CF-StyleGAN: Near-infrared image colorization of SE attention StyleGAN via color features

LINGJUN KONG
University of Shanghai for Science and Technology
XIN YANG
University of Shanghai for Science and Technology
WENJU WANG (✉ wangwenju@usst.edu.cn)
University of Shanghai for Science and Technology

Research Article

Keywords: Color histogram, Colorization of Near-infrared images, Generate adversarial networks (GANs), Attention mechanism

Posted Date: September 28th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3374527/v1

Additional Declarations: No competing interests reported.

CF-StyleGAN: Near-infrared image colorization of SE attention StyleGAN via color features

LINGJUN KONG,[1,2] XIN YANG,[1] AND WENJU WANG [1,*]

[1]College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai 200093, China
[2]Shanghai Publishing and Printing College, Shanghai 200093, China
*Corresponding author(s). E-mail(s): wangwenju@usst.edu.cn;
Contributing authors: klj@sppc.edu.cn; 212573066@st.usst.edu.cn.

Abstract

With current Near-infrared (NIR) image colorization methods, the color and details of the colorized images are not well restored. Thus, in this paper, we propose an unsupervised color feature control SE attention StyleGAN (CF-StyleGAN) method for the NIR image colorization task. The proposed method is based on histogram LAB color and brightness feature extraction, which solves the problem whereby the color and brightness of the results do not match the actual situation. The proposed Squeeze-and-Excitation-based StyleGAN (SE-SGAN) method, which introduces a channel attention mechanism based on StyleGAN and utilizes both standard deviation adaptive normalization and the Mish activation function in the synthesis network, can improve the quality of the output image. The proposed method was evaluated experimentally on the KAIST dataset. We found that the proposed CF-StyleGAN outperformed existing methods and achieved state-of-the-art NIR image colorization results. Experimental results show that the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) values of the colorized images were 27.15 and 0.83, respectively.

Keywords: Color histogram, Colorization of Near-infrared images, Generate adversarial networks (GANs), Attention mechanism

1 Introduction

Near-infrared (NIR) images are widely used in detection and monitoring systems, and many important night vision or low light scenarios, e.g., heat source detection in factories and forests, medical imaging, security checks, agricultural inspections and temperature measurement, and process control in industrial production, require the utilization of NIR imagery to realize comprehensive monitoring processes. However, NIR images are colorless. This lack of color information is not conducive to manual human observation, and large grayscale areas will prevent users from observing details in captured scenes. Thus, colorizing NIR images has important value for a variety of practical applications. However, currently, NIR image colorization methods suffer from several limitations, e.g., low colorization accuracy; thus, the NIR image colorization problem has become an important research topic.

Currently, deep learning (DL)-based [1] NIR image colorization methods can be categorized into supervised, semi-supervised, and unsupervised methods.

Supervised methods require labels to be added to the training set prior to training the corresponding models. Zhang et al. proposed a user-guided image colorization technique that utilizes a convolutional neural network (CNN) DL method [2]. Here, the user's choice is represented by fusing low-level cues and high-level semantic information obtained from large-scale data. However, with this method, the network must guide the user to make valid input choices, which tends to result in large errors in the colorization results. In addition, Kim et al. proposed the Generative Adversarial Networks (GAN)-based Tag2Pix art colorization method [3]. This method takes line art and color label information as the input to generate high-quality color images, which solves the problems associated with the user needing to select color information in Zhang's method [2]; however, the Tag2Pix art colorization method is limited to the colorization of a single line art images. Jheng-Wei Su et al. proposed a method to realize instance-aware coloring [4]. Here, the network structure utilizes a readymade object detector to acquire cropped object images and uses an instance coloring network to extract object-level features, which are then combined to produce good results. With this method, there is need to input grayscale lines and color labeling information as in the method proposed by Kim et al. [3]; however, the model returns to the full image coloring network when examples are not detected, thereby producing various defects, e.g., fading or color overflow. Typically, these supervised NIR image colorization methods require user-guided dataset labeling to realize image colorization. In addition, such methods also suffer from various problems, e.g., high subjectivity, lack of natural color, and human intervention.

