IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 October 2023, accepted 16 November 2023, date of publication 20 November 2023,
date of current version 29 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3335225

RESEARCH ARTICLE

Image Colorization Using Color-Features and Adversarial Learning

HAMZA SHAFIQ AND BUMSHIK LEE, (Member, IEEE)
Department of Information and Communication Engineering, Chosun University, Gwangju 61452, South Korea
Corresponding author: Bumshik Lee (bslee@chosun.ac.kr)

This work was supported by a research fund from Chosun University, 2022.

ABSTRACT In this study, we introduce an innovative image colorization method that not only improves color accuracy and realism but also addresses common issues found in existing methods, such as desaturation and color bleeding. Our proposed method features a novel component called the Color Encoder, which extracts intrinsic color features. Moreover, the proposed Color Encoder aligns essential color features systematically, drawn from a random normal distribution, with real colors. These aligned features are fused at the bottleneck and serve as the foundation for subsequent colorization. Complementing the Color Encoder is our Color Loss mechanism, which aims to align the extracted features from the Color Encoder with the ground-truth color features, enhancing overall color representation accuracy. We also employ a Conditional Wasserstein Generative Adversarial Network (CWGAN) architecture within the framework of a Generative Adversarial Network (GAN) to improve adversarial training and colorization accuracy. To enhance feature representation, we incorporate an attention mechanism at the bottleneck of each encoder layer, further refining our model's ability to capture essential image details. Experimental results show that our approach significantly outperforms other state-of-the-art methods in terms of both realism and precision, striking a well-balanced performance.

INDEX TERMS Image colorization, generative adversarial network, image enhancement.

I. INTRODUCTION

Image colorization is the process of adding color to grayscale images, which can be traced back to their origins in traditional artistic practices. Historically, artists employed a manual process to add color to monochromatic photographs, dedicating meticulous attention to preserving the integrity of the original image and maintaining realism. The procedure required significant labor intensity and relied heavily on artistic skill [1]. The processes for colorizing images also evolved alongside the development of technology. Semi-automated colorization techniques became prevalent in the era of digital images and sophisticated software, relying on user inputs to guide the colorization process. These strategies include scribble-based colorization, in which the user provides color scribbles as input, and the software propagates these colors throughout the image [2]. Although these techniques were able to decrease the amount of manual labor involved, they still required a certain degree of artistic expertise and a knowledge of color theory to achieve the desired results.

Over the past decade, image colorization has significantly evolved owing to the emergence of machine learning, particularly deep learning. This advancement has led to a shift from a manual and artist-dependent approach to an automated approach to image colorization. Various methodologies, such as color transfer [3] and exemplar-based colorization [4], were introduced in the initial stages of automated colorization. These approaches rely on reference colored images to guide the colorization process. The emergence of deep learning has facilitated the utilization of Convolutional Neural Networks (CNNs) [5] and Generative Adversarial Networks (GANs) [6], which learn to colorize images from extensive datasets without explicit instructions [7], [8].

Despite the achievements in image colorization using these techniques, challenges still exist without a comprehensive solution. The process of transitioning from grayscale

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

to color involves converting a single-channel image into a three-channel image using the RGB color model. The mapping process is likely to be non-deterministic, indicating that a grayscale image can be colorized in numerous ways. Human perception is vital for this process. The perception of color labeled as 'realistic' can vary among individuals [9]. Another significant obstacle to image colorization is color bleeding, which can be a severe problem because the assigned colors extend beyond the object borders, resulting in unnatural effects. Color bleeding mainly occurs when colors from one object bleed into nearby areas, with a loss of object differentiation. In addition, the phenomenon of ''hallucinated'' details in image colorization also remains a crucial problem. This phenomenon involves neural networks generating colors or intricate details that were not present in the original image. While such additions can enhance visual appeal, they may deviate from historical accuracy [10].

These issues have made it necessary to perform intensive research and develop novel methods for image colorization. Hence, we introduce a novel image colorization model that tackles the above mentioned issues and improves the overall quality and accuracy using the proposed Color Encoder and GAN architecture. The Color Encoder plays a crucial role in addressing the issue of undersaturation by successfully incorporating color features, resulting in a more realistic and vivid colorization result. However, the integration of the Convolutional Block Attention Module (CBAM) aims to mitigate color bleeding problems by improving the network's ability to focus on crucial image components, leading to more accurate and localized color deployment. Furthermore, the incorporation of Conditional Wasserstein Generative Adversarial Networks (CWGAN) increases the overall visual appeal by guiding the colorization process to align with natural color distributions and enhancing image finer details. These elements collectively empower our method to offer a comprehensive solution to the challenges encountered in image colorization, providing both accuracy and visual appeal. The key contributions of this paper are given below.

- Introduction of a novel Color Encoder, capable of learning and generating essential color features to tackle undersaturation issues and produce colorizations that closely resemble real-world colors.
- Integration of novel Color Loss to measure the disparity between color features generated by the Color Encoder and actual color features obtained from ground truth (GT). The Color Loss plays a crucial role in training the Color Encode.
- Incorporation of the CBAM within the architecture of the CWGAN. CBAM enhances the visual appeal of colorization results by focusing attention on important image features, contributing to more accurate and visually pleasing colorizations.

II. RELATED WORKS

The process of colorizing images extends back to the early twentieth century, when black-and-white photographs were hand-colored with water colors, oils, crayons, or other dyes [11]. Although the manual technique described above was characterized by its subjectivity and reliance on the artist's expertise and understanding, it nonetheless marked the early desire to infuse life into monochrome imagery.

With the introduction of digital image processing, semi-automated image colorization techniques have become available. Levin et al. [12] proposed a semiautomatic image colorization technique in which an interactive and optimization-based strategy with minimal user input was introduced. The user is required to provide a few color scribbles and the framework subsequently distributes these colors across the entire image by considering the similarities between the pixels. Luan et al. [13] introduced an optimization-based method for colorization by extending [12] and enhanced the propagation of colors by reducing user scribbles. Qu et al. [14] further improved the technique in [13] by integrating spatial and range regularization to manage color propagation effectively.

