Monet2Photo: Reverse Style Transfer using CycleGAN with Impressionism-to-Reality Domain

George Kerry Wijaya[1], Shining Sunny Chang[2], Jonathan Saputra[3], Annabelle Chloe[4], Monika Evelin Johan[5,*]

[1,2,3,4,5]*Universitas Multimedia Nusantara, Scientia Boulevard, Tangerang 15810, Indonesia*

(Received: October 18, 2025; Revised: December 10, 2025; Accepted: March 18, 2026; Available online: May 3, 2026)

Abstract

The application of artificial intelligence in artistic image transformation has primarily focused on converting real-world photographs into stylized artworks. In contrast, the inverse task of reconstructing photorealistic images from paintings remains relatively underexplored and presents substantial technical challenges. This study aims to investigate the feasibility and limitations of reverse style transfer by translating impressionist paintings into realistic photographic images, using the works of Claude Monet as a representative case. The main contribution of this research lies in providing a critical examination of reverse image translation under extreme domain gaps, rather than proposing aesthetic enhancement. An unpaired image-to-image translation framework based on CycleGAN is employed to learn mappings between painting and photographic domains without relying on paired data. The methodology is conceptually grounded in adversarial learning combined with cycle consistency constraints to encourage structural preservation while attempting to reconstruct plausible visual features. The experimental setup utilizes a dataset consisting of 300 Monet paintings and 7,028 real photographs, with targeted data augmentation applied to the painting domain to address data imbalance. Prior to model training, exploratory data analysis is conducted to characterize domain discrepancies through visual and statistical comparisons, including color distribution analysis, grayscale intensity patterns, texture descriptors, and dimensionality reduction. Model performance is evaluated through controlled experiments using distribution-based distance measures and qualitative visual inspection. The results indicate that while the model is capable of preserving coarse spatial layouts and generating diverse outputs without memorization, it struggles to recover high-fidelity textures, illumination, and contrast required for photorealistic reconstruction. These findings highlight the inherent limitations of classical CycleGAN architectures for reverse style transfer and suggest the need for more expressive models and stronger constraints in future research on art-to-reality image translation.

Keywords: Adversarial Learning, CycleGAN, Image-To-Image Translation, Monet Painting, Photorealistic Image Generation

1. Introduction

Artificial intelligence based style transfer has become widely used in digital media, particularly for transforming photographs into artistic paintings through mobile applications, creative tools, and visual content platforms [1]. Several previous studies [2], [3], [4], [5] have mainly concentrated on forward style transfer, where real photographs are translated into Monet style images using CycleGAN based architectures. These studies consistently demonstrate that generative adversarial models are effective in capturing artistic color patterns, brushstroke textures, and stylistic appearance while preserving basic structural information from the original photographs.

In contrast, the reverse task of translating paintings back into photorealistic images remains relatively underexplored. This problem is inherently more challenging, as artistic paintings often abstract visual details, soften object boundaries, and omit realistic texture and lighting cues. Consequently, reverse style transfer constitutes an ill posed reconstruction problem rather than a direct image translation task, since the information required for photorealistic recovery is not explicitly preserved in the artistic domain.

This study explores reverse style transfer using the impressionist works of Claude Monet [6], which are known for loose brushstrokes and ambiguous forms [7]. An unpaired image to image translation framework based on CycleGAN is employed to translate Monet paintings into real world photographs without relying on paired datasets. Rather than aiming to achieve full photorealistic reconstruction, this research adopts an exploratory perspective to examine how much structural consistency and visually plausible detail can be recovered under limited data availability and computational constraints. By analyzing both quantitative metrics and qualitative visual outcomes, the results clarify

***Corresponding author: Monika Evelin Johan (monika.evelin@umn.ac.id)**
DOI: https://doi.org/10.47738/jads.v7i2.1222

 ISSN 2723-6471

the practical limitations of classical CycleGAN architectures for reverse style transfer and contribute to a more critical understanding of generative image translation beyond aesthetic stylization.

2. Literature Review

2.1. Claude Monet Painting

Claude Monet was a central figure in the Impressionist movement, an artistic approach that prioritizes the perception of light, color, and atmosphere over precise object representation. As illustrated in figure 1, Monet's paintings predominantly depict natural landscapes, gardens, and outdoor scenes rendered through loose brushstrokes, softened edges, and layered color compositions. Instead of sharply defined forms, objects are often suggested through tonal variation and color contrast, resulting in ambiguous boundaries and reduced structural clarity.

Figure 1. Samples of Claude Monet Painting Style

From a computational perspective, these stylistic characteristics introduce significant challenges for reverse style transfer. The abstraction of object contours, compression of tonal ranges, and deliberate omission of fine textures lead to a substantial loss of information that is essential for photorealistic reconstruction. Unlike forward style transfer, where stylistic patterns can be superimposed onto well-defined photographic structures, reverse translation requires the model to infer missing details such as surface textures, lighting direction, and spatial depth that are not explicitly encoded in the painting domain. Consequently, Monet's artistic style represents an extreme domain gap when mapped back to real-world photographs. This makes it a suitable case for evaluating the limitations of unpaired image-to-image translation models, as successful reconstruction depends not only on stylistic learning but also on the model's capacity to hallucinate plausible visual details from incomplete visual cues. In this context, Monet's paintings serve not as aesthetic targets to be replicated, but as a challenging source domain for assessing the feasibility of reverse style transfer under information loss.

