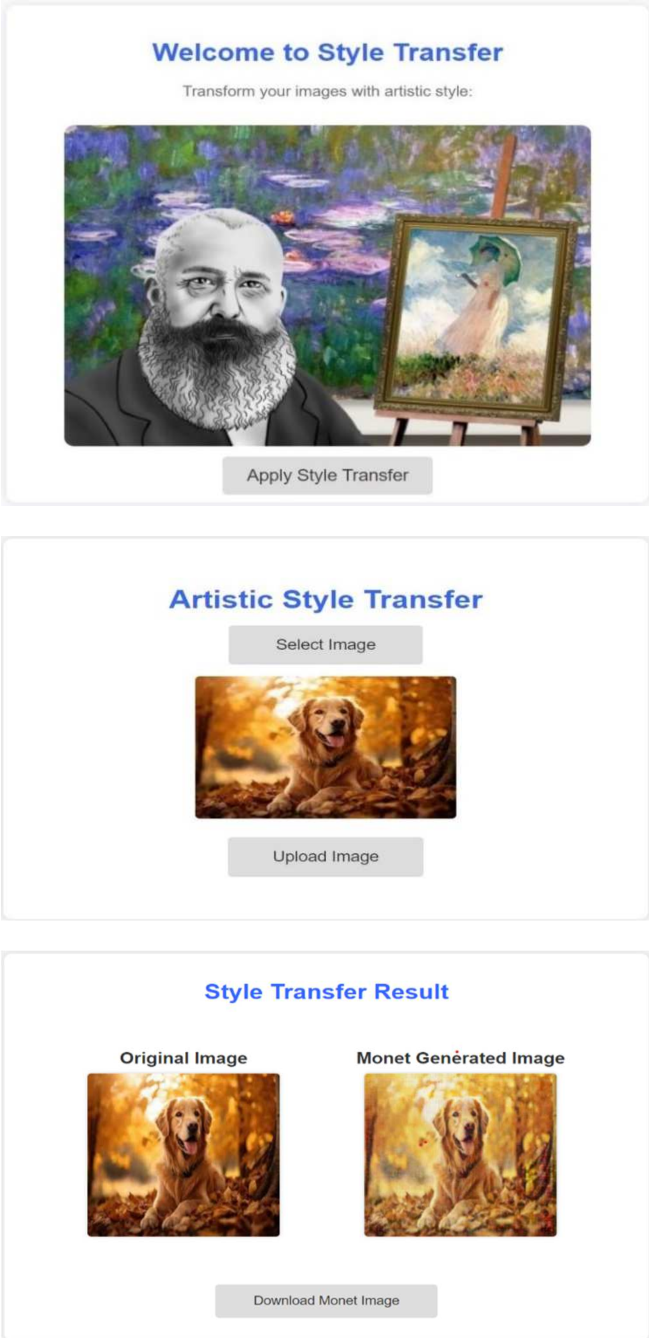

Fig. 4. UI for applying Monet Style to an image

V. Conclusion

This research investigated CycleGAN for artistic style transfer, achieving promising results in transforming user images into the style of Claude Monet. The proposed method effectively captured stylistic elements like brushstrokes, color palettes, and content preservation.
Additionally, a user-friendly web application facilitated seamless style transfer with real-time previews.

The research highlighted the potential of CycleGAN for artistic style transfer. The model's ability to leverage unsupervised learning eliminates the need for extensive paired training data, making it adaptable to various artistic styles.

Furthermore, the successful separation of style and content allows for manipulation without compromising the original image. Quantitative evaluation using the MiFID score supported the qualitative assessment, indicating high similarity between generated and authentic Monet paintings. This research paves the way for further advancements in artistic style transfer tools. Future work could explore incorporating user-specific style preferences and potentially integrating advanced neural architectures for even richer image synthesis and style control. This development would further democratize artistic expression and empower users to become active participants in the digital art landscape.

References

[1] Khan, Asad, Muhammad Ahmad, Nuzhat Naqvi, Faisal Yousafzai, and Jing Xiao. "Photographic painting style transfer using convolutional neural networks." Multimedia Tools and Applications 78 (2019): 19565-19586.

[2] Wang, Zixu, Zhe Quan, Zhi-Jie Wang, Xinjian Hu, and Yangyang Chen. "Text to image synthesis with bidirectional generative adversarial network." In 2020 IEEE International Conference on Multimedia and Expo (ICME), pp. 1-6. IEEE, 2020.

[3] Zhang, Zhiqiang, Jinjia Zhou, Wenxin Yu, and Ning Jiang. "Drawgan: Text to image synthesis with drawing generative adversarial networks." In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4195-4199. IEEE, 2021.

[4] Ruder, Manuel, Alexey Dosovitskiy, and Thomas Brox. "Artistic style transfer for videos and spherical images." International Journal of Computer Vision 126, no. 11 (2018): 1199-1219.

[5] Gatys, Leon A., 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, pp. 2414-2423. 2016.

[6] Gupta, Varun, Rajat Sadana, and Swastikaa Moudgil. "Image style transfer using convolutional neural networks based on transfer learning." International journal of computational systems engineering 5, no. 1 (2019): 53-60.

[7] Ruder, Manuel, Alexey Dosovitskiy, and Thomas Brox. "Artistic style transfer for videos." In Pattern Recognition: 38th German Conference, GCPR 2016, Hannover, Germany, September 12-15, 2016, Proceedings 38, pp. 26-36. Springer International Publishing, 2016.

[8] A. Jang, A. S. Uzsoy, and P. Culliton, "I'm Something of a Painter Myself," Kaggle, 2020. [Online]. Available: https://kaggle.com/competitions/gan-getting-started.

[9] Goodfellow, Ian, 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 (2014).

[10] Creswell, Antonia, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. "Generative adversarial networks: An overview." IEEE signal processing magazine 35, no. 1 (2018): 53-65.

[11] Google AI. (2023, November 22). Overview of GAN Structure. Machine Learning Guide. [Online]. Available: https://developers.google.com/machine-learning/gan/gan_structure

[12] Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks." In Proceedings of the IEEE international conference on computer vision, pp. 2223-2232. 2017.

[13] Jing, Yongcheng, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. "Neural style transfer: A review." IEEE transactions on visualization and computer graphics 26, no. 11 (2019): 3365-3385.

[14] Bai, Ching-Yuan, Hsuan-Tien Lin, Colin Raffel, and Wendy Kan. "A Large-scale Study on Training Sample Memorization in Generative Modeling." (2020).