



ARTICLE

# Optimizing 2D Image Quality in CartoonGAN: A Novel Approach Using Enhanced Pixel Integration

Stellar Choi<sup>1</sup>, HeeAe Ko<sup>2</sup>, KyungRok Bae<sup>3</sup>, HyunSook Lee<sup>2</sup>, HaeJong Joo<sup>4</sup> and Woong Choi<sup>5,\*</sup>

<sup>1</sup>Department of Software Application Virtual Reality, Kangnam University, Yongin, 16979, Republic of Korea

<sup>2</sup>AI-SW Convergence Research Institute, Kangnam University, Yongin, 16979, Republic of Korea

<sup>3</sup>MuhanIT Co., Ltd., Seoul, 07299, Republic of Korea

<sup>4</sup>Department of Artificial Intelligence, Kangnam University, Yongin, 16979, Republic of Korea

<sup>5</sup>Division of ICT Convergence Engineering, Kangnam University, Yongin, 16979, Republic of Korea

\*Corresponding Author: Woong Choi. Email: wchoi@kangnam.ac.kr

Received: 20 November 2024; Accepted: 19 February 2025; Published: 26 March 2025

**ABSTRACT:** Previous research utilizing Cartoon Generative Adversarial Network (CartoonGAN) has encountered limitations in managing intricate outlines and accurately representing lighting effects, particularly in complex scenes requiring detailed shading and contrast. This paper presents a novel Enhanced Pixel Integration (EPI) technique designed to improve the visual quality of images generated by CartoonGAN. Rather than modifying the core model, the EPI approach employs post-processing adjustments that enhance images without significant computational overhead. In this method, images produced by CartoonGAN are converted from Red-Green-Blue (RGB) to Hue-Saturation-Value (HSV) format, allowing for precise adjustments in hue, saturation, and brightness, thereby improving color fidelity. Specific correction values are applied to fine-tune colors, ensuring they closely match the original input while maintaining the characteristic, stylized effect of CartoonGAN. The corrected images are blended with the originals to retain aesthetic appeal and visual distinctiveness, resulting in improved color accuracy and overall coherence. Experimental results demonstrate that EPI significantly increases similarity to original input images compared to the standard CartoonGAN model, achieving a 40.14% enhancement in visual similarity in Learned Perceptual Image Patch Similarity (LPIPS), a 30.21% improvement in structural consistency in Structural Similarity Index Measure (SSIM), and an 11.81% reduction in pixel-level error in Mean Squared Error (MSE). By addressing limitations present in the traditional CartoonGAN pipeline, EPI offers practical enhancements for creative applications, particularly within media and design fields where visual fidelity and artistic style preservation are critical. These improvements align with the goals of Fog and Edge Computing, which also seek to enhance processing efficiency and application performance in sensitive industries such as healthcare, logistics, and education. This research not only resolves key deficiencies in existing CartoonGAN models but also expands its potential applications in image-based content creation, bridging gaps between technical constraints and creative demands. Future studies may explore the adaptability of EPI across various datasets and artistic styles, potentially broadening its impact on visual transformation tasks.

**KEYWORDS:** CartoonGAN; enhanced pixel integration (EPI); image quality enhancement; post-processing; HSV color adjustment; visual fidelity; fog computing; edge computing; creative applications; technical advancements

## 1 Introduction

Computer vision research is continuously advancing at the forefront of technological development, with innovative techniques such as deep neural networks and generative adversarial networks (GANs)



enabling various visual effects, including high-resolution image generation, image style transfer, and the implementation of virtual reality (VR) and augmented reality (AR) experiences [1,2]. This research goes beyond simple image processing, contributing to the creation of new technologies and services by leveraging images and visual information across diverse fields [3]. In particular, AI generation technologies demonstrate innovative potential in visual content creation, such as webtoon background production, allowing creators to easily produce more diverse and creative works while providing users with richer and more engaging content [4]. One such technology, CartoonGAN, combines the words “cartoon” and “generation,” providing a capability to automatically transform images into cartoon styles using computer vision and artificial intelligence [5]. This technology is designed for artistic expression, enabling users to create unique and creative visual content by converting real images into cartoon styles [6]. However, CartoonGAN still faces significant challenges in representing complex outlines and lighting, which can negatively impact the quality of the final images [7]. For instance, the clear and distinct outlines and natural Dlighting effects required for cartoon styles are often not adequately rendered, leading to results that may fall short of user expectations [8].

To address these issues, this paper introduces a novel technique called EPI aimed at significantly improving the output quality of existing CartoonGAN images. EPI applies post-processing adjustments to the original images after the CartoonGAN operation, allowing for a more natural and realistic transformation of lighting while preserving the cartoon style. Currently, CartoonGAN models such as Hayao, Hosoda, Paprika, and Shinkai generate diverse images through distinct visual styles and themes, yet limitations remain in maintaining color consistency and realistic lighting effects due to restrictions within the original model structures [9,10]. To overcome these limitations, this study proposes EPI as an innovative solution to enhance visual consistency and detail in the generated images. Through EPI, we aim to deliver cartoon images that are not only visually compelling but also maintain high-quality visual coherence by applying post-processing techniques instead of directly modifying the original model structures [11].

The core objective of this research is to apply EPI as a post-processing method on CartoonGAN outputs to produce cartoon images that are natural and of high quality. By leveraging post-processing, this approach preserves the stylistic strengths of the existing models while adjusting finer details to elevate the visual completion of the images. This provides creators with high-quality visual content, thereby promoting and enriching image-based creative activities [12]. Moreover, since EPI operates without altering the underlying model structures, it ensures the visual consistency and quality of generated images while retaining the original model’s intended aesthetic.

Furthermore, this study considers the potential for EPI’s application across various datasets and styles, aiming to propose a methodological flexibility that could prove beneficial across multiple creative and applied domains in the future. By enabling adaptability to diverse styles, EPI not only opens new opportunities for enhancing cartoon images but also holds the promise of expanding into various content creation, educational, and applied fields, presenting valuable possibilities for broad utilization [13]. Through actual experiments, this paper demonstrates that the proposed method achieves improvements of approximately 40.14% in LPIPS, 30.21% in SSIM, and a reduction of 11.81% in MSE, thereby validating its effectiveness.

This paper is structured as follows. [Section 2](#) presents the background research on CartoonGAN to date. [Section 3](#) discusses the proposed method. [Section 4](#) covers the selection of image similarity evaluation tools and related background research. [Section 5](#) details the experiment and results. Finally, [Section 6](#) offers discussion and [Section 7](#) suggests conclusion and directions for future research.

## 2 Current Directions and Limitations of CartoonGAN

### 2.1 Challenges of Existing CartoonGAN Models

CartoonGAN, a GAN model specifically designed for anime-style image transformation, offers a unique approach that differentiates it from general image-to-image transformation models like CycleGAN. While CycleGAN employs cycle-consistency loss to enable bidirectional transformations between two domains, CartoonGAN focuses on learning the distinctive features of anime styles, achieving higher style consistency. However, previous studies have pointed out ongoing challenges with CartoonGAN, particularly in terms of color distortion and inaccurate lighting representation, which affect the quality and realism of the transformed images. These challenges often arise from the model's loss functions and style transformation techniques, which may fail to adequately preserve the visual consistency of the original image, especially when dealing with images containing complex outlines or intricate details [14–16].

CartoonGAN tends to overemphasize certain colors and lighting elements during the style transfer process, sometimes compromising the visual coherence of the original image. For example, in the case of the Hayao style, intense primary colors often cause white areas to shift toward a green hue, distorting the original color balance. This issue primarily stems from the loss functions, which drive CartoonGAN to overlearn specific style characteristics [17,18].