2.1. Deep Learning in Image Translation

Deep learning enables models with multiple hierarchical layers to learn complex visual representations from data, including edges, textures, color distributions, and spatial relationship [8]. In computer vision, these capabilities have driven significant advances in image understanding and generation tasks, forming the foundation for modern image-to-image translation techniques. Image-to-image translation aims to learn a mapping between two visual domains, allowing images from a source domain to be transformed into a target domain without explicit pixel-level correspondence [9]. In particular, generative adversarial network based models have demonstrated strong performance in forward style transfer, where realistic photographs are translated into artistic representations. However, reverse image translation, where abstract or stylized images are mapped back to photorealistic domains, presents a fundamentally different challenge. Artistic paintings often discard fine-grained visual cues that are essential for realistic reconstruction, making the translation problem ill posed and under constrained. As a result, the model must infer missing information rather than merely transferring style patterns, which significantly increases the difficulty of the task. Although more recent image generation frameworks, such as diffusion-based and transformer-based models, have shown superior performance in high-fidelity synthesis, they typically require large-scale datasets and substantial computational resources. In contrast, CycleGAN offers a lightweight and well-established baseline for unpaired image translation under limited data and hardware constraints. In this study, CycleGAN is therefore adopted not as a state-of-the-art solution, but as an analytical tool to examine the feasibility and limitations of reverse style transfer across large domain gaps.

ISSN 2723-6471

3. Methodology

3.1. Generative Adversarial Networks (GAN)

Generative Adversarial Networks (GANs) are a class of deep learning models designed for data generation through adversarial training between two components: a generator and a discriminator [10]. The generator aims to synthesize images that resemble samples from a target distribution, while the discriminator evaluates whether an input image is real or generated. Through this competitive learning process, GANs are able to model complex visual distributions without requiring explicit supervision.

As illustrated in figure 2, GANs have demonstrated strong capability in capturing local texture patterns, color distributions, and spatial structures in image synthesis tasks. However, adversarial training alone does not guarantee semantic correctness or perceptual realism, particularly when the target domain contains visual information that is not explicitly preserved in the source domain. This limitation is especially relevant for reverse style transfer, where the generator must infer missing textures, lighting cues, and structural details rather than directly transferring stylistic features.

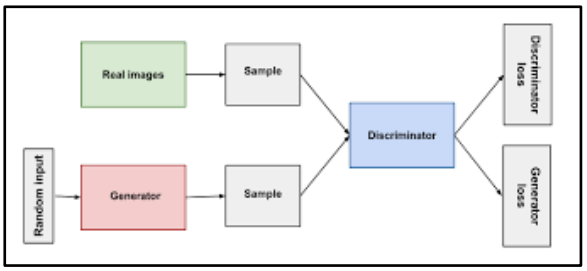


Figure 2. GAN Algorithms

3.2.CycleGAN

In this study, CycleGAN is adopted due to its suitability for scenarios where direct correspondences between source and target images are unavailable [11]. The objective is to translate images between the photographic domain and Monet's painting domain using unpaired data. CycleGAN learns two bidirectional mappings: from photographs to Monet-style paintings and from Monet paintings to photographs. A cycle consistency constraint is imposed to ensure that an image translated from one domain to the other can be reconstructed back to its original form with minimal distortion[3], [12], [13], [14], [15], [16]. This constraint encourages the preservation of structural content while allowing stylistic variation. The overall CycleGAN workflow is illustrated in figure 3.

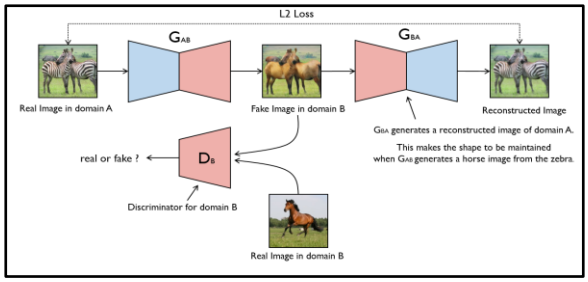


Figure 3. CycleGAN Algorithm

Alternative approaches such as Pix2Pix rely on paired image data, where each photograph must have a corresponding stylized counterpart, which is not available in this study. CycleGAN overcomes this limitation by learning from independent collections of Monet paintings and photographs, making it a suitable baseline for analyzing reverse style transfer under unpaired conditions [17], [18]. CycleGAN consists of two generator–discriminator pairs, as shown in figure 4:

The generators are implemented using ResNet-based architectures with instance normalization and nine residual blocks, while the discriminators adopt a PatchGAN design that evaluates realism at the patch level. Cycle consistency is enforced by mapping images forward and backward across domains during training. G: Monet → Photo, F: Photo → Monet, D_Y: distinguishes real vs generated photos, D_X: distinguishes real vs generated paintings.

 ISSN 2723-6471

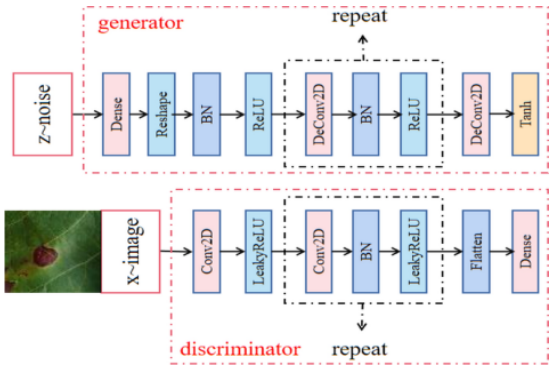


Figure 4. CycleGAN Architecture: Generator vs. Discriminator

3.3. Memorization-informed Fréchet Inception Distance (MiFID)

To evaluate model performance, this study employs Memorization-informed Fréchet Inception Distance (MiFID), an extension of the standard Fréchet Inception Distance (FID) [19], [20], [21]. In this context, a lower MiFID score indicates that the generated images more closely resemble the target photographic domain. FID measures the similarity between feature distributions of generated images and real photographs from the target domain, where lower values indicate closer alignment.

$$MiFID = FID \,.\frac{1}{d_{thr}} \tag{1}$$

The calculation of MiFID is derived from the product of two primary components. The FID value is obtained by calculating the Fréchet distance between the feature representations of the generated photorealistic images and the real photographic dataset, extracted through a pre-trained Inception network. Meanwhile, the memorization distance $\frac{1}{d_{thr}}$is calculated based on the minimum distance between generated samples and the training images in the feature space, where d_{thr} serves as a threshold to penalize models that simply replicate training data.