The emergence of machine learning has had a profound influence on image colorization. Ironi et al. introduced automated techniques for colorization in which colors are transferred from a reference image to a grayscale target image using feature similarities [3]. Welsh et al. also used color transfer but matched the source and target images using texture descriptors [4]. Particularly, deep-learning algorithms allow image colorization to be more automated, accurate, and consistent. Cheng et al. proposed an early deep-learning model that utilized a deep CNN as a regression task for colorization to predict the chrominance value of each pixel [15]. Zhang et al. made a notable contribution with an end-to-end model that employed a classification objective function rather than regression loss, leading to the production of more vivid and realistic colorizations [6]. Taking advantage of the nature of human color perception, this work used the CIELAB color space to anticipate the 'a' and 'b' channels from the input 'L' channel. The instance-aware image colorization method proposed by Su et al. focuses on achieving precise colorization using object instances [16].

GANs have also been investigated for image-colorization tasks. The Pix2Pix framework [17] employs a conditional GAN (cGAN) to acquire knowledge of the mapping between the input and output images. This framework has also been used effectively for colorization [17]. Moreover, there has been a surge in the popularity of open-source projects such as ''DeOldify'' that focus on adding color to historical images [18]. A further noteworthy contribution in the field is the research conducted by Vitoria et al., titled ''ChromaGAN.'' This study uses adversarial learning techniques to achieve colorization, with a specific emphasis on the distribution of semantic classes [19]. Furthermore, CBAM [20] is integrated into the ChromaGAN architecture to improve colorization [21]. Wang et al. [22] proposed DualGAN, a dual-path generative adversarial network for image colorization, which consists of two generator streams, one for low-frequency and one for high-frequency

colorization. The two streams are trained jointly with a discriminator to generate photorealistic colorizations. Subsequently, Kim et al. presented "BigColor," a technique that utilizes a generative color prior to enhancing natural image colorization [23]. Moreover, the authors propose a GAN-based image colorization method for self-supervised visual feature learning [38]. This GAN-based image colorization method is based on the cGAN architecture, with a new loss function, a multi-scale discriminator, and a channel and spatial attention mechanism. Furthermore, Liu and Tu [39] propose a PatchGAN-based image colorization model incorporating a CBAM. Liu et al. show that CBAM can help the model focus on important regions of the image, leading to improved performance on benchmark datasets.

Recently, the popularity of transformers [24] in vision-based applications has increased. Transformers are a class of neural networks that are designed to effectively process and manage long-range relationships. The Colorization Transformer, by Kumar et al. emerged as an early example of transformers employed in image colorization [25]. The grayscale image was coarsely colored using a colorization transformer composed of a conditional autoregressive transformer. Subsequently, two fully parallel networks are employed to upsample the coarse color, resulting in a finely colored high-resolution image.

In addition to colorization transformers, subsequent studies have employed transformers for image colorization. CT2 [26] is another image colorization technique that employs a transformer-based approach. In this method, colors are encoded as tokens, and the interaction between grayscale image patches and color tokens is guided by the color attention and query modules. Furthermore, DDColor [27] proposes a dual decoder GAN architecture for image colorization. The first decoder generates a coarse colorization, while the second decoder refines the colorization and adds semantic details.

Despite considerable advancements in image colorization, the existing techniques have several limitations. Manual and semiautomated techniques are associated with significant time consumption, require a certain level of artistic expertise, and may yield inconsistent outcomes [28]. Although powerful, deep-learning models are computationally expensive, require large amounts of data, and often produce desaturated outputs owing to the conservative nature of loss functions [8]. Furthermore, they struggle with multiple complicated images because of the difficulty in learning high-level semantics [29]. The issue of color ambiguity remains unresolved in most existing models. Models also tend to 'hallucinate' details, which can be problematic for historical and archival image colorization, where color authenticity is crucial [30].

To address these problems, we propose a novel image colorization method based on a CWGAN, which uses attention and Color Encoder in the generator. In our proposed method, we propose a novel colorization element, the Color Encoder, which significantly amplifies the colorization process by incorporating a comprehensive array of color features. Unlike conventional techniques, the proposed Color Encoder enhances the colorization process with a deeper understanding of the network with contextual color information, resulting in significantly improved outcomes. Moreover, our novel Color Loss mechanism plays a crucial role in training the Color Encoder to generate color features that closely resemble the original color features. The focus on the Color Encoder and the specific Color Loss significantly improves the effectiveness and quality of image colorization, setting our model apart from conventional approaches.

III. PROPOSED METHODOLOGY

The proposed image colorization model leverages the power of deep learning, specifically using GAN to achieve accurate and visually appealing colorization results. The proposed architecture consists of a generator network that learns to map the luminance (L) channel of the input grayscale image to the chrominance (a and b) channels in the CIELAB color space [31], and a discriminator network that learns to identify real images from the GT and the fake image generated by the generator. The CIELAB color space was chosen in this study because it closely approximates human vision, making it ideal for perceptually meaningful color transformations [31]. With the L^* component for luminance and the a^* and b^* components for color along the green-red and blue-yellow axes, respectively, CIELAB successfully isolates color information from intensity information. Moreover, the CIELAB color space provides perceptually uniform representations of colors, ensuring that the Euclidean distance between colors aligns more precisely with human perception of color distinctions.

Let x^L be the input grayscale image with a luminance channel (L) and x^{ab} be the GT image with chrominance channels (ab). x^L is the input to the generator, and the output chrominance image y^{ab} is generated using (1).

$$y^{ab} = G(x^L) \tag{1}$$

where G represents the generator function trained to produce an image containing the chrominance channels y^{ab}. These chrominance channels produce a colorized image, referred to as y^{Lab} when combined with the original input luminance channel, x^L. In other words. y^{Lab} is the final colorized image produced by the generator G, which combines the luminance information x^L with the chrominance channel y^{ab}. The output colorized image is characterized by its coherency and visual appeal. Subsequently, it is converted from the CIELAB color space to the RGB color space for the final output. Figure 1 shows the overall architecture of the proposed network. As shown in Figure 1, the luminance image is extracted from the input image and fed to the generator. The generator contains an encoder, a decoder, and a Color Encoder. The generator produced a y^{ab} image containing chrominance channels. The generated image is fed to the discriminator along with the input image as a condition. The discriminator generates scores based on real and false images.