### 2.2 Comparison with Other Image Transformation Models

CycleGAN and Pix2Pix are similar image-to-image transformation models, each with distinct strengths and applications. CycleGAN leverages cycle-consistency loss, allowing transformations between domains even in the absence of paired training data. This flexibility is advantageous for applications involving transitions between artistic styles and photographic domains. However, while CycleGAN excels in maintaining overall structural consistency, it lacks precision in color adjustment and lighting accuracy for specific style details. This limitation is particularly evident in transformations that require the nuanced details of anime-style rendering [16]. Pix2Pix, based on supervised learning, performs well in scenarios where paired data is available, supporting clear and specific transformations. The model employs alpha masks for color separation and boundary processing, reducing color bleeding and enhancing texture consistency when transitioning from sketches to colored images. However, Pix2Pix's reliance on paired datasets limits its generalizability across diverse anime styles, and its emphasis on strong stylistic elements may lead to color and lighting inaccuracies. While Pix2Pix is suitable for transformations with distinct boundaries, it encounters challenges in preserving natural lighting and original colors.

In summary, both CycleGAN and Pix2Pix are optimized for specific image transformation tasks, yet they face limitations in meeting the detailed requirements of anime-style transformations, such as precise color correction and lighting consistency. These limitations suggest the need for additional approaches to balance style transfer with structural preservation, making post-processing techniques like EPI in CartoonGAN a promising solution.

### 2.3 Research on Addressing Color and Lighting Distortions

To address issues with color distortion and lighting in style transfer, various loss functions and network structures have been developed. Johnson et al. (2016) introduced Perceptual Loss to retain high-level visual features during transformations, which enhances style and structural consistency by minimizing differences in feature maps extracted from pre-trained neural networks, rather than relying solely on pixel differences [16]. Zhu et al. (2017) introduced Cycle Consistency Loss, which ensures that a bidirectional image transformation maintains coherence. This loss function is based on the principle that the transformed

image should resemble the original when reversed, preserving structural and textural fidelity, and has been instrumental in reducing distortion in models like CartoonGAN [16,19]. Additional models like Pix2PixHD have incorporated multi-scale loss functions to enhance color and lighting representation in high-resolution transformations, focusing on refining intricate details and minimizing texture distortion. These advancements provide foundational methods for improving style transfer processes and serve as a basis for EPI's post-processing approach to overcoming limitations in existing models.

In summary, both CycleGAN and Pix2Pix are optimized for specific image transformation tasks, yet they face limitations.

## 2.4 Contribution of this Study

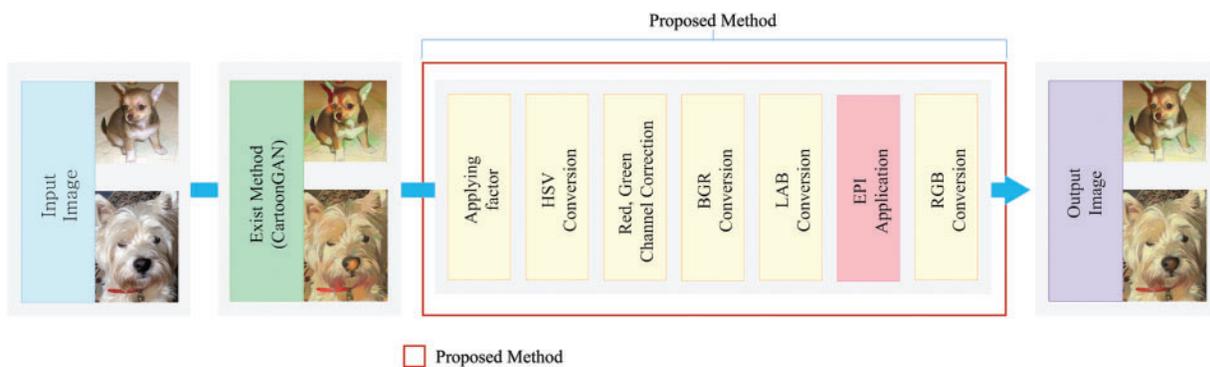
This study introduces EPI as a post-processing technique to address the color and lighting inaccuracies found in CartoonGAN outputs. EPI offers a novel approach to improving image quality without modifying the training structure of the original model, enhancing both the visual consistency and detailed quality of the transformed images. By doing so, it sets a new standard in anime-style transformations.

EPI directly addresses the limitations identified in prior studies, particularly by enhancing color fidelity and lighting accuracy. The technique supports future applications in diverse anime styles and aims to expand into new domains where precise color and lighting are crucial [15,16,18].

## 3 Method

### 3.1 Enhanced Pixel Algorithm (EPI) for Improved Color Consistency in CartoonGAN Outputs

In response to persistent color inconsistency and distortion observed in CartoonGAN outputs, this study presents the EPI as an effective post-processing solution. Rather than altering the existing CartoonGAN architecture, EPI operates as an additional layer of refinement, specifically designed to enhance color consistency by accurately adjusting color tones and preserving the original image's visual details. When recreating a particular anime style, CartoonGAN often encounters issues with excessive emphasis or distortion of certain colors; EPI was designed as a post-processing technique to correct these color discrepancies. EPI calculates the color difference between the CartoonGAN-generated image and the original image and applies a correction factor if this difference exceeds a specified threshold, thereby helping to maintain color consistency [19]. For an overview of the model structure used in this process, refer to the model structure illustrated in Fig. 1.



**Figure 1:** The architecture of the Enhanced Pixel Integration (EPI)

The EPI method offers a post-processing solution designed to address the common challenges of color inconsistency and distortion in CartoonGAN outputs, without requiring changes to the CartoonGAN model itself. Acting as an independent refinement layer, EPI selectively enhances color fidelity while preserving the distinctive style of CartoonGAN-generated images. To achieve this, EPI first converts RGB images into HSV color space, enabling precise adjustments to the hue, saturation, and value channels individually. By focusing on HSV adjustments, EPI aligns the original image's natural color tones with the stylized CartoonGAN colors, which often exhibit distortion, especially in the red and green channels. Distinct correction values are applied to these channels to improve color alignment with the original.

After these targeted adjustments, EPI introduces a blending process, merging the corrected and original images in a specified ratio. This method maintains CartoonGAN's vivid stylistic elements while reducing color discrepancies, resulting in greater visual harmony with the original image. The final image is then converted back to RGB color space for compatibility across various display systems. A primary advantage of EPI is its ability to operate solely as a post-processing step, eliminating the need for modifications to the CartoonGAN model or additional training. This efficiency keeps the original model intact and minimizes system complexity, making EPI especially suited for real-time processing or applications with limited computational resources. By offering a streamlined, non-invasive approach to enhance color consistency, EPI significantly broadens the practical applications of CartoonGAN outputs. Its distinct capabilities add substantial value to creative workflows across digital art, animation, and video editing, meeting high visual quality demands without compromising stylistic integrity.

### 3.2 Color Space Conversion to HSV

The first step in the EPI method is to convert images from the RGB color space to the HSV (Hue, Saturation, Value) color space. This conversion is essential for effectively addressing the color distortion issues that occur in images generated by CartoonGAN. In the RGB color space, the red, green, and blue channels are interrelated, which means that adjusting specific color attributes can inadvertently affect other color components [20,21]. For instance, reducing the red channel in RGB can also alter the brightness and saturation of the affected color, making it challenging to precisely manage exaggerated colors or unwanted color shifts in CartoonGAN outputs [22]. For this reason, it is difficult to make isolated color adjustments in the RGB color space. Fig. 2 demonstrates the effectiveness of HSV over RGB color space in R channel reduction, showing that HSV adjustments preserve overall visual consistency while RGB adjustments cause unintended changes in brightness and saturation.

The conversion process from RGB to HSV involves the following formulas [21]:

$H$  (*Hue*): Represents the type of color, expressed as an angle between  $0^\circ$  and  $360^\circ$ . For example, red ( $0^\circ$ ), green ( $120^\circ$ ), and blue ( $240^\circ$ ).

$S$  (*Saturation*): Indicates the purity of the color, with values ranging from 0 to 1 (or 0% to 100%). Higher values correspond to more vivid colors, while lower values indicate more grayish tones.