MiFID augments this metric by incorporating memorization distance, which assesses whether the model reproduces training samples rather than generating novel outputs. This distinction is particularly important in low-data regimes, where generative models are prone to overfitting. By using MiFID, this study evaluates not only the visual similarity to photorealistic textures but also the model's ability to produce diverse and non-memorized outputs. [17].

3.4. Data Collection

The dataset used in this study is sourced from the Kaggle competition *"I'm Something of a Painter Myself"* [22]. It consists of two unpaired image domains: a fine-art domain, with 300 impressionist paintings attributed to Claude Monet; a photographic domain, comprising 7,028 standard photographs depicting natural and urban scenes. All images are provided at a resolution of 256 × 256 pixels and are available in both JPEG and TFRecord formats. JPEG images are used for visual inspection, exploratory analysis, and preprocessing, while TFRecords are employed for efficient data loading during model training in TensorFlow. This separation allows interpretability during analysis while maintaining computational efficiency during training.

3.5. Data Preparation

Data augmentation was applied specifically to Monet paintings due to their significantly smaller dataset size compared to original photos (300 vs. 7,038). Without augmentation, the model risks overfitting on Monet images and becoming biased toward patterns in the photo dataset [23], [24]. To address this, three augmentation techniques Color Jitter, Random Crop, and Horizontal Flip were used to increase the variety of Monet images [25]. Figure 5 show some of the augmentation results.

 ISSN 2723-6471

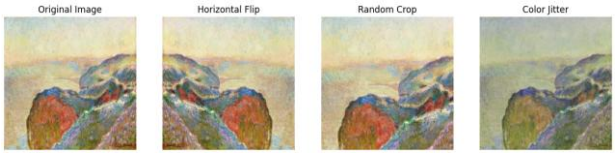


Figure 5. Augmentation Result

Color Jitter introduced variations in color, brightness, contrast, and saturation, helping the model recognize Monet's style under diverse lighting and artistic effects. Random Crop extracted random 224x224 pixel portions of images, introducing compositional diversity and teaching the model to focus on different perspectives. Horizontal Flip mirrored images horizontally, which is particularly useful since many Monet paintings lack fixed orientation, improving the model's generalization ability.

Each Monet image was augmented five times using these techniques, increasing the total number of Monet images from 300 to 4,800. The augmented dataset was then combined with the original photo dataset to train the model more effectively. This ensured a balanced representation of both classes, enabling the model to better capture Monet's Impressionist characteristics without bias toward the photo dataset.

After augmenting the Monet dataset to balance the representation of both classes, the next step involved preprocessing all images both Monet paintings and original photos to ensure consistency and optimize model training. This included resizing and normalizing the images. Resizing standardized all images to 256x256 pixels, enabling efficient processing by the model. Normalization adjusted pixel values to the range [-1, 1], stabilizing the learning process and accelerating training by reducing the impact of extreme color intensity variations between Monet paintings and photos. The results shown in Figure 6.

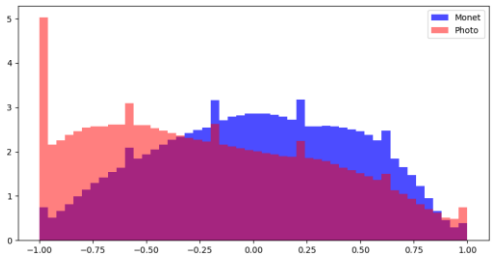


Figure 6. Pixel Distribution After Normalization

The normalization process revealed distinct characteristics in the pixel distributions of the two datasets. Monet's paintings exhibited a smoother and more evenly distributed histogram, reflecting the softer gradients typical of Impressionist art. In contrast, the original photos showed spikes in the negative range, likely due to dominant shadows or darker tones. These differences highlight that even after normalization, the unique color and texture qualities of each dataset remain distinguishable. By addressing these variations through normalization, the model can more effectively learn the visual differences between Monet's style and real photos without being biased by the initial intensity disparities. This preprocessing step complements the data augmentation process by ensuring that the expanded Monet dataset and the original photo dataset are uniformly prepared, allowing the model to focus on capturing stylistic nuances rather than being affected by inconsistent image properties.

3.6. Data Modelling

This study adopts the CycleGAN framework to perform unpaired image-to-image translation from ordinary photographs to Monet-style paintings [26], [27]. As previously discussed, CycleGAN is particularly well-suited for this task due to its ability to operate without paired datasets, allowing the model to learn stylistic mappings using separate collections of Monet paintings and real photos. In this stage, the focus shifts from theoretical foundations to the actual implementation architecture and training strategies applied in this project. The whole formula of CycleGAN can be seen below.