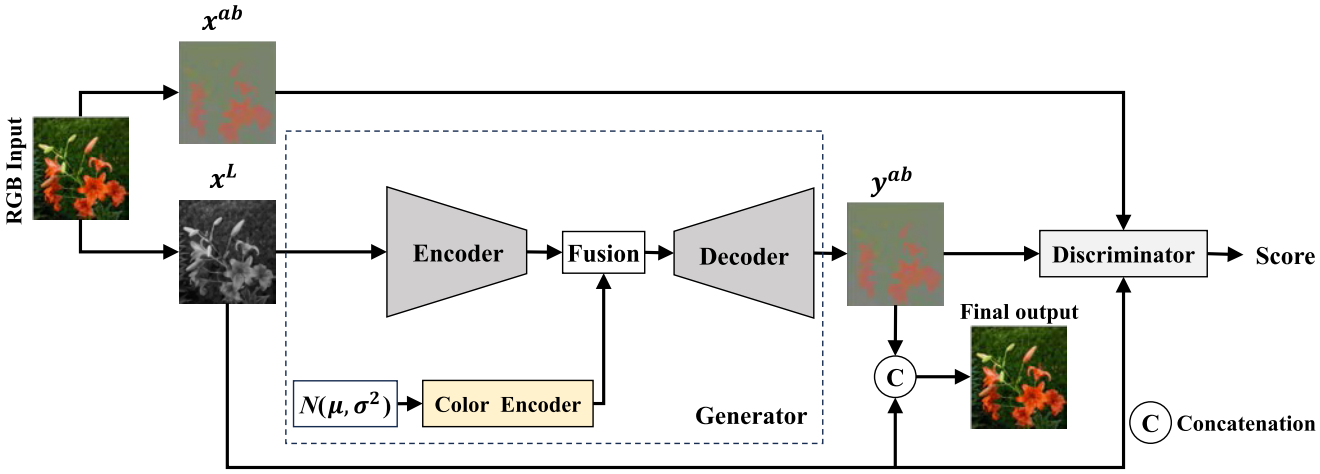

FIGURE 1. Overall architecture of the proposed colorization network.

Furthermore, a CWGAN architecture was utilized during training to enhance the colorization performance of the proposed architecture. The CWGAN expands the conventional GAN framework by integrating the Wasserstein distance as the training criterion in the cGAN [17], which can facilitate stable and interpretable gradients during training. The conditional approach enables the generator to learn colorization patterns specific to the input grayscale images, thereby ensuring that the colorization process is contextually relevant.

A. GENERATOR ARCHITECTURE

Figure 2 shows the architecture of the generator for the proposed image-colorization model. The generator is specifically designed to capture complex color relationships and accurately maintain spatial coherence. It comprises an encoder-decoder structure that effectively utilizes both content and color information. The encoder begins with the input grayscale image x^L and gradually reduces its spatial dimensions over the layers while simultaneously extracting increasingly detailed and intricate features from the image. The encoder consists of six convolutional blocks, each followed by batch normalization and ReLU activation. The convolutional block sequentially reduces the size of an image and increases the channels to 64, 128, 256 and 512, respectively. The decoder, a mirror of the encoder, upsamples low-resolution feature maps obtained from the encoder. Like the encoder, the decoder consists of six convolutional blocks, utilizing batch normalization and ReLU activation. The convolutional block sizes in reverse order are 512, 256, 128, and 64 channels, respectively. Moreover, as shown in Figure 2, CBAM [20] is used at each encoding stage, where the attention mechanism enhances feature representations by selectively focusing on salient regions while disregarding unnecessary information. The encoder outputs are subsequently transmitted to the decoder through skip connections as in U-Net [32], which retains fine-grained details and structural information to facilitate the colorization procedure.

Colorization was improved using a novel Color Encoder that creates color characteristics by sampling from a normal distribution, effectively modeling the color distribution. The Color Encoder consists of several convolutional layers that extract intrinsic color features from the input grayscale image. The convolutional block sizes in the Color Encoder correspond to 64, 128, 256, and 512 channels, sequentially, and it helps to enhance to capture and represent color-related information effectively. The color features of the Color Encoder are integrated with the output of the encoder at the network bottleneck. The integration of Color Encoder and encoder features ensures that the process of colorization integrates both the local context and global color information, allowing the model to generate colors that are coherent and realistic. The Color Encoder, a key component of our model, not only generates rich color features but also undergoes a crucial training process to ensure their accuracy. This training is facilitated by a specialized metric known as Color Loss. Color Loss serves as a guiding mechanism, quantifying the disparity between the features produced by the Color Encoder and those derived from GT chrominance images. Subsequently, the decoder network utilizes the fused features from the encoder and Color Encoder, and gradually upsamples the features to the original resolution. Skip connections from the encoder layers enable the decoder to access detailed spatial information and preserve fine textures in the colorized output. Finally, the output of the decoder is a chrominance image.

1) COLOR ENCODER

In our image colorization model, we propose a novel Color Encoder module that enhances colorization performance by producing detailed color characteristics. The Color Encoder module with a normally distributed input and convolution blocks is shown in Figure 2. The functionality of the Color Encoder is based on the integration of learned color features into the colorization process, which is a crucial step in attaining precise and realistic colorization.