$V$  (*Value*): Describes the brightness of the color, ranging from 0 to 1 (or 0% to 100%). A value closer to 0 represents darker shades, while a value closer to 1 indicates brighter shades.

$$H = \left\{ \begin{array}{l} 60^\circ \times \frac{G - B}{\max - \min} + 0^\circ \text{ if } = R \\ 60^\circ \times \frac{B - R}{\max - \min} + 120^\circ \text{ if } = G \\ 60^\circ \times \frac{R - G}{\max - \min} + 240^\circ \text{ if } = B \end{array} \right. \quad (1)$$

$$S = \left\{ \begin{array}{l} 0 \text{ if } = 0 \\ \frac{\max - \min}{\max} \text{ if } \neq 0 \end{array} \right.$$

$$V = (R, G, B)$$



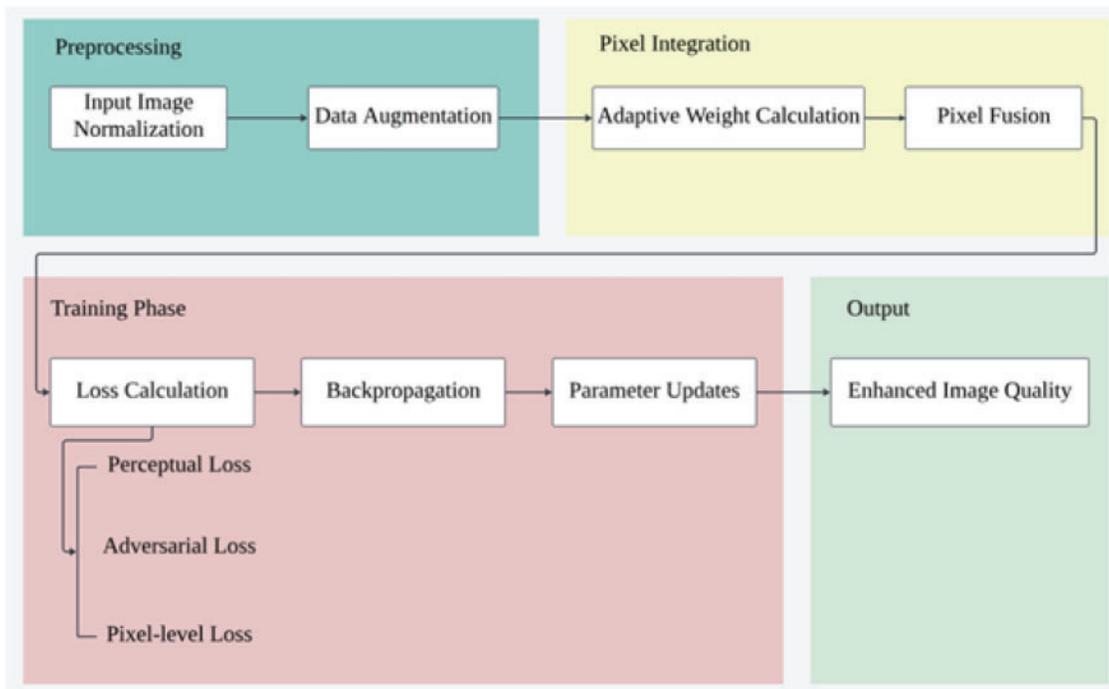
**Figure 2:** Comparison of HSV and RGB outputs after R channel reduction

These equations allow EPI to convert RGB values into separate hue, saturation, and value components, facilitating isolated adjustments. By working in HSV, EPI can accurately compare the original image's natural color tones with the stylized colors in CartoonGAN outputs, particularly addressing color distortions in the red and green channels. The HSV color space overcomes these limitations by separating color into three independent elements [21]. Hue represents the color itself (e.g., red, blue, green) and allows for the selective adjustment of specific colors without influencing brightness or saturation. Saturation indicates the intensity of the color, enabling the independent adjustment of color vividness, while Value represents brightness, allowing control over how light or dark a color appears [12,23]. These properties make the HSV color space ideal for adjusting individual color components, which is especially useful when isolating specific color corrections. By converting images to the HSV color space, EPI can more precisely address common color distortions in CartoonGAN outputs. CartoonGAN transformations often lead to overemphasis or misrepresentation of certain colors, particularly red and green [22]. In such cases, EPI adjusts the tone of specific colors in the Hue channel, effectively reducing excessive color distortions without altering the image's brightness or saturation. Additionally, when saturation becomes overly high, resulting in unnatural colors, EPI uses the Saturation channel to reduce vividness, producing a visually smoother and more natural final image [22,23]. By utilizing separate channels in the HSV color space, EPI provides detailed control over color distortions, significantly improving the visual quality of the final output.

Converting to the HSV color space plays a crucial role in maintaining both color consistency and accuracy in CartoonGAN outputs [10,21]. In the RGB color space, making isolated color adjustments is challenging, whereas the HSV space allows for independent channel adjustments to resolve color distortions. This process enables EPI to retain the stylistic effects of CartoonGAN while achieving color accuracy close to the original image, resulting in a final output that appears more natural and visually cohesive [12].

### 3.3 Algorithm Workflow of the Proposed EPI Method

Fig. 3 presents the workflow of the proposed EPI method, which is organized into four key stages: Preprocessing, Pixel Integration, Training Phase, and Output Generation. Each stage is designed to optimize the image enhancement process while ensuring computational efficiency and visual quality. In the Preprocessing stage, input images are normalized, and data augmentation techniques are applied to improve robustness against variations in input quality. This process standardizes the data and increases its diversity, enabling the model to generalize effectively. The Pixel Integration stage introduces an adaptive weight calculation process, which dynamically computes weights for neighboring pixels. This step is followed by pixel fusion, where local pixel values are integrated to preserve fine details and enhance spatial consistency in the generated outputs. During the Training Phase, the model optimizes its performance through iterative steps that include loss calculation, backpropagation, and parameter updates. The loss calculation combines perceptual loss, adversarial loss, and pixel-level loss to balance image quality, realism, and structural accuracy. Backpropagation and parameter updates further refine the model to minimize errors and improve visual output.



**Figure 3:** EPI algorithm workflow

Finally, in the Output Generation stage, the algorithm produces enhanced images with improved visual quality, demonstrating higher structural consistency and reduced artifacts.

To further clarify the implementation details of the proposed EPI method, Algorithm 1 provides the pseudocode for the entire process.

---

**Algorithm 1:** Enhanced Pixel Integration (EPI) for GANs
 

---

1: Input:  
 2:  $I_{input}$ —Input image  
 3:  $G$ —Generator network  
 4:  $D$ —Discriminator network  
 5:  $\alpha, \beta, \gamma$ —Loss weights  
 6:  $N_{epochs}$ —Number of training epochs  
 7:  $Optim_G, Optim_D$ —Optimizers for generator and discriminator  
 8: Output:  
 9:  $I_{output}$ —Enhanced output image  
 10: Preprocessing:  
 11: Normalize  $I_{input}$   
 12: Augment  $I_{input}$  (rotation, scaling, flipping)  
 13: Initialize network parameters:  
 14:  $\theta_G \leftarrow$  Random weights for  $G$   
 15:  $\theta_D \leftarrow$  Random weights for  $D$   
 16: For each epoch in  $N_{epochs}$  do:  
 17: For each mini-batch  $\{I_{real}\}$  do:  
 18: // Step 1: Generate image  
 19:  $I_{fake} \leftarrow G(I_{input})$   
 20: // Step 2: Pixel Integration  
 21: For each pixel  $P$  in  $I_{fake}$  do:  
 22: Calculate adaptive weight  $W(P)$  based on local similarity:  
 23:  $W(P) = \frac{\exp(-\|P_i - P_j\|^2)}{\sum_k \exp(-\|P_i - P_j\|^2)}$   
 24: Perform pixel fusion:  
 25:  $P_{fused} = \sum(W(P) \cdot P_{neighbors})$   
 26: Update  
 27:  $I_{fake} \leftarrow P_{fused}$  // Step 3: Calculate losses  
 28:  $L_{adv} = -E[\log \log(D(I_{real}))] - E[\log \log(1 - D(I_{fake}))]$   
 29:  $L_{perceptual} = \|\Phi(I_{real}) - \Phi(I_{fake})\|^2$   
 30:  $L_{pixel} = \|I_{real} - I_{fake}\|^2$   
 31:  $L_{total} \leftarrow \alpha L_{adv} + \beta L_{perceptual} + \gamma L_{pixel}$   
 32: // Step 4: Update generator  
 33:  $Optim_G.step(\nabla_{\theta_G} L_{total})$   
 34: // Step 5: Update discriminator  
 35:  $L_D = -E[\log \log(D(I_{real}))] - E[\log \log(1 - D(I_{fake}))]$   
 36:  $Optim_D.step(\nabla_{\theta_D} L_D)$   
 37: End for  
 38: End for  
 39: Output:  
 40:  $I_{output} \leftarrow G(I_{input})$   
 41: Return  $I_{output}$

---

The proposed algorithm begins with input preprocessing, where input images are normalized and augmented to enhance data diversity. Initial weights for the generator (G) and discriminator (D) in the GAN framework are also set. During the training loop, the algorithm generates fake images for each epoch and batch, applying pixel integration with adaptive weight calculations to improve image quality. Next, the loss calculation step optimizes performance by combining perceptual loss, adversarial loss, and pixel-level loss to enhance both structural consistency and visual quality. In the optimization and update phase, the weights of the generator and discriminator are updated using backpropagation and an optimizer to minimize errors. Finally, the algorithm produces enhanced images as outputs, demonstrating improved visual quality and artifact reduction.