Semi-supervised methods construct labeled and unlabeled data for training, which helps address the high subjectivity problem present in supervised methods. Deshpande et al. proposed a method that combines handmade edge features and residual network features [5] to generate color images. This method overcomes the problem of object color overflow due to a limited understanding of boundaries by automatic coloring. However, this method is ineffective when attempting to colorize images with multiple objects. He et al. proposed an end-to-end convolutional neural network C (CNNC) for exemplar-based local colorization [6]. This method effectively mitigates the problem of poor image colorization when handling multiple objects. In addition, it can select, spread, and predict colors in large-scale data, and it exhibits good robustness and generalizability when using reference images that are independent of the input grayscale image. However, with this method, the network cannot color objects with unusual colors. Lu et al. proposed the end-to-end Gray2ColorNet colorization network based on reference images [7]. This method learns the color information in the training data in addition to the semantic and global colors in the reference image, which effectively solves the problem of the CNNC method [6], i.e., special colors cannot be applied to objects. However, the Gray2ColorNet colorization network is only applicable to the colorization of low-resolution images, which is a significant limitation. While the results of this type of semi-supervised NIR image colorization method are dependent on the sample image, there are unique color problems that are difficult to handle in cases where the samples may lack similar colors.

Unsupervised methods do not require labeled datasets, which not only solves the problem of excessive manual intervention in the generated image results present in supervised method but also solves the problem where the results of the semi-supervised methods are overly dependent on the sample images, which can lead to large color differences in the image colorization results. Dong et al. proposed a nonreference method to color NIR images using SNet [8], and this method can effectively enhance the edges and stabilize the color regions. In addition, Suárez et al. proposed a method to colorize NIR images based on a superimposed conditional GAN [9]. Suárez's structure uses multiple loss functions in a conditional probabilistic generative model. However, the methods proposed by Dong et al. and Suárez et al. suffer from several problems, e.g., the training dataset is too small, the experimental results need to be optimized, and the color variety is insufficient. Sekiguchi et al. proposed a deep neural network for NIR image colorization [10] that utilizes a CNN based on encoder-decoder for colorization that has some predictive power, and Valsesia et al. proposed a neural network based on a graphical evolutionary layer [11] to solve the NIR image colorization problem. Both of these methods solve the problem of poor color diversity due to the lack of appropriate datasets; however, they generate blurred images and lack vivid colors. Yang et al. proposed a cycle-consistent generative adversarial network with cross-scale dense connection [12] to learn color transformations from the NIR domain to the RGB domain based on paired and unpaired data. This method improves the image resolution; however, the results obtained by this method exhibit significant differences between the generated images and the ground truth images. Afifi et al. proposed the HistoGAN [13] method to generate high-quality colorized images based on a color histogram-controlled GAN. Here, the color histogram provides an intuitive way to describe image colors while maintaining a connection to domain-specific semantics. However, this method is primarily only applicable to the colorization of color and grayscale images. Generally, unsupervised NIR image colorization methods do not require human interaction; however, they still suffer from chromatic aberrations from the real image, and the overall image quality is insufficient.

In summary, compared to supervised and semi-supervised methods, NIR image colorization by unsupervised methods can reduce the chromatic aberration problem due to artificial effects. However, existing unsupervised methods suffer from low colorization accuracy. Thus, this paper proposes a method to improve the accuracy of NIR image colorization. Our main contributions are summarized as follows.

(1) We propose an unsupervised color feature control SE attention StyleGAN (CF-StyleGAN) and apply it to the NIR image colorization task. The proposed CF-StyleGAN improves the accuracy and image quality of NIR image coloring. The network utilized in the proposed method comprises three main parts, i.e., histogram LAB color feature extraction and image brightness extraction, a U-Net encoder and decoder based on an attention mechanism, and the squeeze-and excitation-based StyleGAN (SE-SGAN).

(2) A histogram-based LAB color feature and brightness extraction method is proposed to solve the problem where the color and luminance of the colorization result that are not realistic. This method converts the RGB color space to CIELAB color space, which has a wide color range. This process effectively avoids the influence of cross-conversion between different devices and represents and conveys color information through two-dimensional (2D) histogram parameterization. In addition, NIR images lack luminance information; thus, this method can also generate luminance layers to provide luminance information for the reconstructed colorized images.