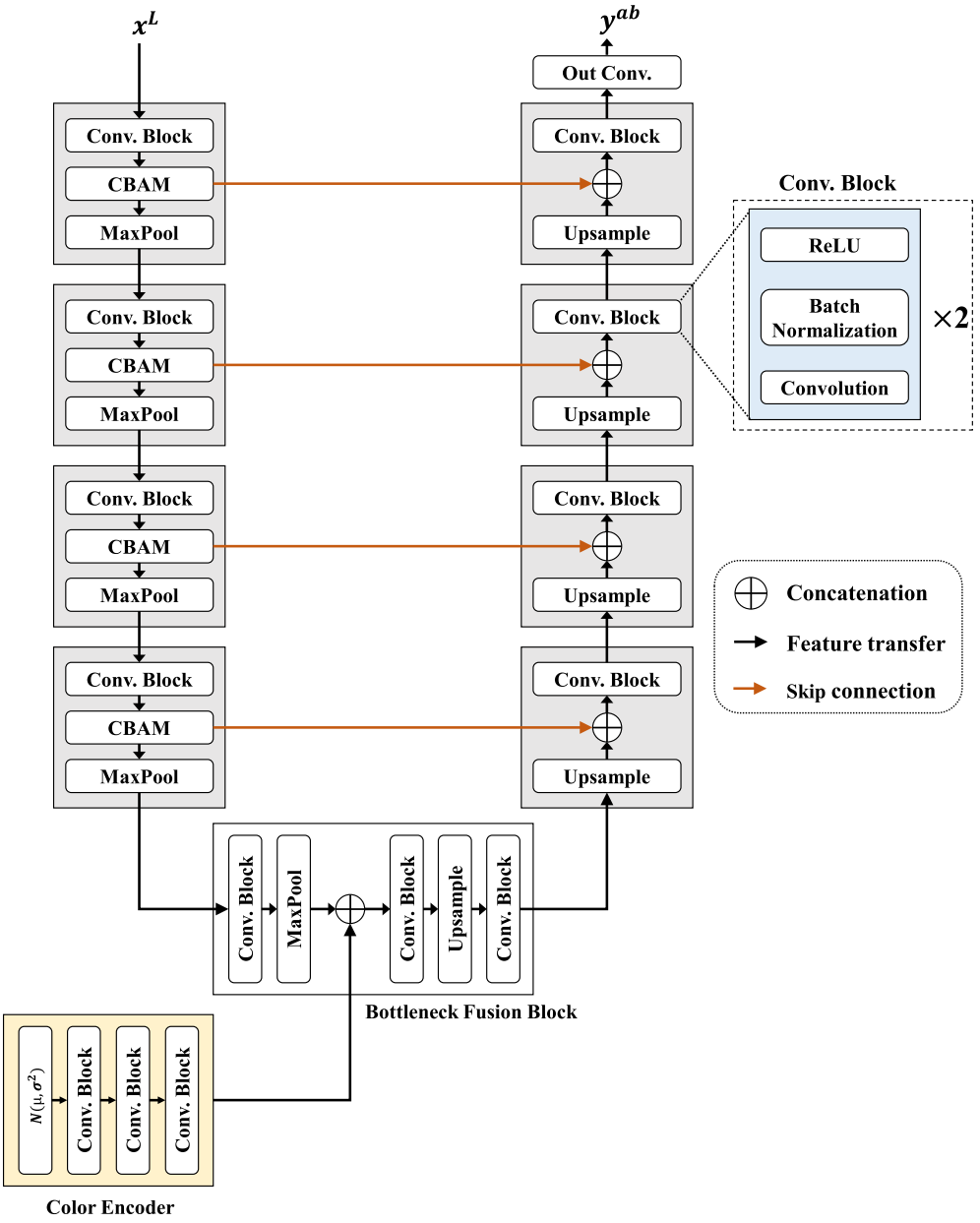

FIGURE 2. Proposed generator architecture.

Colorization is initiated using a random normal distribution that serves as a representation of the possible colors. The random normal distribution goes through a sequence of convolutional layers designed to extract and enhance the color features. The process of transitioning over these layers helps remove randomness because the initial state of randomness gradually becomes an organized color feature. An essential step is required to guarantee that these created features genuinely correlate with the real-world color features. The color features generated by the Color Encoder are compared with those obtained from a pretrained VGG network with GT color images, where a set of reference features that capture the essence of authentic color compositions is extracted. The primary goal of the comparison was to ensure that the features generated by the Color Encoder were in accordance with the reference features derived from the GT images. The similarity between two features is evaluated using a loss function that quantifies the degree of similarity between the features of the Color Encoder and the reference features. The Color Encoder can then generate features that resemble the color of the GT images by minimizing the loss. This approach converts a random normal distribution into color features that are not only semantically meaningful, but also inherently aligned with real color patterns. Moreover, the Color Encoder is trained jointly with the GAN. In the proposed colorization model, the Color Encoder plays a crucial role in generating colorized outputs that are visually appealing and realistic.

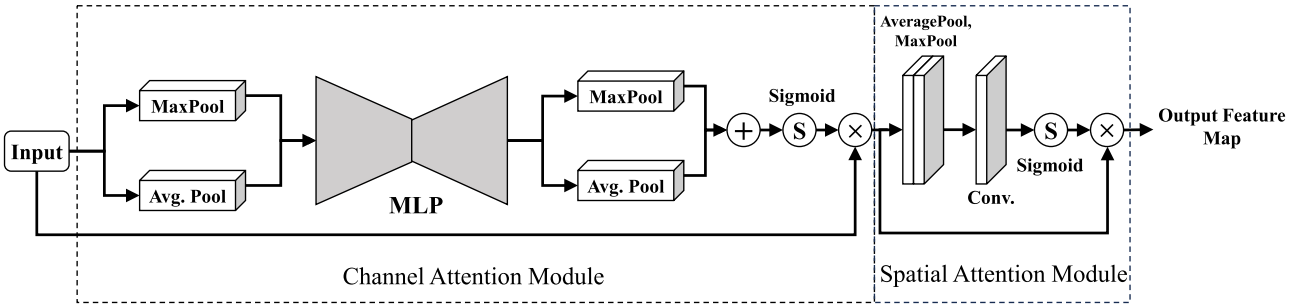

FIGURE 3. Convolution block attention module (CBAM) [20].

2) CBAM

In our colorization model, the CBAM was used to improve the ability of the model to recognize complex color patterns. The CBAM framework comprises two modules: Channel Attention Module (CAM) and Spatial Attention Module (SAM) [20].

The CAM operates on the channel dimension and involves a multi-layer perceptron (MLP) that learns to recalibrate channel-wise features. The MLP consists of linear transformations followed by non-linear activations (ReLU), focusing on channel-level details and aiding in better feature selection. The SAM, on the other hand, concentrates on spatial attention. It first compresses the input using a channel pooling mechanism. Subsequently, a basic convolutional layer processes the compressed data to generate a spatial attention map, which is then utilized to weigh the original features. Channel-wise attention helps identify significant patterns across various channels, emphasizing crucial image features. Simultaneously, spatial attention facilitates the model's focus on crucial spatial regions, thus ensuring the precise capture of significant details. These modules work together to selectively emphasize important color features, while also considering the spatial interactions between them. By integrating both channel-wise attention and spatial attention maps, CBAM ensures that the model directs its attention towards crucial features while maintaining the overall structure and consistency of the image. As a result, the CBAM plays a crucial role in enhancing the colorization process by guiding the attention of the model toward important image features.