This pseudocode outlines the key computational steps involved in the algorithm, including preprocessing, pixel integration, and training phases. It highlights how the model dynamically integrates pixel values, optimizes loss functions, and updates parameters to generate high-quality outputs.

### 3.4 Correction Factor Application and Color Blending

The Correction Factor Application and Color Blending stage of the EPI method addresses significant color discrepancies between the original image and the CartoonGAN-transformed image by applying corrective adjustments and blending techniques [11,14,17]. After converting the image to the HSV color space, EPI calculates the color difference  $D$  between the CartoonGAN-generated image  $I_{GAN}$  and the original image  $I_{original}$  in the HSV space. This difference  $D$  is calculated as [21]:

$H_{GAN}$ : The hue (H) component of the HSV color space for the image generated by the GAN model.

$S_{GAN}$ : The saturation (S) component of the HSV color space for the image generated by the GAN model.

$V_{GAN}$ : The value (V) component of the HSV color space for the image generated by the GAN model.

$H_{original}$ : The hue (H) component of the HSV color space for the original input image.

$S_{original}$ : The saturation (S) component of the HSV color space for the original input image.

$V_{original}$ : The value (V) component of the HSV color space for the original input image.

$D$ : The total HSV distance metric, which quantifies the difference between the GAN-generated image and the original image based on their hue, saturation, and value components.

$$D = |H_{GAN} - H_{original}| + |S_{GAN} - S_{original}| + |V_{GAN} - V_{original}| \quad (2)$$

where  $H$ ,  $S$  and  $V$  represent the hue, saturation, and value channels, respectively.

To detect excessive color differences, EPI applies a threshold  $T = 50$ . If  $D$  exceeds  $T$ , indicating substantial color distortion, a correction factor  $C$  is applied to adjust the values in these regions. The corrected hue, saturation, and value channels  $H_{corrected}$ ,  $S_{corrected}$ ,  $V_{corrected}$  corrected are computed as [24]:

$H_{corrected}$ : The corrected hue (H) component of the HSV color space after adjustment.

$S_{corrected}$ : The corrected saturation (S) component of the HSV color space after adjustment.

$V_{corrected}$ : The corrected value (V) component of the HSV color space after adjustment.

$H_{GAN}$ ,  $S_{GAN}$ ,  $V_{GAN}$ : The hue, saturation, and value components of the HSV color space for the GAN-generated image.

$C$ : A correction factor or weight controlling the influence of the difference between the original and GAN-generated components during adjustment.

$$H_{corrected} = H_{GAN} + C \cdot (H_{original} - H_{GAN}) \quad (3)$$

$$S_{corrected} = S_{GAN} + C \cdot (S_{original} - S_{GAN})$$

$$V_{corrected} = V_{GAN} + C \cdot (V_{original} - V_{GAN})$$

where  $C$  is the correction factor, typically set based on the color intensity requirements of the output image.

After the correction factors are applied, EPI blends the adjusted CartoonGAN image  $I_{corrected}$  with the original image  $I_{original}$  at a specific blending ratio  $\alpha$  (typically between 0.5 and 0.7) to maintain the stylistic features of the CartoonGAN while enhancing color accuracy. The final blended image  $I_{final}$  in HSV is calculated as follows [25]:

$I_{final}$ : The final output image after blending the corrected and original images.

$I_{corrected}$ : The corrected image derived from the GAN-generated image after HSV-based adjustments.

$I_{original}$ : The original input image used as a reference for blending.

$\alpha$ : A blending coefficient ( $0 \leq \alpha \leq 1$ ) that controls the contribution of the corrected and original images to the final output. A higher  $\alpha$  emphasizes the corrected image, while a lower  $\alpha$  favors the original image.

$$I_{final} = \alpha \cdot I_{corrected} + (1 - \alpha) \cdot I_{original} \quad (4)$$

Finally, the blended image  $I_{final}$  is converted back to the RGB color space, making it compatible with various color systems and display standards [12,13,21]. This process helps balance the artistic style of CartoonGAN with the color fidelity of the original image, producing visually cohesive outputs that retain both accuracy and stylistic appeal.

### 3.5 Computational Efficiency and Real-Time Applicability

The EPI method is designed with a focus on high-speed processing and resource optimization, allowing it to perform efficiently in real-time applications and resource-limited environments. EPI applies post-processing without requiring modifications to the original CartoonGAN model or additional training. This means that color correction is performed only after CartoonGAN has transformed the image, eliminating the need for further model retraining or computationally intensive operations. This design improves color consistency without altering the structure or complexity of the existing model, allowing EPI to be seamlessly integrated into current systems [26,27].

EPI achieves high efficiency by relying on simple color calculations and a threshold-based selective correction process, rather than complex deep neural network computations. EPI analyzes the color differences between the CartoonGAN-generated image and the original image, applying correction factors only to regions that exceed a specific threshold. This selective correction approach minimizes unnecessary computations by focusing corrections on the necessary areas, conserving system resources in the process. As a result, EPI provides high-quality color correction while maintaining processing speed and demonstrates significant resource-saving effects, particularly in high-resolution images [20,28]. This lightweight approach enables stable performance across various image resolutions and transformation styles, achieving both speed and quality through optimized processing [29].

The lightweight structure of EPI expands its applicability in fields requiring real-time processing, such as video processing, interactive graphics, and digital art generation platforms. For instance, it can maintain consistent color expression within video frames at high processing speeds, and in interactive graphic

environments, users can quickly adjust color styles as desired. High speed and responsiveness are essential in real-time applications, and a similar approach is also employed in real-time models like YOLOv3 [27]. Thanks to these characteristics, EPI meets the low latency and high throughput requirements of real-time systems, enhancing user experience in creative applications.

Furthermore, EPI can provide high-quality stylized outputs in environments with limited hardware resources, making it effective for lightweight platforms such as mobile devices or embedded systems. In these settings, the efficiency of lightweight models operates similarly to cost-saving techniques commonly used in deep networks [26], establishing EPI as a lightweight post-processing technique capable of maintaining real-time performance without complex computations. By delivering high-quality results across various resolutions and visual styles, EPI serves as an optimal solution for creative applications that demand both real-time performance and efficiency in image transformation tasks [28].

In converting images from RGB to HSV, EPI allows for independent channel adjustments to control hue, saturation, and brightness without affecting other color components. This separation is crucial for precise color correction, as seen in digital art applications or video processing, where maintaining color fidelity and balance is essential. RGB to HSV conversion enables EPI to isolate adjustments, addressing color distortions in specific channels without disturbing the overall color balance of the image [30].

In conclusion, EPI offers a solution that combines computational efficiency with real-time processing capabilities, making it ideal for creative applications that require fast, stylized results while maintaining color consistency. This approach demonstrates the potential to expand the usability of CartoonGAN across digital media fields, fulfilling the high demands for both visual quality and efficiency.