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda \mathcal{L}_{cyc}(G, F) + \lambda_{id} \mathcal{L}_{identify}(G, F) \quad (2)$$

3.6.1. Generator Architecture

The generator follows a U-Net-inspired encoder–decoder structure with skip connections, which help preserve spatial information lost during downsampling. The encoder progressively reduces image resolution while increasing depth using convolutional layers with stride 2, instance normalization, and LeakyReLU activation. The bottleneck consists of nine residual blocks, enabling deeper feature extraction without vanishing gradients. The decoder then reconstructs the image through transposed convolutional layers, instance normalization, and ReLU activation. The final layer applies a tanh activation to output an image with pixel values in the range [-1, 1]. This architecture allows the generator to capture both global structure and fine artistic textures, which is essential in modeling the soft, blended brushstrokes of Monet's style. To support training stability and reduce overfitting, dropout layers are selectively included after the residual blocks. The total parameter count is approximately 54 million, reflecting the model's capacity to learn complex style mappings Generator Architecture across domains.

3.6.2. Instance Normalization

A key technique used throughout the generator and discriminator is Instance Normalization. Unlike batch normalization, instance normalization operates on individual images, helping maintain consistent contrast and color style within each generated output. This is particularly useful for style transfer tasks where global consistency is less important than stylistic fidelity. By normalizing feature maps channel-wise, instance normalization accelerates convergence, improves generalization, and mitigates the effects of varying lighting and color intensity across images.

3.6.3. Discriminator Architecture

The discriminators follow the PatchGAN design, which classifies image patches rather than full images as real or fake. This encourages the generator to produce high-frequency detail and texture fidelity rather than merely approximating global structure. Each discriminator processes the input image through a series of convolutional layers with increasing filter depth and decreasing spatial size, using LeakyReLU activation and instance normalization. The output is a 30×30×1 matrix of patch-level realism scores, promoting fine-grained learning of brush texture, stroke density, and localized contrast. This design allows the discriminator to act as a stylization critic, effectively judging whether small regions of the image resemble Monet's brushwork.

3.6.4. Training Strategy

Training was conducted using the Adam optimizer with a learning rate of 0.0002, $\beta 1 = 0.5$, and a batch size of 1. Due to hardware limitations, training was limited to 6 epochs. Nevertheless, the model displayed convergence behavior by the fifth epoch, especially in the photo-to-Monet direction. To improve generalization despite limited data and training time, several augmentation strategies were used in the Monet domain: Color Jitter, Random Cropping, and Horizontal Flipping. These expanded the Monet dataset from 300 images to over 4,800, helping the generator experience stylistic diversity. To monitor progress during training, loss values from both generator and discriminator were plotted. The generator loss showed a downward trend before stabilizing, while the discriminator loss remained relatively flat, indicating that the generator began to successfully deceive the discriminator as training progressed.

4. Results and Discussion

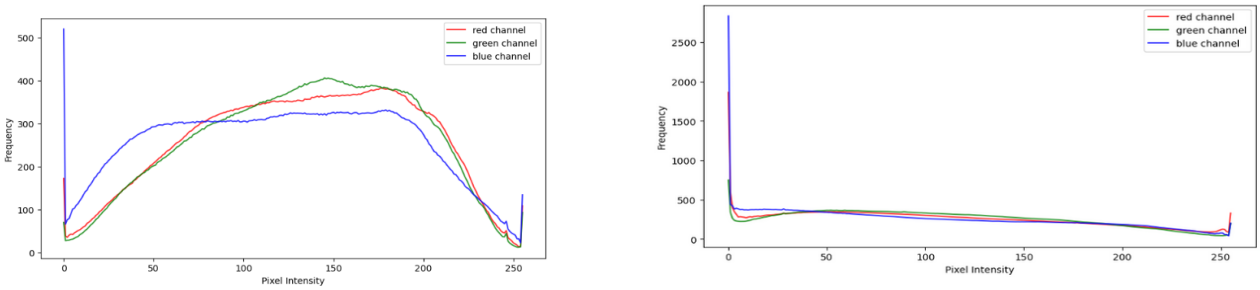
4.1. Exploratory Data Analysis

Exploratory data analysis is conducted to characterize the visual differences between the Monet and photographic domains prior to model training.

4.1.1. Color Histogram Analysis

Pixel-level color histograms are computed in both RGB and HSV color spaces[28], [29]. RGB histograms are generated using the full dataset, while HSV histograms are computed from a stratified random sample of 100 images per domain to manage computational cost. All histograms are normalized to enable cross-domain comparison. As shown in Figure 7, Monet paintings exhibit well-distributed midtone intensities across color channels, reflecting controlled color usage and gradual tonal transitions. In contrast, photographs show a skew toward lower intensity values, indicating higher contrast and frequent shadow regions. HSV analysis further reveals dominant warm hues and moderate saturation in Monet's works, while photographic images display more irregular hue distributions and stronger value variations.

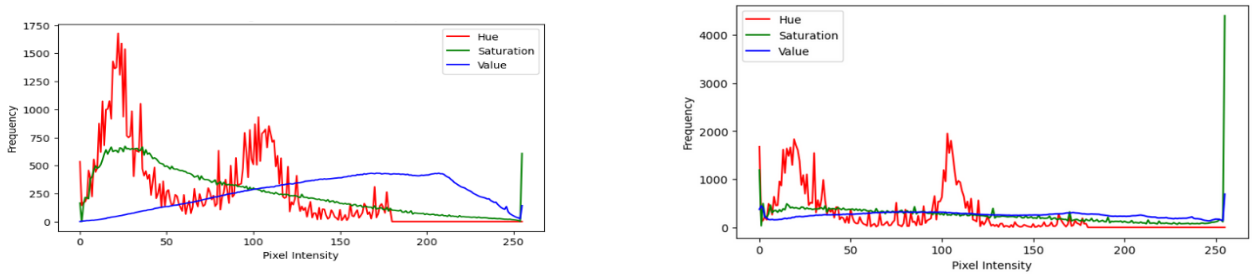
ISSN 2723-6471



(a) (b)

Figure 7. Average Color RGB Distribution : a) Monet; b) Photo

To further analyze chromatic composition, HSV histograms were generated from a random sample of 100 images in each domain from Figure 8. In the Monet dataset, hue values concentrate primarily within the 10–80 pixel intensity range, indicating dominant warm tones such as reds, oranges, and yellows. Saturation follows a unimodal curve peaking in the mid-range, while the value channel increases gradually, suggesting brighter overall compositions. The photo dataset, in contrast, reveals more erratic hue distributions with sharp peaks and broader variation. Saturation levels are more uniform and less pronounced, while value skews sharply toward the upper end of the intensity scale, reflecting areas of strong lighting or overexposure.