B. DISCRIMINATOR

In our proposed colorization model, we adopted the PatchGAN [17] discriminator, which plays a crucial role in enhancing the realism of colorization. The discriminator of PatchGAN operates at a localized level with small patches within the images rather than in the entire image, improving spatial coherence and facilitating high-resolution colorization. In a practical implementation, the PatchGAN discriminator is trained to differentiate between two distinct categories: real patches extracted from the color images of the GT and colorized patches generated by the model. The number of filters starts with 64 in the initial convolutional layer. Subsequent convolutional blocks use the doubled number of filters. This progression increases the number of filters from 64 to 128 and subsequently from 128 to 256 in the following convolutional blocks. In addition, the stride is set to 2 in each convolutional block, effectively halving the spatial dimensions of the feature maps at each stride. This design allows for effective feature extraction while progressively reducing the spatial resolution in the network. The adversarial framework motivates the generator to generate colors that closely resemble actual images, whereas the discriminator enhances its ability to differentiate between real and fake outputs. Our model can enhance the realism and authenticity of colorized outputs using the PatchGAN discriminator.

C. OBJECTIVE FUNCTION

Our colorization model's ability to learn and get better is largely the result of the different types of "loss" we've used in the objective function. This combination includes adversarial, perceptual, and Color Loss, each of which contributes differently to improving the colorization process. The objective function of the proposed network is defined as

$$L_{total} = \lambda_g L_G + \lambda_1 L_{L1} + \lambda_p L_p + \lambda_c L_c \tag{2}$$

where L_{total} is the final loss, and L_G, L_{L1}, L_p *and* L_c are the adversarial, L1, perceptual, and Color Losses, respectively. λ_g, λ_p, λ_1, and λ_c in (2) are fixed and empirically set to $\{\lambda_g, \lambda_p, \lambda_1, \lambda_c\}=\{0.1, 1000, 100, 100\}$, respectively. The loss term L_G represents adversarial loss. The Wasserstein adversarial loss is used to resolve the vanishing gradient problem and enhance the colorization process, and is defined as

$$L_D = E_{x^{ab}}\left[D\left(x^{ab}, x^l\right)\right] - E_{y^{ab}}\left[D\left(y^{ab}, x^L\right)\right] + \lambda \times GP \tag{3}$$

$$L_G = -E_{y^{ab}}\left[D\left(y^{ab}, x^L\right)\right] \tag{4}$$

where L_D and L_G represent the discriminator and generator losses, respectively. x^{ab} and y^{ab} represent the real and generated chrominance images, respectively. GP is the gradient penalty (GP), x^L is the input grayscale image that is passed to the discriminator as a condition, E represents the expected

value or mean, and λ is a hyperparameter denoting the weight or coefficient for the *GP* term. In our context, the Wasserstein loss is a crucial part of our model's training, guiding the generator to create more realistic colorizations by minimizing the difference between the distribution of generated and real chrominance images.

In (2), L_c refers "Color Loss," which is a particular metric that measures difference between feature produced by the Color Encoder and one from a pretrained VGG network. The difference is calculated by L_1 distance. The Color Loss based on L_1 distance is distinct from the L_{L1} loss mentioned in the overall loss function of the model, which evaluates the model accuracy in replicating real colors. Here L_1 distance is only used to quantify the difference between actual color features (from pretrained VGG network) and generated ones (from Color Encoder). The Color Loss measures the alignment between the learned color features and representations obtained from the GT color images. Color Loss is defined by (5).

$$L_c = E\left|\left|G_f\left(N\left(\mu, \sigma^2\right)\right) - VGG\left(x^{ab}\right)\right|\right|_1 \tag{5}$$

where L_c is the Color Loss while $N(\mu, \sigma^2)$ is the random normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 0.1$ and G_f, VGG are the functions for Color Encoder through which random normal distribution is passed and pretrained VGG network, respectively. We use a two-step process, initially extracting features from the GT chrominance image using a pretrained VGG network, and then comparing these features with those generated by the Color Encoder, which receives a random normal distribution as input, ensuring that obtained color features closely resemble the GT.

L_{L1} is the conventional L1 loss defined as (6).

$$L_{L1} = \left|\left|x^{ab} - y^{ab}\right|\right|_1 \tag{6}$$

L1 loss measures the accuracy of the colorization model in replicating real colors. We also used perceptual loss to evaluate the perceptual similarity between the colorized and GT images. It leverages high-level features from a pretrained VGG network to ensure that the colorized output captures the overall structure, textures, and patterns of the real image. The VGG loss can be defined by the rectified linear unit activation layer of the pretrained VGG network, as in (7).

$$L_p = \left|\left|\varphi_k\left(x^{ab}\right) - \varphi_k\left(y^{ab}\right)\right|\right|_2^2 \tag{7}$$

k denotes the layer index with 0, 1, 2, and 3, signaling the layers from which features are extracted in the VGG network and φ_k represents the features of the k-th layer of the pretrained VGG network.

IV. EXPERIMENTAL RESULTS

A. IMPLEMENTATION DETAILS

The experiment was conducted using the PASCAL-VOC dataset [33], which consists of 17,125 publicly available images. The supplied images underwent a resizing process, in which they were transformed to a resolution of 256×256 pixels using bilinear interpolation. The experiment showed that scaling outperformed random cropping, owing to the potential adverse impact of cropping on color learning. The input images underwent normalization, resulting in their values being adjusted to fall within the range of -1 to 1. During the training process, the input images were split into L and ab channels, with the L channel serving as the input and the ab channel as the GT. We set 80% training, 10% validation, and 10% testing sets, respectively, which are taken from different image categories. The other methods were trained on the same training set for fair comparison. The epoch for training is set to 800, and the entire training takes approximately 2.5 days with GeForce RTX 3090 GPU.

The ADAM optimizer was used in our experiment with learning rates of 1×10.4 and 2×10.4 for the generator and discriminator, respectively. The values for the exponential decay rates β_1 and β_2 in the ADAM optimizer [34] were set to 0.5 and 0.999, respectively. The training process involved iteratively optimizing the generator and discriminator until the network converged.