(3) The SE-SGAN method is proposed to improve the quality of the output images. This method is based on StyleGAN, and it implements a channel attention mechanism and utilizes standard bias adaptive normalization and the Mish activation function in the synthesis network. The addition of the channel attention mechanism and standard bias adaptive normalization makes the output images more vivid in detail, and the Mish activation function improves the network's robustness. This method can integrate the image information obtained in the first stage and train the colorization results of the NIR image.

(4) Extensive experiments were conducted on the KAIST dataset to demonstrate the efficiency of the proposed method. We found that the proposed CF-StyleGAN method outperformed SNet and Deep CNN. The results indicate that the proposed method improved the peak signal-to-noise ratio (PSNR) values by 17.48% and 4.54%, and the structural similarity (SSIM) values were improved by 7.79% and 16.90%, respectively.

2 Our method

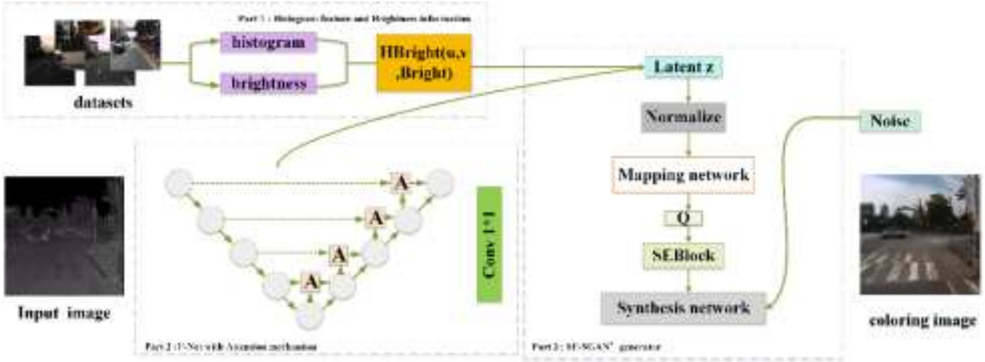


Fig. 1. Architecture of proposed CF-StyleGAN

In this paper, we propose the CF-StyleGAN. The architecture of the proposed network is shown in Figure 1. The network structure comprises three main components, i.e., (1) histogram-based color feature extraction and extraction of image brightness information, (2) an attention mechanism-based U-Net encoder and decoder, and (3) SE-SGAN-based image quality enhancement. Component 1 involves the extraction of the histogram LAB color and brightness information from RGB color image training dataset. The extracted information can be input to the generator of the SE-SGAN (Component 3). Component 2 involves inputting an NIR image to the U-Net encoder and decoder with an attention mechanism. Here, the encoder–decoder is employed to obtain and retain the fine local area features of the image to reduce loss of details. We refer to the U-Net network [14] and add the jump connection and the attention mechanism [15]. This step can accelerate the learning efficiency. Component 3 is an SE-SGAN trained together on the information output from Components 1 and 2 to generate a pair of color RGB images.

2.1 Histogram Lab color feature and brightness feature extraction

2.1.1 Histogram Lab color space features extraction

First, we convert RGB color images into LAB color space images. The LAB color space is a device-independent color system based on physiological characteristics. This means that the LAB color space can be used to describe human visual sensing and is unaffected by the interconversion between different devices. The LAB color space has a wide color field, including all of the color fields of the RGB color space, while also expressing colors that cannot be expressed by the RGB color space. As a result, the LAB color space is conducive to improving NIR image colorization accuracy. In reference to color constancy literature [16], we utilize histogram features to minimize the effect of brightness variations, and we construct the histogram as a differentiable histogram of colors in logarithmic chromaticity space [17]. Here, the feature is a 2D histogram of the image colors projected into the log-chromatic space. This 2D histogram is represented parametrically to convey the color information of the image, and it is more compact than a typical three-dimensional histogram defined in the RGB color space. The log-chromatic space is defined by the intensity of a single channel as normalized by the other two channels. As a result, three possible options can be utilized to construct three different histograms, which are then combined to form the histogram feature *H*.

The histogram LAB color space feature extraction involves the following three-step process.