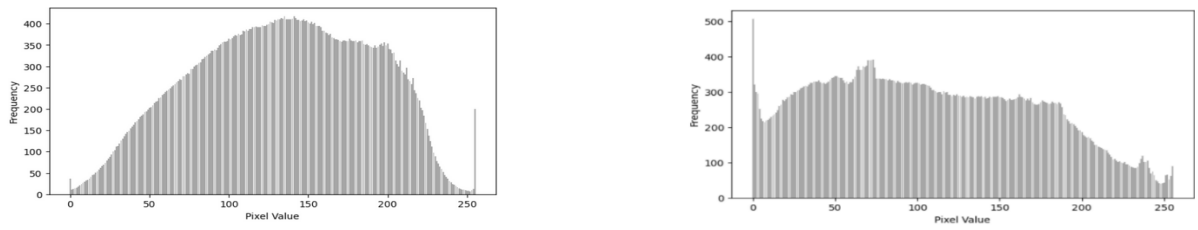


(a) (b)

Figure 8. HSV Histogram: a) Monet; b) Photo

4.1.2. Grayscale Intensity

Next, to analyze tonal structure and brightness variation, all images were transformed into grayscale using a luminance-weighted method. The resulting grayscale intensity values were collected and visualized through normalized histograms to capture the distribution of light and dark regions across the datasets. This representation provides a simplified yet informative view of tonal composition, supporting subsequent processes such as texture extraction and informing potential adjustments like brightness normalization or contrast enhancement during pre-processing. Figure 9 shows the grayscale histogram distributions for both the Monet and photographic datasets, derived from a random sample of 100 images per domain. The Monet paintings exhibit a relatively even distribution centered around mid-intensity values, with a gentle rise and fall across the pixel range. This structure suggests a dominance of midtones and an intentional suppression of extremes, resulting in low contrast and smooth tonal transitions, which characteristics aligned with the aesthetic of impressionist art.



(a) (b)

 ISSN 2723-6471

Figure 9. Grayscale Histogram : a) Monet; b) Photo

In contrast, the photographic domain displays a histogram skewed toward lower intensity values, with a notable peak in the 0–50 pixel range. This pattern indicates the presence of deeper shadows and greater contrast, commonly found in natural scene captures. The histogram also shows a broader spread and higher variation, reflecting the diversity of lighting conditions and exposure levels inherent in real-world photography. These differences in grayscale intensity profiles emphasize the divergence in tonal composition between the two domains. While Monet's work favors visual softness and luminance balance, photographs preserve shadow detail and high dynamic range. As with color histograms, this disparity reinforces the need for preprocessing techniques—such as histogram equalization or contrast normalization, to address tonal imbalance before model training in style transfer pipelines.

4.1.3. Texture Analysis using GLCM

Texture quantification was performed using Gray-Level Co-occurrence Matrix (GLCM) metrics, focusing on contrast and correlation to assess spatial intensity relationships within grayscale images [30]. As presented in Table 1 and visualized in Figure 10, the Monet dataset demonstrates a higher average contrast value (283.20) compared to the photographic domain (220.06), indicating more pronounced local intensity variation. This elevated contrast reflects the layered brushstroke patterns and surface irregularities characteristic of impressionist painting techniques.

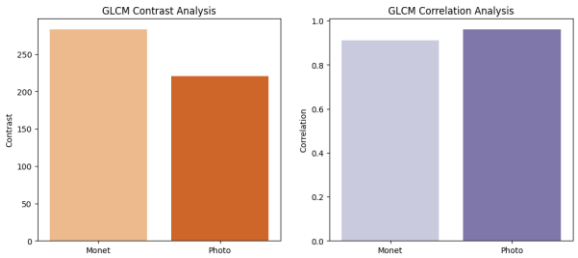


Figure 10. Texture Analysis: GLCM Contrast & Correlation

Table 1. Texture Analysis using GLCM

Domain	Texture Analysis	
	Contrast	Correlation
Photo	220.06	0.96
Monet	283.20	0.96

Correlation values remain consistent across both domains (0.96), suggesting a similar level of spatial dependency between neighboring pixels. While this metric indicates comparable structural regularity in statistical terms, it does not necessarily reflect perceptual texture differences, especially those driven by artistic abstraction.

The contrast differences observed in Figure 10 prove the existence of richer surface textures in the Monet domain, highlighting the effectiveness of GLCM features in capturing style differences. These findings provide quantitative support for incorporating texture descriptors into style-aware preprocessing or discriminator design in generative modeling.

4.1.4. Color Palette

The dominant color palette was extracted using the K-Means clustering method [27] on pixel values taken from 100 images in each domain. The resulting cluster centers represent the most frequently occurring color groups in the dataset, providing a compact summary of style preferences. Figure 11 shows the five-cluster palette generated from the Monet and photography datasets. The Monet palette is characterized by soft tones such as desaturated blue, gray, olive green, and warm beige. These colors reflect the Impressionist aesthetic, which emphasizes softness, tonal harmony, and atmospheric mood over visual accuracy. The presence of these soft colors also aligns with previous histogram findings, which showed a dominance of mid-tones and limited saturation in Monet's works. In contrast, the photographic palette includes dark grays, sharp blacks, and a wider range of chromatic variations. Although there is overlap in the color families, the photographic palette appears less uniform, with higher visual contrast and more saturated tones. This

 ISSN 2723-6471

distribution reflects the diversity of lighting conditions, object materials, and exposure levels in capturing the real world.