B. RESULTS AND DISCUSSION

We used a variety of well-established assessment metrics including the Peak Signal-to-Noise Ratio (PSNR) [35], Structural Similarity Index (SSIM) [36], and colorfulness [37] to thoroughly analyze the performance of the proposed image colorization model. The Peak Signal-to-Noise Ratio (PSNR) is a commonly used metric in image colorization that measures the accuracy of colorized images by calculating the ratio between the maximum possible pixel value and mean-squared error between the generated chrominance image and the corresponding GT chrominance image. A higher PSNR signifies a higher degree of similarity in pixel values, indicating improved image quality. The Structural Similarity Index (SSIM) is a perceptual metric used to assess the degree of structural similarity between colored images and their corresponding GT. This assessment considers luminance, contrast, and structure, thereby providing a more human-centered evaluation of the visual fidelity. SSIM is a metric that quantifies the similarity between two images in the range of -1 to 1, where a value of 1 indicates a perfect match between the images. We computed PSNR and SSIM values with the resulting chrominance images (ab channel) and their corresponding GTs (ab channel). This approach evaluates the quality of colorization with a focus on the chrominance aspect, which is a key component of the color information, and provides a comprehensive assessment that fully incorporates color information when comparing the performance of different image colorization methods. The metrics of PSNR and SSIM are valuable in this context as they are well-established metrics for quantifying image quality, and they allow for a quantitative comparison of our method with existing techniques. PSNR and SSIM, designed for grayscale images, are commonly used metrics for assessing colorization models. Despite their original intent, they effectively capture key aspects of image quality, providing valuable insights into

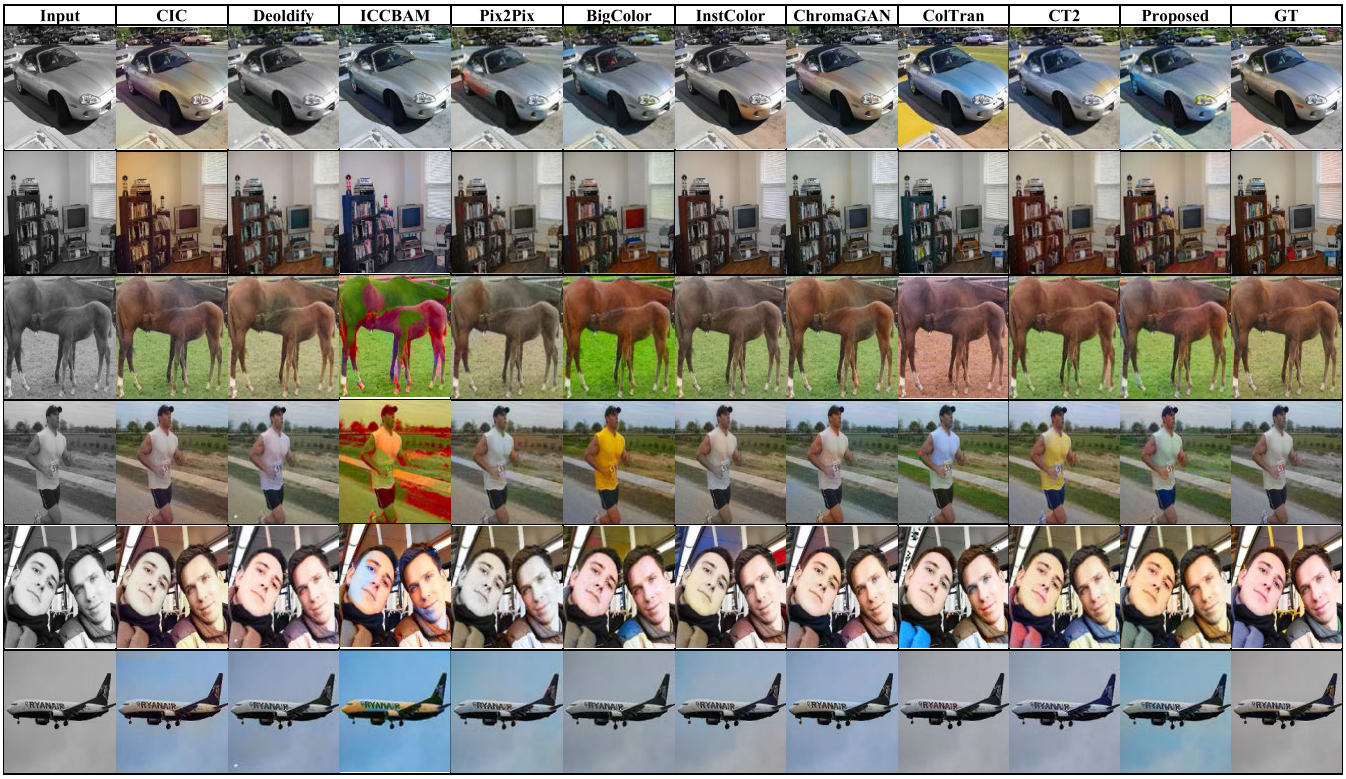

FIGURE 4. Visual results and comparisons.

the fidelity of colorized outputs. However, the colorfulness metric [37] is used to measure the color diversity and saturation of an image. A colorfulness metric [37] quantifies the level of colorfulness using colorized images and measures the degree of saturation and vividness exhibited by the colors present in the images. Higher values indicate that the colorized images demonstrate a greater abundance and intensity of color. The colorfulness of an image is computed by evaluating the standard deviation of its color channels. A higher standard deviation implies more color diversity and, hence, higher colorfulness.

The △Colorfulness metric is also proposed to alleviate the drawback of the colorfulness metric, which could favor vibrancy over realism and accuracy, and quantifies the difference in colorfulness values between the output and GT images. △Colorfulness is obtained by calculating the absolute difference in the colorfulness values between output and GT images and quantitatively measures how much the perceived colorfulness changes between these images. A lower △Colorfulness value indicates a higher level of color accuracy.

The performance of the proposed image colorization model was compared with that of a comprehensive set of seven state-of-the-art colorization models: CIC [8], Deoldify [18], Image colorization with CBAM (ICCBAM) [39], Pix2Pix [17], BigColor [22], InstColor [16], ChromaGAN [19], ColTran [25], and CT2 [26]. Table 1 presents a quantitative comparison between the proposed model and the other models. As shown in Table 1, our proposed method demonstrated significantly improved performance compared with other methods in terms of PSNR values, with higher similarity to the original color images. The capacity of the model to replicate the structural attributes of real images with perceptual accuracy was further highlighted using SSIM.

Desaturation, which causes muted or less vibrant colors, can be mitigated by integrating the proposed Color Encoder. The Color Encoder effectively incorporates the learned color features into the network at the bottleneck, enhancing the intensity and vibrancy of the generated colors. Furthermore, color bleeding, where colors can inadvertently spread beyond object boundaries, can be resolved by integrating the CBAM. The CBAM attention mechanism operates at the bottleneck of each layer in the encoder part, allowing the model to focus on local and global features. The selective attention of CBAM helps confine colorization to appropriate regions, significantly lowering the bleed effect with accurate and realistic colors. With the help of these strategic integrations, our model demonstrates improved color accuracy and realism while effectively reducing desaturation and color bleeding, as shown in Figure 4.