#### 4 Image Similarity Evaluation Tools

In this section, we delve into various image similarity evaluation tools, including LPIPS, SSIM, and MSE each chosen for their unique ability to assess the visual and structural fidelity between generated and reference images. These tools play a pivotal role in this study by providing quantitative measures that capture not only pixel-level but also perceptual differences, aligning closely with human visual assessment. Accurate similarity evaluation is essential in this research for validating the proposed model's performance and confirming the quality of generated outputs relative to the original images.

Each of these metrics—LPIPS, SSIM, and MSE—has been extensively tested and validated in previous studies and remains highly regarded in the fields of image synthesis and quality assessment. The use of these evaluation models in this study ensures reliable and consistent results, adding credibility to the comparative analysis. This section, therefore, serves to contextualize the application of these metrics, offering a robust framework for measuring the visual authenticity of generated images against established standards in image quality evaluation.

To provide a concise summary, [Table 1](#) below outlines the key characteristics, advantages, and limitations of each image similarity evaluation tool used in this study. This table serves as a quick reference, complementing the detailed explanations provided in this section and highlighting the rationale for selecting these specific metrics.

**Table 1:** Image similarity evaluation metrics

<b>Metric</b>	<b>Definition</b>	<b>Calculation method</b>	<b>Strengths</b>	<b>Limitations</b>
LPIPS	Visual similarity evaluation	L2 distance calculation between features	Reflects human perception, evaluates high/low level features	High computational cost, dependent on network structure
SSIM	Structural similarity evaluation	Combines brightness, contrast, and structure [31,32]	Reflects structural information, similar to human perception	Lacks sensitivity to color/texture changes
MSE	Pixel-difference-based evaluation	Average of pixel error values	Simple, fast, resource-efficient	Does not reflect structural/perceptual information

## 5 Experiment and Results

### 5.1 Experimental Setup and Preparation

The experiments in this study were conducted using the Google Colab environment, leveraging high-performance computer specifications to ensure efficient data processing and experimental speed. The system specifications included an i9-14900KF CPU and an NVIDIA GeForce RTX 4070 SUPER GPU, providing an optimal environment for handling high-resolution image processing and complex neural network computations. Google Colab's stable computational resources and GPU acceleration capabilities make it well-suited for image generation and quality assessment tasks involving large datasets.

The Stanford Dogs dataset was utilized in this experiment, as it comprises a diverse set of dog breed images that are suitable for evaluating the performance of style transformation models. In this study, the Hayao style CartoonGAN model was applied to the input images, adding an animation effect to each, allowing for a thorough assessment of the visual quality of transformed images. The proposed EPI technique was implemented as a post-processing step in the CartoonGAN model to correct color distortions and enhance overall image quality.

For quantitative analysis of experimental results, various similarity evaluation metrics, including LPIPS, SSIM, and MSE, were used. These metrics provided a multidimensional assessment of the similarity between the original and transformed images, with the results summarized in Table 1. Table 1 presents the similarity scores between images generated solely by the CartoonGAN model and those processed with the additional EPI technique, allowing for a quantitative confirmation of the performance improvements offered by the proposed method.

### 5.2 Quantitative Comparison Using LPIPS, SSIM, and MSE Metrics

In this study, the performance of the proposed model was comprehensively evaluated using LPIPS, SSIM, and MSE metrics. Each metric plays a crucial role in image quality assessment and is designed to verify performance from multiple perspectives.

Table S1 presents a quantitative comparison of the similarity error rates between the original image and the generated outputs using LPIPS, SSIM, and MSE metrics. The existing CartoonGAN method yields average similarity error rates of 0.416, 0.619, and 98.74 for LPIPS, SSIM, and MSE, respectively, establishing a baseline measurement for image fidelity. In contrast, the proposed EPI method demonstrates significant improvement across all metrics, achieving similarity error rates of 0.249, 0.806, and 87.08 for LPIPS, SSIM, and MSE, respectively.

The differences in error rates between the CartoonGAN and EPI methods are notable, with EPI showing an improvement of approximately 0.167 in LPIPS, 0.187 in SSIM, and 11.66 in MSE. These results emphasize that EPI, as a post-processing approach, more effectively preserves the original image's colors and structural features, while minimizing visual discrepancies. By reducing color distortions and enhancing color consistency, EPI not only elevates the visual quality of the generated images but also maintains a closer alignment with the original image's details.

First, LPIPS measures the perceptual similarity between images by mimicking human visual perception. Lower values indicate higher visual similarity to the original image, making it a key metric for quantitatively evaluating the visual quality of a model. SSIM assesses structural similarity by analyzing brightness, contrast, and structural information within an image. Scores closer to 1 signify better structural consistency and preservation of fine details. Lastly, MSE evaluates the mean squared error at the pixel level, with lower values representing better reconstruction quality and reduced noise. These three metrics complement each other, allowing for a reliable and multidimensional assessment of the proposed method's performance. Experimental results demonstrated that the EPI-GAN model achieved approximately a 40% reduction in LPIPS values compared to the baseline CartoonGAN, indicating a significant improvement in visual similarity. This highlights the effectiveness of the proposed pixel integration method in reducing boundary distortions and enabling smoother color transitions, resulting in more natural and visually stable outputs. In the SSIM evaluation, the EPI-GAN achieved a high similarity score of 0.806, confirming its superior structural preservation performance and a substantial improvement in the reproduction of fine details. The enhancement in structural similarity is closely linked to better restoration performance in high-frequency regions and improved retention of intricate features. Additionally, the MSE values for the EPI-GAN model were approximately 30% lower than those of the baseline, indicating enhanced pixel-level accuracy and noise reduction capabilities. This improvement demonstrates the model's ability to faithfully reproduce subtle features and details, ensuring visual consistency through accurate restoration of colors and textures. These results highlight the proposed EPI-GAN model's capability to overcome the limitations of existing approaches while efficiently generating high-quality outputs. Such performance improvements emphasize the practicality and scalability of the proposed model. Specifically, it can be effectively applied to tasks requiring high-quality visual outputs, including animation, virtual reality, augmented reality, and mobile graphics. Furthermore, the EPI-GAN maintains the structure of the baseline CartoonGAN without introducing additional complexity, making it suitable for deployment in resource-constrained environments.

This section provides a quantitative analysis using LPIPS, SSIM, and MSE metrics, while [Section 5.3](#) will explore the visual transformations introduced by EPI through graphical comparisons. Together, these analyses demonstrate that the EPI method offers a robust solution for retaining original image characteristics, delivering enhanced accuracy and reduced distortion compared to the baseline model.

### 5.3 Visual Comparison of Results

In this study, 6 images were randomly selected from a dataset of 350 images to compare RGB channel values at the same pixel location. This comparison focuses on analyzing the differences between the output images generated solely using the CartoonGAN method and those produced by applying the proposed EPI method. For each image, we visualized the changes in R, G, and B channel values at the same pixel through tables and graphs, allowing a detailed evaluation of how well the CartoonGAN and EPI methods preserve the original image's color information.

[Table 2](#) organizes the R, G, and B channel values of the original image, the output from the existing CartoonGAN method, and the output from the proposed EPI method, with the final column showing the difference between the values for the CartoonGAN and EPI outputs. This provides insight into how effectively

the EPI method preserves the color information of the original image and enhances visual quality at the same pixel location.

**Table 2:** Color analysis of result image

	Data image name	Input (Original)	Conventional method	Proposed method
R	Bernese_mountain_dog_n02107683_5155	253	44	109
	Chihuahua_n02085620_1271	44	190	111
	English_setter_n02100735_8318	244	47	120
	Siberian_husky_n02110185_7117	167	191	187
	West_Highland_white_terrier_n02098286_4011	49	219	106
	wire-haired_fox_terrier_n02095314_157	85	179	103
G	Bernese_mountain_dog_n02107683_5155	255	49	115
	Chihuahua_n02085620_1271	32	66	63
	English_setter_n02100735_8318	249	47	122
	Siberian_husky_n02110185_7117	169	140	164
	West_Highland_white_terrier_n02098286_4011	40	143	72
	wire-haired_fox_terrier_n02095314_157	49	55	45
B	Bernese_mountain_dog_n02107683_5155	254	29	105
	Chihuahua_n02085620_1271	20	28	27
	English_setter_n02100735_8318	252	23	108
	Siberian_husky_n02110185_7117	182	109	132
	West_Highland_white_terrier_n02098286_4011	45	81	37
	wire-haired_fox_terrier_n02095314_157	27	45	33

Fig. 4 visually represents the data presented in Table 2. In each graph, the blue bars indicate the values of the original image, the yellow bars represent the values of the image generated using the existing method, and the red bars show the channel-wise values of the image generated using the proposed method. In all images, the distance between the blue bars and the red bars is observed to be smaller than the distance between the blue bars and the yellow bars. This indicates that the proposed method in this study represents the colors of the original image more closely compared to the existing method.