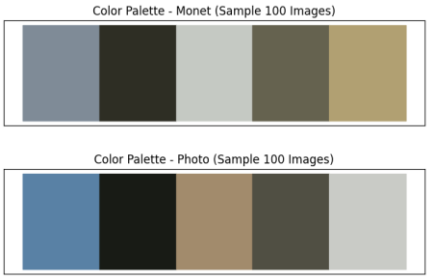


Figure 11. Color Pallette Monet and Photo (100 Samples)

The clustering results reinforce the chromatic differences between domains, supporting the hypothesis that color use in artistic images tends toward aesthetic abstraction, while photographs maintain a more literal and high-contrast representation. These palette differences may provide insights into style conditioning, perceptual loss adaptation, or palette-based color augmentation strategies in model development.

4.1.5. Principle Component Analysis

Principal Component Analysis (PCA) was applied to grayscale images in both domains to generate low-dimensional reconstructions that preserve overall structure while discarding fine visual detail [31]. Figure 12 illustrate example outputs from each domain, reconstructed using a reduced number of principal components.

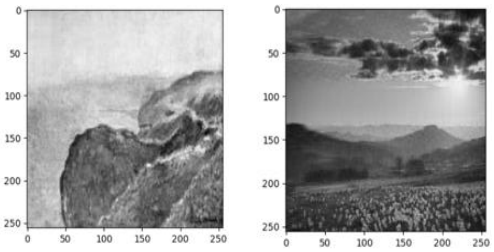


Figure 12. Reconstruction PCA

The Monet reconstruction appears smooth and visually softened, with lower edge clarity and a more homogeneous tonal distribution. This outcome reflects the impressionist style, where brushstrokes and compositional flow dominate over precise boundaries and detail. In contrast, the photographic reconstruction retains sharper edges and finer textures, especially in areas of high contrast, such as sky contours and object boundaries.

These observations suggest that photographs maintain localized visual information more effectively under dimensional compression, while paintings emphasize broader tonal regions and stylistic abstraction. PCA thus proves useful for revealing how visual complexity and structural encoding differ between artistic and natural imagery, and can serve as a diagnostic tool for assessing dataset characteristics in generative modeling tasks [19].

4.2. Performance Results

The performance of the CycleGAN model was assessed through three primary approaches: quantitative, qualitative, and statistical. Quantitatively, the reduction in loss values for the generators and discriminators served as key indicators of model improvement. Specifically, the generator loss for the Monet to photo transformation exhibited a significant decrease followed by stabilization. This trend indicates the model's enhanced capability to reconstruct images that align with the target photographic distribution. Regarding the discriminator for the photographic domain, a stable loss value suggests that the generator effectively produced outputs that challenge the discriminator's ability to distinguish real from synthetic photos. Figure 13 below shows the result of generator loss and discriminator loss.

ISSN 2723-6471

(a) (b)

Figure 13. Loss Graphs : a) Generator Loss; b) Discriminator Loss

The generator loss graph shows a consistent decline in both directions during the early stages of training, with the Monet generator (photo → Monet) achieving slightly lower and more stable loss values by epoch 6. This indicates that the generator was increasingly successful in producing Monet-style outputs that could fool the discriminator. On the other hand, the discriminator for the Monet domain displays relatively flat and higher loss values, suggesting that it struggled to distinguish between real Monet paintings and generated ones, further confirming the generator's effectiveness in mimicking Monet's style.

In contrast, the photo discriminator (Monet → photo) exhibits a more pronounced and steady decrease in loss, implying that it could still reliably differentiate generated photo-like images from real ones. This imbalance highlights that the model more effectively learned to generate Monet-style paintings than to reconstruct realistic photographs from Monet inputs. From a technical perspective, this disparity confirms the inherent difficulty of the reverse style transfer task, as the model faces greater challenges in synthesizing the high-frequency details and sharp textures required for photorealistic reconstruction compared to the stylistic abstraction of impressionist art. The Monet to Photo transformation yielded a FID score of 283.96 and a Memorization Score of 15.12, indicating that while the model moderately diverged from the distribution of real photographic images, it still produced a fair degree of variation and creativity as shown in figure 14.

```
FID Score: 283.96
Memorization Score (Avg NN Dist): 15.1153
```

Figure 14. Monet-to-Photo FID Evaluation Results

The calculation of these metrics involves comparing the feature distributions extracted via a pre-trained Inception network from both the generated outputs and the real photographic dataset. The relatively high FID score suggests that the model's ability to reconstruct realistic photo textures from Monet inputs remains limited. This is likely due to the abstract and stylized nature of Monet paintings, which lack detailed structure and sharp edges.

However, the final MiFID score reflects the integration of a memorization penalty into the evaluation. While the FID measures visual divergence, the memorization distance component $\frac{1}{d_{thr}}$ assesses whether the generator is producing novel outputs or merely replicating training samples. Since the final MiFID score of 15.12 is significantly lower than the raw FID value, it indicates that the model achieved a high degree of generalization. This shows that the outputs are not simply overfitted replicas of the training set, but rather novel interpretations that aim to approximate realism. These results reinforce the known difficulty of reverse style transfer from art to photo and highlight the importance of combining FID with memorization based metrics to evaluate both accuracy and generalization in generative models.