Our proposed method has a lower △Colorfulness value than existing state-of-the-art models, indicating that the outputs of the proposed method are more coherent to GTs. The obtained value indicates that our colorizations achieve a better balance between color intensity (vibrance) and realism.

TABLE 1. Quantitative comparison.

Models	PSNR (dB)	SSIM	Colorfulness	△Colorfulness
CIC[8]	21.000	0.925	30.43	2.55
Deoldify [18]	22.972	0.911	16.60	16.38
ICCBAM[39]	19.305	0.857	**46.65**	13.67
Pix2Pix[17]	23.886	0.932	18.66	14.38
BigColor [22]	21.473	0.883	35.71	2.73
InstCol [16]	22.911	0.910	22.21	10.77
ChromaGAN [19]	23.636	0.882	21.89	11.09
ColTran [25]	23.839	0.868	35.74	2.76
CT2 [26]	19.304	0.912	36.04	3.06
Proposed	**24.157**	**0.941**	30.92	**2.06**

FIGURE 5. Failure cases of the proposed method.

Although some models obtained very higher colorfulness scores with rare and vivid colors such as ICCBAM, the output images appeared to be overly saturated and unrealistic as shown in Figure 4. In contrast, our method achieves a lower △Colorfulness value, as shown in Table 1, indicating that our proposed model can produce outputs that closely resemble real-world GT image color perceptions. This result highlights that our model tends to make colorized images appear both realistic and visually appealing. This strikes a good balance between the two, so the colors look real, and the images are pleasing to the eye.

Figure 4 presents the visual results and comparisons of the proposed colorization model with other state-of-the-art methods. As shown in Figure 4, the outputs of our proposed network achieved more naturalness and realism, which is noticeable in the fourth row of Figure 4, where our method shows natural colors in numbers on a t-shirt. In contrast, the other models struggle with the task of colorization and do not achieve the same level of precision. BigColor shows oversaturation in the output color images, resulting in an exaggerated and unrealistic visual appearance. In comparison to the Pix2Pix architecture, which also employs a GAN-based approach and utilizes a U-Net as the generator, our proposed architecture shares similarities but notably demonstrates improvements in mitigating undersaturation as shown in Figure 4. CT2 also showed artifacts, such as bleeding artifacts. Moreover, CT2 and ColTran were oversaturated. In contrast, our proposed method achieves a delicate balance between vividness and accuracy in the output images. The output images from the proposed method have a higher level of realism and effectively preserve a vivid quality that is similar to the actual colors of the GT.

From Table 1 and Figure 4, it is evident that the proposed method demonstrates an exceptional performance in generating colorized images that exhibit a blend of realism and naturalness. The proposed method achieved superior performance compared with other techniques in precisely representing the fundamental characteristics of colors and effectively displaying both vivid and accurate colorizations.

C. ABLATION STUDIES

Ablation studies were performed to investigate the impact of two key elements (Color Encoder and CBAM modules) in our image colorization model. To evaluate the effect of the Color Encoder and CBAM modules, the model was trained and tested using two turning-on/off modules. Table 2 presents the results of these ablation studies.

The ‘‘Base’’ model, which turns off CBAM and the Color Encoder, shows relatively lower performance for multiple metrics. The model only with Color Encoder (‘‘A’’ model) by turning-off CBAM shows improvement in PSNR and SSIM with worse Colorfulness.

The model with only CBAM by the turning-off Color Encoder shows a similar improvement as case A. However, our proposed method surpasses these variants, indicating its capacity for superior image colorization quality.

This observation indicates that the inclusion of the CBAM has a notable impact on enhancing the attention mechanism of the model, leading to improved precision in the generated outcomes. In addition, the Color Encoder significantly enhances the realism and colorfulness of images by capturing

TABLE 2. Ablation studies.

Models	PSNR (dB)	SSIM	Colorfulness	△Colorfulness
Base (w/o CBAM and Color Encoder)	23.367	0.932	22.64	10.34
A (only with Color Encoder)	23.726	0.936	20.26	12.72
B (only with CBAM)	23.927	0.938	19.88	13.1
Proposed (Color Encoder + CBAM)	**24.157**	**0.941**	**30.92**	**2.06**

TABLE 3. Effect of the color encoder on the other colorization methods.

Models	PSNR (dB)	SSIM	Colorfulness	△Colorfulness
Pix2Pix[17]	23.886	0.932	18.66	14.38
Pix2Pix with Color Encoder	**23.919**	**0.939**	**26.10**	**6.88**
ChromaGAN [19]	23.636	0.882	21.89	11.09
ChromaGAN with Color Encoder	**23.805**	**0.936**	**25.52**	**7.46**

and incorporating essential color features. This performance improvement is attributed to the ability of the Color Encoder to extract and integrate crucial color information, which enriches the colorization process and results in more realistic and vibrant colorized images.

To investigate the efficacy of the Color Encoder, we performed experiments by adding the Color Encoder in the Pix2Pix and ChromaGAN frameworks. Table 3 shows the effect of the Color Encoder on other colorization methods. As shown in Table 3, the Color Encoder leads to improvements in terms of PSNR, SSIM, and Colorfulness when it is integrated into both methods. The improvement is more noticeable in Colorfulness and △Colorfulness for the methods, highlighting the effect of Color Encoder in creating more vibrant and visually appealing colorized images. These results indicate that the Color Encoder can improve the colorization performance even for the other method as well as our proposed method.

Despite the promising results, our proposed colorization model often shows failure cases in a specific condition. Figure 5 shows the examples. The proposed model shows color bleeding and incoherent color allocations with complex or cluttered scenes for the object boundaries and shows hallucinated details and unsaturated colorization for low-light images with faded contents or less context information. These failure cases underline the necessity for further enhancements, particularly in discerning complex scenes and refining the adaptation of colors in faded images.

V. CONCLUSION

This study introduces a novel image colorization model that utilizes a deep learning approach to generate accurate and realistic colorized images. The proposed model incorporates a proposed Color Encoder and CBAM module into a CWGAN architecture, yielding compelling results in both quantitative and qualitative evaluations. The effectiveness of the proposed model is demonstrated through experimentation and evaluation. The incorporation of Color Loss into the comprehensive objective function allowed the generation of colorizations that demonstrated both a strong resemblance to GT images and a perceptual appeal aligned with human visual perception. The significance of the proposed Color Encoder and CBAM module in producing precise color details and emphasizing important features was demonstrated through ablation studies. In conclusion, the proposed image colorization model outperformed state-of-the-art methods by producing colorized images that were both visually appealing and realistic.