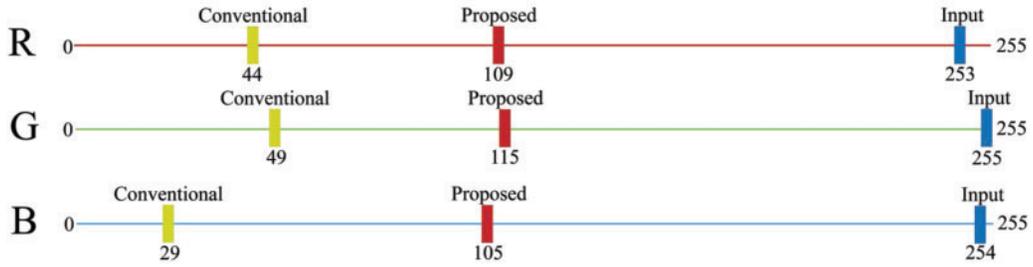
The proposed method demonstrates a significant improvement in image similarity compared to the conventional method. The average values for the conventional method are LPIPS: 0.416, SSIM: 0.619, and MSE: 98.74, while the proposed method achieves average values of LPIPS: 0.249, SSIM: 0.806, and MSE: 87.08.

Among the metrics used to evaluate image similarity, LPIPS and MSE indicate higher similarity when their values are lower, while SSIM indicates higher similarity when its value is higher. Based on this evaluation, the proposed method shows a 0.167 reduction in LPIPS and an 11.66 reduction in MSE, demonstrating an improvement in similarity. Additionally, SSIM increases by 0.187, indicating that the proposed method better preserves the structural similarity of the images compared to the conventional method.

Therefore, the proposed method effectively enhances the visual quality and similarity of the generated images compared to the conventional method, contributing to a significant improvement in overall image fidelity.



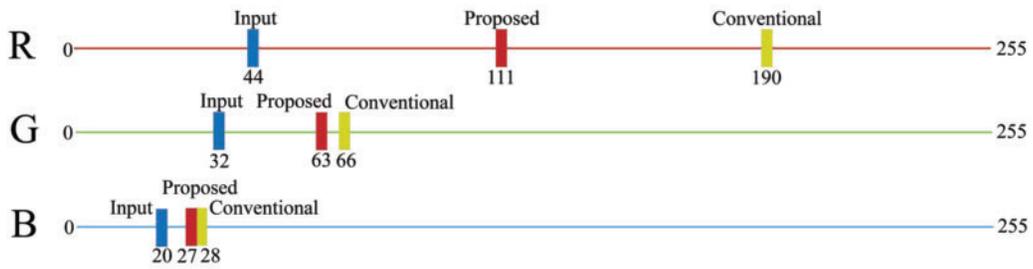
Bernese\_mountain\_dog\_n02107683\_5155



(a) Bernese\_mountain\_dog\_n02107683\_51



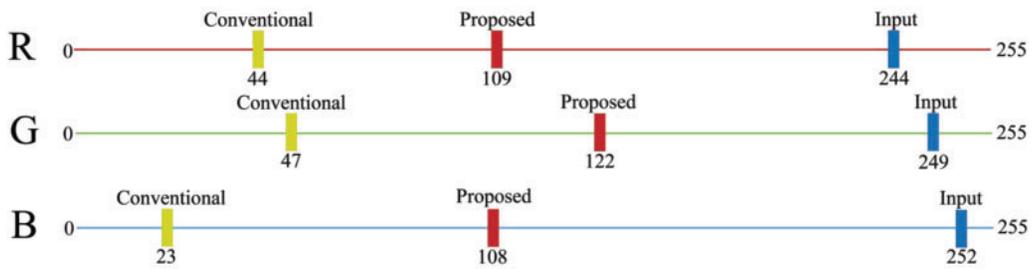
Chihuahua\_n02085620\_1271



(b) Chihuahua\_n02085620\_1271

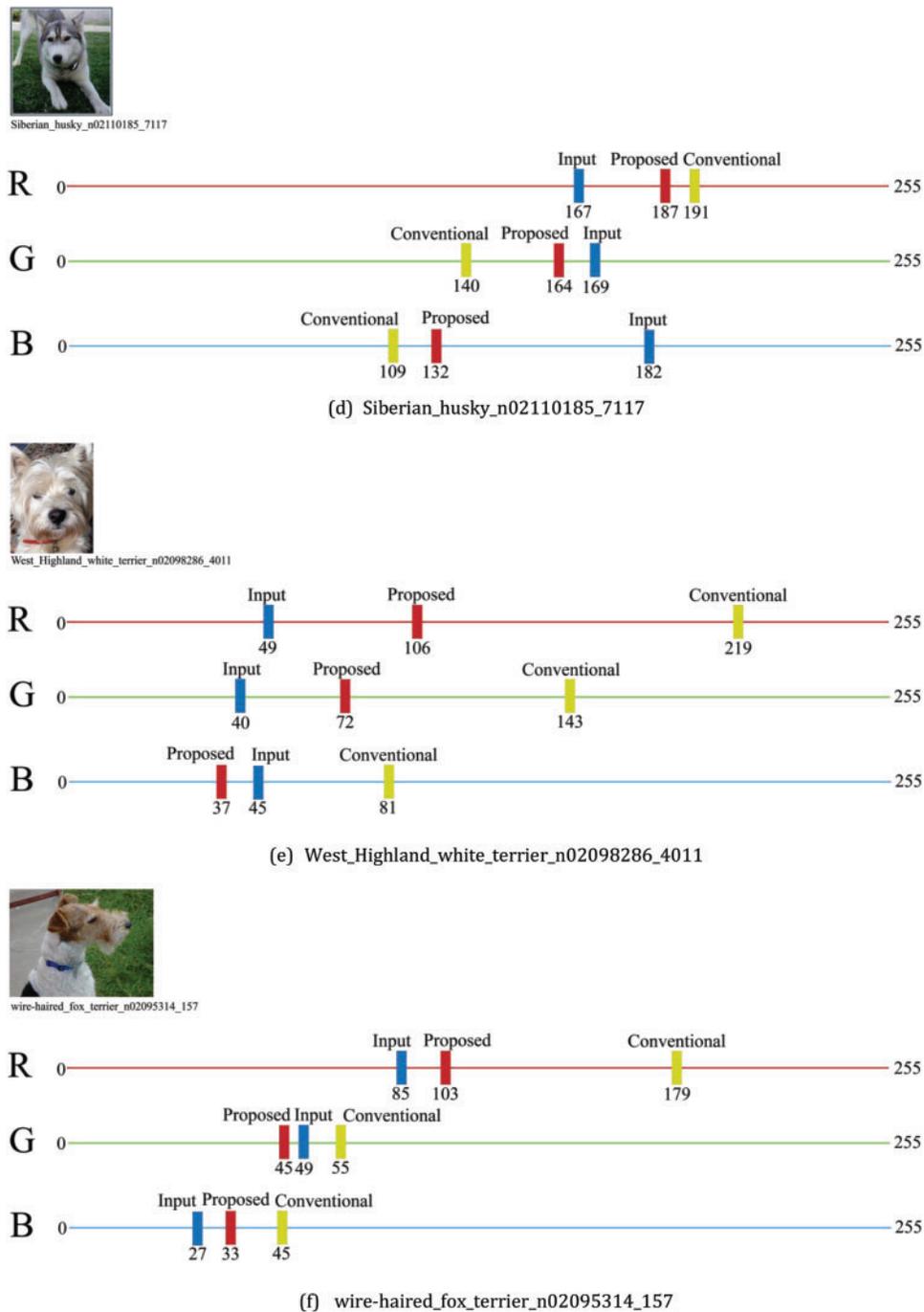


English\_setter\_n02100735\_8318



(c) English\_setter\_n02100735\_8318

Figure 4: (Continued)