The visual results from the Monet to Photo transformation reveal several challenges in reconstructing photorealistic images from artistic inputs. As shown in figure 15, many of the generated photos appear significantly darker than their Monet counterparts, with muted brightness and limited contrast range. In several examples, key textures such as water reflections, foliage, or architectural edges are either underdeveloped or completely absent, resulting in visually incomplete scenes. This outcome aligns with the findings from the RGB analysis during the data understanding phase, where the real photo domain showed strong shadows and dark pixel concentrations, especially in the blue channels. The model appears to have struggled in mapping Monet's soft brushstrokes and high midtone balance into the sharp, high contrast realism typical of photographs. These results suggest that while the model attempts to preserve structure, it fails to recover the clarity and texture density expected from real world photo outputs. This reinforces the inherent

 ISSN 2723-6471

difficulty of reverse style transfer from abstract to realistic domains, as the information required for photorealistic recovery is not explicitly present in the source artistic domain.

Figure 15. Photo Generated Images

5. Conclusion

This project successfully explored the challenging task of reverse style transfer by transforming Claude Monet's impressionist paintings into photorealistic images using the CycleGAN framework. The model demonstrated a modest capability in reconstructing spatial structure and recovering partial texture details from abstract artworks. Despite limitations in photorealistic sharpness, which primarily stem from the inherent ambiguity in impressionist inputs, the generated outputs retained recognizable scene layouts and natural color gradients. Quantitative evaluation using FID and MiFID metrics confirmed that while the realism of the generated images remains limited with a FID score of 283.96, the model successfully avoided memorization and produced creative yet structurally coherent outputs as indicated by the MiFID score of 15.12. The application of targeted data augmentation techniques to the Monet dataset successfully mitigated dataset imbalance, increasing the sample size from 300 to 4,800 images and enabling better learning from limited artistic samples.

These findings highlight the inherent difficulty of generating reality from abstraction and suggest that future work could focus on deeper models, multi scale attention, or hybrid architectures to better recover high frequency details. Overall, this study advances the understanding of reverse style transfer and opens possibilities in historical reconstruction, visual restoration, and art based data synthesis.

6. Declarations

6.1. Author Contributions

Conceptualization: G.K.W., S.S.C., J.S., A.C., and M.E.J.; Methodology: G.K.W., S.S.C., J.S., A.C., and M.E.J.; Software: G.K.W., S.S.C., J.S., A.C., and M.E.J.; Validation: G.K.W., S.S.C., J.S., A.C., and M.E.J.; Formal Analysis: G.K.W., S.S.C., J.S., A.C., and M.E.J.; Investigation: G.K.W., S.S.C., J.S., A.C., and M.E.J.; Resources: G.K.W., S.S.C., J.S., A.C., and M.E.J.; Data Curation: G.K.W., S.S.C., J.S., A.C., and M.E.J.; Writing Original Draft Preparation: G.K.W., S.S.C., J.S., A.C., and M.E.J.; Writing Review and Editing: G.K.W., S.S.C., J.S., A.C., and M.E.J.; Visualization: G.K.W., S.S.C., J.S., A.C., and M.E.J.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data used in this study is public data available in the Kaggle repository [22].

6.3. Funding

This research does not use funding from any party, other than facilities support by Universitas Multimedia Nusantara.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

This research has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] L. Wang, L. Wang, and S. Chen, "ESA-CycleGAN: edge feature and self-attention based cycle-consistent generative adversarial network for style transfer," *IET Image Process*., vol. 16, no. 1, pp. 176–190, 2022, doi: 10.1049/ipr2.12342.

[2] V. D, S. Manandhar, S. Naikodi, M. G, M. V. V. L, and M. Z. A. N, "Generating Monet style images using DCGAN based CycleGAN," *SSRN Electron. J.*, vol. 2023, no. Jan., pp. 1–12, 2023, doi: 10.2139/ssrn.4527418.

[3] M. Li, L. Lin, G. Luo, and H. Huang, "Monet style oil painting generation based on cyclic generation confrontation network," *J. Appl. Opt. Comput. Eng*., vol. 2024, no. Feb., pp. 1–15, 2024, doi: 10.1117/12.3025245.

[4] A. Agarwal, I. Sangal, and A. Shrotriya, "The artistic transformation: Monet style transfer using cycle GAN," *J. ICT Sustain. Dev*., vol. 2026, no. Jan., pp. 215–224, 2026, doi: 10.1007/978-3-032-06677-0_22.

[5] F. Neha, D. Bhati, D. K. Shukla, and M. Amiruzzaman, "A tiered GAN approach for Monet-style image generation," *Int. J. Comput. Theory Appl.*, vol. 2024, no. Dec., pp. 232–237, 2024, doi: 10.1109/ICCTA64612.2024.10974888.

[6] G. Yao and Z. Abindinhazir, "Study and appreciation of Claude Monet's artistic creation and life experience," *Idealogy J.*, vol. 8, no. 2, pp. 1–12, 2023, doi: 10.24191/idealogy.v8i2.445.

[7] T. Baumeister, "Space, time, and change in Claude Monet's late paintings," *J. Mod. Contemp. Art Stud*., vol. 2024, no. Jan., pp. 215–233, 2024, doi: 10.1007/978-3-031-39598-7_12.

[8] N. H. Shabrina, "Explainable ensemble deep learning for potato leaf pest and diseases identification," *Int. J. Adv. Soft* Comput. *Appl.,* vol. 17, no. 2, pp. 1–12, 2025, doi: 10.15849/IJASCA.250730.09.

[9] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *IEEE Access*, vol. 2017, no. Oct., pp. 2242–2251, 2017, doi: 10.1109/ICCV.2017.244.

[10] I. Goodfellow, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020, doi: 10.1145/3422622.