REFERENCES

[1] A. R. Smith, "Color gamut transform pairs," in *Proc. 5th Annu. Conf. Comput. Graph. Interact. Techn.*, Aug. 1978, pp. 12–19.
[2] L. Yatziv and G. Sapiro, "Fast image and video colorization using chrominance blending," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1120–1129, May 2006.
[3] R. Irony, D. Cohen-Or, and D. Lischinski, "Colorization by example," in *Proc. Eurographics Symp. Rendering Techn.*, 2005, pp. 201–210.
[4] T. Welsh, M. Ashikhmin, and K. Mueller, "Transferring color to greyscale images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 277–280, Jul. 2002.
[5] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. NeurIPS*, 2014, pp. 2672–2680.
[7] G. Larsson, M. Maire, and G. Shakhnarovich, "Learning representations for automatic colorization," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 577–593.
[8] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 649–666.
[9] A. Deshpande, J. Rock, and D. Forsyth, "Learning large-scale automatic image colorization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 567–575.
[10] L. Chen, C. Doersch, A. Kornblith, G. Hinton, and T. D. Bui, "BigGAN: Large scale adversarial representation learning," in *Proc. NeurIPS*, 2018, pp. 10824–10839.
[11] J. Hill and A. Zakia, *The New Manual of Photography*. London, U.K.: Dorling Kindersley Limited, 2008.
[12] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 689–694, Aug. 2004.
[13] Q. Luan, F. Wen, D. Cohen-Or, L. Liang, Y.-Q. Xu, and H.-Y. Shum, "Natural image colorization," in *Proc. 18th Eurographics Conf. Rendering Techn.*, 2007, pp. 309–320.
[14] Y. Qu, T.-T. Wong, and P.-A. Heng, "Manga colorization," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 1214–1220, Jul. 2006.
[15] Z. Cheng, Q. Yang, and B. Sheng, "Deep colorization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 415–423.
[16] J.-W. Su, H.-K. Chu, and J.-B. Huang, "Instance-aware image colorization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7965–7974.

[17] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.
[18] *Deoldify*. Accessed: Aug. 2023. [Online]. Available: https://github.com/jantic/DeOldify
[19] P. Vitoria, L. Raad, and C. Ballester, "ChromaGAN: Adversarial picture colorization with semantic class distribution," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 2434–2443.
[20] S. Woo, J. Park, J. Lee, and M. Yang, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 427–436, doi: 10.1007/978-3-030-01234-2_29.
[21] Y. Lu, X. Huang, Y. Zhai, L. Yang, and Y. Wang, "ColorGAN: Automatic image colorization with GAN," in *Proc. IEEE 3rd Int. Conf. Inf. Technol., Big Data Artif. Intell. (ICIBA)*, vol. 3. Chongqing, China, May 2023, pp. 212–218, doi: 10.1109/ICIBA56860.2023.10164924.
[22] X. Wang, Y. Wu, Y. Li, H. Zhang, X. Zhao, and Y. Shan, "DualGAN: Dual-path generative adversarial network for image colorization," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2021, pp. 15679–15686.
[23] G. Kim, K. Kang, S. Kim, H. Lee, S. Kim, J. Kim, S.-H. Baek, and S. Cho, "BigColor: Colorization using a generative color prior for natural images," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2022, pp. 350–366.
[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. 31st Conf. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, 2017, pp. 6000–6010.
[25] M. Kumar, D. Weissenborn, and N. Kalchbrenner, "Colorization transformer," in International Conference on Learning Representations, 2021.
[26] S. Weng, J. Sun, Y. Li, S. Li, and B. Shi, "CT^2: Colorization transformer via color tokens," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2022, pp. 1–16.
[27] X. Kang, T. Yang, W. Ouyang, P. Ren, L. Li, and X. Xie, "DDColor: Towards photo-realistic and semantic-aware image colorization via dual decoders," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Dec. 2023, pp. 1–10.
[28] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.
[29] X. Liu, M. Gao, Y. Wang, and Q. Chen, "A local–global L_1-norm based variational model for image colorization," *Neurocomputing*, vol. 212, pp. 86–98, Sep. 2016.
[30] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 1857–1865.
[31] *Colorimetry—Part 1: CIE Standard Colorimetric Observers*, document CIE Publication 15.2, International Commission on Illumination (CIE), 2004.
[32] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent (MICCAI)*, 2015, pp. 234–241.
[33] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.
[34] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 1–13.
[35] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.
[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[37] D. Hasler and S. Susstrunk, "Measuring colourfulness in natural images," in *Proc. IS&T/SPIE Electron. Imag.*, vol. 12, Jun. 2003, pp. 87–95.
[38] S. Treneska, E. Zdravevski, I. M. Pires, P. Lameski, and S. Gievska, "GAN-based image colorization for self-supervised visual feature learning," *Sensors*, vol. 22, no. 4, p. 1599, Feb. 2022, doi: 10.3390/s22041599.
[39] C. Liu and Y. Tu. (2021). *Image Colorization With Convolution Block Attention Modules*. GitHub repository. [Online]. Available: https://github.com/kliu513/Image-Colorization

HAMZA SHAFIQ received the B.S. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2020. He is currently pursuing the M.S. degree with the Department of Information and Communications Engineering, Chosun University, South Korea. From 2020 to 2021, he was a Research Assistant with the University of Engineering and Technology, Lahore. He is a Graduate Research Assistant with the Multimedia Information Processing Laboratory, Chosun University. His research interests include image restoration and enhancement, image-to-image translation, and medical image processing.

BUMSHIK LEE (Member, IEEE) received the B.S. degree in electrical engineering from Korea University, Seoul, South Korea, in 2000, and the M.S. and Ph.D. degrees in information and communications engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2006 and 2012, respectively. He was a Research Professor with KAIST, in 2014, and a Postdoctoral Scholar with the University of California at San Diego, San Diego, CA, USA, from 2012 to 2013. He was a Principal Engineer with the Advanced Standard Research and Development Laboratory, LG Electronics, Seoul, from 2015 to 2016. In 2016, he joined the Department of Information and Communications Engineering, Chosun University, South Korea. His research interests include video processing, video security, and medical image processing.