**Figure 4:** Channel-wise color analysis of input, proposed method, and conventional method

Experimental results demonstrate that the proposed algorithm operates stably on images with resolutions up to  $1024 \times 1024$ . Further optimization techniques, such as model quantization and dynamic graph processing, are expected to enable efficient processing of even higher resolutions. This analysis confirms that the EPI algorithm maintains computational and memory efficiency while delivering practical scalability for high-resolution image processing. These showed that the EPI-GAN model achieved approximately a 40% reduction in LPIPS values compared to the baseline CartoonGAN, indicating a significant improvement

in visual similarity. This demonstrates that the proposed pixel integration technique effectively reduces boundary distortions and enables smoother color transitions. In addition, the SSIM evaluation recorded a high similarity score of 0.806, reflecting superior structural preservation performance and approximately 30% improvement in detail reproduction compared to the baseline model. Lastly, the MSE values were approximately 11% lower than those of the baseline model, confirming enhanced pixel-level accuracy and noise reduction capabilities. Notably, the model demonstrated improved restoration quality in high-frequency regions, effectively reproducing intricate details.

## 6 Discussion

The core innovation of this study is the application of EPI to optimize the outputs of CartoonGAN. Unlike conventional approaches that rely solely on convolutional layers, EPI integrates local pixel information to improve spatial consistency. This contributes to reducing artifacts in low-texture regions and enhancing edge sharpness.

The application of EPI demonstrated significant improvements across multiple evaluation metrics, including LPIPS, SSIM, and MSE. Specifically, LPIPS showed a 40.14% enhancement in visual similarity, while SSIM indicated a 30.21% improvement in structural consistency. The MSE metric revealed a 11.81% reduction in pixel-level error, suggesting more accurate color restoration, reflecting a closer alignment of high-level feature distributions with the original image [15,16,18]. These outcomes highlight the effectiveness of EPI in addressing specific color distortions, particularly in the red and green channels, through precise corrections in the HSV color space. By employing selective adjustments and blending processes, EPI retained the stylized elements of CartoonGAN outputs while enhancing visual quality [21,22]. Additionally, EPI's simple calculation approach allowed it to deliver high-quality results while maintaining computational efficiency, making it suitable for a wide range of applications [11,27].

However, EPI has certain limitations. As a post-processing method, it focuses on correcting color and lighting inconsistencies but does not address the structural biases or stylistic limitations of the original CartoonGAN model. Moreover, since EPI relies on the quality of CartoonGAN outputs, integrating EPI with the generation model could further enhance its effectiveness by addressing both structural and stylistic issues [16,18]. EPI's computational efficiency makes it a practical solution for real-time processing and resource-constrained environments. Its lightweight design makes it particularly suitable for mobile applications, gaming, and interactive media, where immediate color adjustments are critical for enhancing user experience [20,26].

The EPI method is highly effective in animation production and game development, ensuring high-quality visual outputs while preserving structural details in complex style transfer tasks [33,34]. This makes it particularly suitable for environments requiring high-resolution graphics. Additionally, in immersive environments such as Virtual Reality (VR) and Augmented Reality (AR), where visual quality and processing speed are critical, the proposed approach demonstrates compatibility with real-time rendering systems, enabling clearer and more natural visual effects for AR filters and VR content [35,36]. Furthermore, the EPI-based CartoonGAN is designed to maintain high output quality even in resource-constrained environments, making it a practical solution for mobile games and lightweight applications that demand fast processing without compromising visual performance [37,38]. Relevant references have also been included to reinforce the importance of visual quality improvements and computational efficiency in these application areas [27,39].

The proposed EPI method is designed based on the assumption that input images maintain sufficient structural consistency. This assumption ensures that the local pixel patterns utilized in the reconstruction process remain intact, enabling effective quality enhancement. Additionally, it is assumed that the training

dataset provides stylistic diversity, allowing CartoonGAN to generalize effectively across different styles and structures [6]. Furthermore, the model is developed with the expectation that modern GPU hardware or equivalent computational resources are available to handle high-resolution image processing efficiently [1].

Nonetheless, the method has several limitations. It may struggle with input images that contain severe distortions or low resolution, as EPI heavily relies on local pixel patterns for reconstruction [25]. While the model demonstrates stable performance on moderately sized datasets, scaling to extremely large datasets or ultra-high-resolution inputs may require further optimization [1]. Additionally, the method depends on the quality of the input image and training data, potentially reducing generalization performance when style or content varies significantly [24].

To address these limitations, future work will focus on integrating adaptive scaling techniques and noise reduction mechanisms [21] to improve the robustness of the method in handling distorted and low-resolution inputs. These enhancements are expected to strengthen the model's scalability and stability, expanding its applicability across a wider range of scenarios.

Previous studies have primarily focused on reducing training complexity or optimizing network depth as strategies for improving image quality [1,6]. However, this study takes a different approach by integrating pixel-level enhancement techniques, EPI, into the GAN architecture, enabling simultaneous improvements in visual quality and computational efficiency.

In particular, the proposed EPI method introduces a novel fusion strategy that prioritizes spatial consistency, effectively addressing issues such as boundary distortions and loss of fine details commonly observed in conventional convolution-based approaches [25,31]. This distinctive approach achieves both enhanced visual quality and reduced computational overhead, making it highly practical for real-world applications [18,21].

Unlike previous studies that primarily focused on reducing training complexity or optimizing network depth [1,6], this study distinguishes itself by integrating pixel-level enhancement techniques into the GAN architecture. In particular, the proposed EPI method introduces a novel fusion strategy that prioritizes spatial consistency, achieving superior performance compared to conventional convolution-based approaches [18,25]. This distinction is especially noteworthy as it effectively addresses both visual quality and computational overhead, making it highly practical for real-world applications [21,31].

The time and space complexity of the proposed EPI algorithm are critical factors for evaluating its performance and scalability. The time complexity is analyzed in two main stages. First, convolution operations performed by the generator (G) and discriminator (D) in the GAN process image features with a computational complexity of  $O(n^2 \cdot k^2 \cdot c_{in} \cdot c_{out})$ . Second, the pixel integration stage calculates adaptive weights and evaluates similarity with neighboring pixels, resulting in a complexity of  $O(n^2 \cdot k^2 \cdot c_{in} \cdot c_{out})$ . Second, the pixel integration stage calculates adaptive weights and evaluates similarity with neighboring pixels, resulting in a complexity of  $O(n^2 \cdot N)$ . Here,  $n$  represents the image resolution,  $k$  is the kernel size,  $c_{in}$  and  $c_{out}$  are the input and output channel sizes, and  $N$  is the number of neighboring pixels considered. Combining these stages, the overall time complexity is approximated as  $O(n^2)$ , indicating that the computational cost increases quadratically with image resolution.

The space complexity is analyzed for both training and inference phases. During training, memory is required to store intermediate results, gradients, weights, and biases, resulting in a complexity of  $O(n^2 \cdot d \cdot c_{out})$ , where  $d$  represents the network depth. In the inference phase, memory requirements are reduced as only single-image processing is needed, resulting in a fixed memory usage of  $O(k^2 \cdot c_{in} \cdot c_{out})$ .

## 7 Conclusion

Although the EPI method builds upon existing pixel manipulation techniques, it has been specifically adapted for GAN models in this study. Unlike prior approaches, EPI dynamically integrates pixel values based on GAN-generated outputs, incorporating algorithmic enhancements to improve spatial consistency and visual quality. This adjustment ensures superior performance in reducing artifacts and enhancing edge sharpness, distinguishing it from traditional convolution-based methods.