[11] X. Ma, "A comparison of art style transfer in Cycle-GAN based on different generators," *J. Phys. Conf. Ser*., vol. 2711, no. 1, pp. 1–12, 2024, doi: 10.1088/1742-6596/2711/1/012006.

[12] M. Fu, Y. Liu, R. Ma, B. Zhang, L. Wu, and L. Zhu, "A model integrating attention mechanism and generative adversarial network for image style transfer," *PeerJ Comput. Sci*., vol. 10, no. Sep., pp. 1–12, 2024, doi: 10.7717/peerj-cs.2332.

[13] L. Lakshmi, "Performance analysis of Cycle GAN in photo to portrait transfiguration using deep learning optimizers," *IEEE Access,* vol. 11, no. Jan., pp. 136541–136551, 2023, doi: 10.1109/ACCESS.2023.3337430.

[14] Y. Liao and Y. Huang, "Deep learning-based application of image style transfer," *Math. Probl. Eng*., vol. 2022, no. 1, pp. 1–10, 2022, doi: 10.1155/2022/1693892.

[15] Y. Liu, "Improved generative adversarial network and its application in image oil painting style transfer," Image Vis. Comput., vol. 105, no. Jan., pp. 1–12, 2021, doi: 10.1016/j.imavis.2020.104087.

[16] L. Vela, F. Fuentes-Hurtado, and A. Colomer, "Improving the quality of image generation in art with top-k training and cyclic generative methods," *Sci. Rep*., vol. 13, no. 1, pp. 17764–17774, 2023, doi: 10.1038/s41598-023-44289-y.

[17] Y. Zhao, "Application of CycleGAN-based image style transfer algorithm in visual communication design," *J. Comput. Methods Sci. Eng*., vol. 25, no. 4, pp. 3152–3164, 2025, doi: 10.1177/14727978251318804.

[18] M. Wang, "Mask CycleGAN: unpaired multi-modal domain translation with interpretable latent variable," *arXiv,* vol. 2022, no. May, pp. 1–12, 2022, doi: 10.48550/arXiv.2205.06969.

[19] C.-Y. Bai, H.-T. Lin, C. Raffel, and W. C. Kan, "On training sample memorization: lessons from benchmarking generative modeling with a large-scale competition," *ACM Trans. Knowl. Discov. Data,* vol. 2021, no. Aug., pp. 2534–2542, 2021, doi: 10.1145/3447548.3467198.

[20] C.-Y. Bai, H.-T. Lin, C. Raffel, and W. Kan, "A large-scale study on training sample memorization in generative modeling," *arXiv*, vol. 2020, no. Jan., pp. 1–12, 2020, doi: 10.48550/arXiv.2001.06046.

[21] J. Kim and J. Song, "A review on the generative models and its performance metrics," *Commun. Stat. Appl. Methods*, vol. 32, no. 2, pp. 235–248, 2025, doi: 10.29220/CSAM.2025.32.2.235.

[22] 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

[23] S.-A. Rebuffi, S. Gowal, D. A. Calian, F. Stimberg, O. Wiles, and T. A. Mann, "Data augmentation can improve robustness," *Adv. Neural Inf. Process. Syst.*, vol. 34, no. Dec., pp. 29935–29948, 2021, doi: 10.48550/arXiv.2111.05328.

[24] K. Maharana, S. Mondal, and B. Nemade, "A review: data pre-processing and data augmentation techniques," *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 91–99, 2022, doi: 10.1016/j.gltp.2022.04.020.

[25] V. C. Tjokro and S. A. Sanjaya, "Enhancing data traceability: a knowledge graph approach with retrieval-augmented generation," *Int. J. Inf. Technol. Intell. Syst*., vol. 2024, no. Jan., pp. 473–478, 2024, doi: 10.1109/ISRITI64779.2024.10963652.

[26] S. Gupta, "Far3Det: towards far-field 3D detection," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 2023, no. Jan., pp. 692–701, 2023, doi: 10.1109/WACV56688.2023.00076.

[27] J. Ao, Z. Ye, W. Li, and S. Ji, "Impressions of Guangzhou city in Qing dynasty export paintings in the context of trade economy: a color analysis of paintings based on k-means clustering algorithm," *Herit. Sci*., vol. 12, no. 1, pp. 77–87, 2024, doi: 10.1186/s40494-024-01195-4.

[28] N. Phuangsaijai, J. Jakmunee, and S. Kittiwachana, "Investigation into the predictive performance of colorimetric sensor strips using RGB, CMYK, HSV, and CIELAB coupled with various data preprocessing methods: a case study on an analysis of water quality parameters," *J. Anal. Sci. Technol.*, vol. 12, no. 1, pp. 19–29, 2021, doi: 10.1186/s40543-021-00271-9.

[29] I. Maity and S. Samanta, "Utilization of image processing tools for a comparative study on RGB and HSV color space: targeting feature extraction of banana fruit image during different ripeness stages," *IEEE Silchar Subsect. Conf. J*., vol. 2024, no. Nov., pp. 1–6, 2024, doi: 10.1109/SILCON63976.2024.10910858.

[30] N. Iqbal, R. Mumtaz, U. Shafi, and S. M. H. Zaidi, "Gray level co-occurrence matrix (GLCM) texture based crop classification using low altitude remote sensing platforms," *PeerJ Comput. Sci*., vol. 7, no. May, pp. 1–12, 2021, doi: 10.7717/peerj-cs.536.

[31] B. Zhao, X. Dong, Y. Guo, X. Jia, and Y. Huang, "PCA dimensionality reduction method for image classification," *Neural Process. Lett.*, vol. 54, no. 1, pp. 347–368, 2022, doi: 10.1007/s11063-021-10632-5.