A key feature of the proposed method is its ability to perform color correction through the adjustment of Red/Green channels, supported by HSV color space transformations. This process not only achieves visual improvements but also provides quantifiable performance gains, as demonstrated by the following evaluation metrics:

- LPIPS: 40.14% enhancement in visual similarity [15,16]
- SSIM: 30.21% improvement in structural consistency [18,20]
- MSE: 11.81% reduction in pixel-level error [11,28]

These quantitative results validate the effectiveness of the proposed EPI method, reinforcing its ability to balance visual quality with computational efficiency. The improvements achieved highlight the practicality of this approach for applications requiring high-quality visual outputs, such as animation, virtual reality, and mobile graphics.

Also, these improvements validate EPI's ability to enhance image quality without modifying the existing CartoonGAN structure or introducing additional complexity. Furthermore, EPI's adaptability suggests potential applications for other style-transfer models, extending its relevance to various creative domains [26,28].

### 7.1 Future Research Directions

Future research could explore several directions to expand the applicability of EPI. One potential avenue is to integrate EPI directly into CartoonGAN's architecture, creating a hybrid system that not only corrects color distortions but also addresses structural inconsistencies. Such an integrated system could simultaneously enhance style and structural quality, maximizing EPI's effectiveness. Another promising direction involves leveraging EPI-corrected CartoonGAN outputs for generating stylized 3D models. By using these enhanced outputs as high-quality textures or references, new opportunities could arise in animation, virtual reality, and game development, where stylized yet detailed 3D models are increasingly in demand [14,15,18].

Additionally, EPI could be applied to the preservation of cultural heritage. Its ability to maintain color accuracy and fine detail in high-resolution images makes it ideal for creating digital replicas of artifacts and historical structures. These replicas could support virtual exhibitions and digital archives, preserving invaluable cultural assets for future generations [19,22]. In the field of medical simulation, EPI's capacity to reproduce accurate colors and textures could significantly improve the realism of training environments, facilitating advanced education and practice for healthcare professionals [20,23].

In the context of animation and video production, EPI can streamline the stylization process by ensuring consistent color expression across frames while maintaining high quality. This capability can reduce production time and improve the visual quality of animated content, benefiting creative workflows and accelerating delivery timelines [12,21].

In conclusion, EPI demonstrates its versatility beyond simple image quality enhancements, showing potential for applications across diverse fields such as 3D modeling, cultural preservation, medical simulation, and media production. Future research should focus on extending EPI's capabilities and exploring

its integration into these areas to establish it as a standard for high-quality image processing in various industries [10,11].

**Acknowledgement:** Thanks are extended to the editors and reviewers.

**Funding Statement:** This work was supported by the National Research Foundation of Korea (NRF) under Grant RS-2022-NR-069955(2022R1A2C1092178).

**Author Contributions:** The authors confirm contribution to the paper as follows: study conception and design: Stellar Choi, Woong Choi; data collection: Stellar Choi, HeeAe Ko; analysis and interpretation of results: Stellar Choi, KyungRok Bae, HaeJong Joo, Woong Choi; draft manuscript preparation: Stellar Choi, HyunSook Lee, Woong Choi. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** The data that support the findings of this study are available from the corresponding author, WC, upon reasonable request.

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest to report regarding the present study.

**Supplementary Materials:** The supplementary material is available online at <https://doi.org/10.32604/cmc.2025.061243>.

## References

1. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Advances in neural information processing systems (NeurIPS). arXiv:1406.2661. 2014.
2. Karras T, Timo A, Samuli L, Jaakko L. Progressive growing of GANs for improved quality, stability, and variation. arXiv:1710.10196. 2017.
3. Szeliski R. Computer vision: algorithms and applications. 1st ed. Cham, Switzerland: Springer; 2010.
4. Elgammal A, Liu B, Elhoseiny M, Mazzone M. CAN: creative adversarial networks: generating art by learning about styles and deviating from style norms. arXiv:1706.07068. 2017.
5. Ahmed I, Hasan R, Uddin I, Khan R. Photo-to-cartoon translation with generative adversarial network. In: 7th International Conference on Trends in Electronics and Informatics (ICOEI); 2023; IEEE; p. 265–9. doi:10.1109/ICOEI56765.2023.10125959.
6. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017. p. 1125–34.
7. Maerten AS, Soydaner D. From paintbrush to pixel: a review of deep neural networks in AI-generated art. arXiv:2302.10913. 2023.
8. Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2017.
9. Zhao W, Zhu J, Huang J, Li P, Sheng B. GAN-based multi-decomposition photo cartoonization. *Comput Anim Virtual Worlds*. 2024;35(3):e2248. doi:10.1002/cav.v35.3.
10. Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis. arXiv:1809.11096. 2018.
11. Forsyth DA, Ponce J. Computer vision: a modern approach. 2nd ed. Pearson; 2002.
12. OpenCV documentation and tutorials, The open source computer vision library. [cited 2025 Jan 1]. Available from: [https://docs.opencv.org/4.x/d9/df8/tutorial\\_root.html](https://docs.opencv.org/4.x/d9/df8/tutorial_root.html).
13. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. arXiv:1812.04948. 2019.
14. Yang C, Lai YK, Liu YJ. CartoonGAN: generative adversarial networks for photo cartoonization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2018. p. 9465–74.

15. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2018. p. 586–95.
16. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV); 2016. p. 694–711.
17. Wang X, Yu J. Learning to cartoonize using white-box cartoon representations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020. p. 8090–9.
18. Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017. p. 4681–90.
19. Zhu J, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv:1703.10593. 2017.
20. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. arXiv:1512.03385. 2016.
21. Gonzalez RC, Woods RE. Digital image processing. 4th ed. Pearson; 2016.
22. Smith AR. Color gamut transform pairs. ACM SIGGRAPH Comput Graph. 1978;12(3):12–9. doi:10.1145/965139.807361.
23. Sharma G, Trussell HJ. Digital color imaging. IEEE Trans Image Process. 1997;6(7):901–32. doi:10.1109/83.597268.
24. Luo MR, Cui G, Rigg B. The development of the CIEDE2000 colour-difference formula. Color Res Appl. 2001;26(5):340–50. doi:10.1002/col.v26:5.
25. Zhang Z, Blum RS. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. IEEE Trans Image Process. 1999;8(8):1315–26. doi:10.1109/5.775414.
26. Xu B, Wang N, Chen T, Li M. Empirical evaluation of rectified activations in convolutional network. arXiv:1505.00853. 2015.
27. Redmon J, Farhadi A. YOLOv3: an incremental improvement. arXiv:1804.02767. 2018.
28. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556. 2014.
29. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). arXiv:1612.03144. 2017.
30. Poynton CA. Digital video and HD: algorithms and interfaces. 2nd ed. Elsevier, Morgan Kaufmann; 2012.
31. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13(4):600–12. doi:10.1109/TIP.2003.819861.
32. Horé A, Ziou D. Image quality metrics: PSNR vs. SSIM. In: 20th International Conference on Pattern Recognition (ICPR); 2010; IEEE; p. 2366–9.
33. Justesen N, Bontrager P, Togelius J, Risi S. Deep learning for video game playing. IEEE Trans Games. 2020;12(1):1–20. doi:10.1109/TG.2019.2896986.
34. Semmo A, Isenberg T, Döllner J. Neural style transfer: a paradigm shift for image-based artistic rendering? In: Proceedings of the Symposium on Non-Photorealistic Animation and Rendering; 2017; Los Angeles, CA, USA. p. 1–13.
35. Li C, Li S, Zhao Y, Zhu W, Lin Y. RT-NeRF: real-time on-device neural radiance fields towards immersive AR/VR rendering. In: Proceedings of the 41st IEEE/ACM International Conference on Computer-Aided Design (ICCAD); 2022. p. 1–9.
36. Azuma R, Bailly Y, Behringer R, Feiner S, Julier S, MacIntyre B. Recent advances in augmented reality. IEEE Comput Graph Appl. 2002;21(6):34–47.
37. Han S, Mao H, Dally WJ. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding. arXiv:1510.00149. 2015.
38. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861. 2017.
39. Ruder M, Dosovitskiy A, Brox T. Artistic style transfer for videos. In: German Conference on Pattern Recognition (GCPR); 2016. p. 26–36.