(1) A RGB color image in dataset I_{RGB} is converted to the XYZ color space, which is denoted I_{XYZ}. This process is expressed as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \cdot \tag{1}$$

(2) The XYZ color space representation of the image I_{XYZ} information is converted to an LAB color space representation denoted. This process is expressed as follows:

$$\begin{cases} L^{*} = 116 f(\frac{Y}{Y_n}) - 16 \\ a^{*} = 500\left[f(\frac{X}{X_n}) - f(\frac{Y}{Y_n}) \right], f(\mathrm{t}) = \begin{cases} t^{1/3}, (\mathrm{t} > (\frac{6}{29})^3) \\ \frac{1}{3}(\frac{29}{6})^2 t + \frac{16}{116}, (\mathrm{others}) \end{cases} \\ b^{*} = 200\left[f(\frac{Y}{Y_n}) - f(\frac{Z}{Z_n}) \right] \end{cases} \cdot \tag{2}$$

(3) The image information represented by the LAB color space is transformed to a uv factors representation.

① First, the input image I_{Lab} is converted to the log-chromatic space, and then the histogram can be calculated. For example, we can select the *a* channel as the main color and normalize the *L* and *b* channels to obtain I_{ua}, I_{va}. With the *a* channel as the main color, I_{ua} and I_{va} expressed by uv factors are obtained as follows:

$$I_{ua}\left(x\right)=\log_2^{\left(\frac{I_a(x)+\varepsilon}{I_L(x)+\varepsilon}\right)},I_{va}\left(x\right)=\log_2^{\left(\frac{I_a(x)+\varepsilon}{I_b(x)+\varepsilon}\right)}, \tag{3}$$

where the *L*, *a*, and *b* subscripts denote the color channels of image I_{Lab}, *x* is the pixel index, (ua,va) is the uv coordinate based on using *a* as the primary channel, and ε is a small constant added to facilitate numerical stability.

② Then, the $Lab-uv$ histogram of the uv coefficient representation is acquired by thresholding the colors and calculating the contribution value of each pixel based on the intensity $I_y(\mathrm{x})=\sqrt{I_L^2\left(x\right)+I_a^2\left(x\right)+I_b^2\left(x\right)}$. To represent differentiability, a previous study [18] weighted the kernel of each threshold by the obtained contribution value, and the final unstandardized histogram was calculated as follows:

$$H(u,v,a)\propto\sum_x k(I_{ua}(x),I_{va}(x),u,v)I_y(x). \tag{4}$$

Here, $k(\cdot)$ is the inverse multiquadric kernel, which is defined as:

$$k(I_{ua},I_{va},u,v)=(\|I_{ua}-u\|^2+\tau^2)^{-\frac{1}{2}}\times(\|I_{va}-v\|^2+\tau^2)^{-\frac{1}{2}}. \tag{5}$$

Here, τ is the attenuation parameter used to control the smoothness of the histogram bin. Note that the other components I_{uL}, I_{vL}, I_{ub}, and I_{vb} are the same as in Equations (3), (4), and (5), and are calculated by projecting the *L* and *b* color channels into the logarithmic chromaticity space.

2.1.2 Image brightness extraction

For normal grayscale image colorization, using the input grayscale image as the brightness, only the chromaticity needs to be estimated; thus, the output image contains the original detail information and is not blurred by the detail information lost during the convolution of the neural network. However, the pixel values of NIR images depend on the reflection of the NIR light by the measurement material; therefore, rather than the brightness information observed by the human eye, they cannot be used as a brightness layer directly. This means that NIR image colorization requires the generation of both chromaticity and brightness layers. The luminance information $Bright$ extracted from the RGB color images I_{RGB} in the training set is given in Equation (6):

$$Bright=\sqrt{0.241\times R^2+0.691\times G^2+0.068\times B^2}, \tag{6}$$

where *R*, *G*, and *B* are the color information of the three channels of I_{RGB}.

The calculated brightness information can be combined with the image color histogram information obtained from the calculation Equation (4) in 2.1.1 calculation, Equation (7):

$$HBright(u, v, Bright) = concat[u, v, Bright]\,, \tag{7}$$

where u and v are the coefficients calculated from Equation (3), and $Bright$ is the brightness information calculated from Equation (6).

Finally, the histogram features are normalized to 1, i.e., $\sum_{u,v,Bright} HBright(u, v, Bright) = 1$.

2.2 SE-SGAN

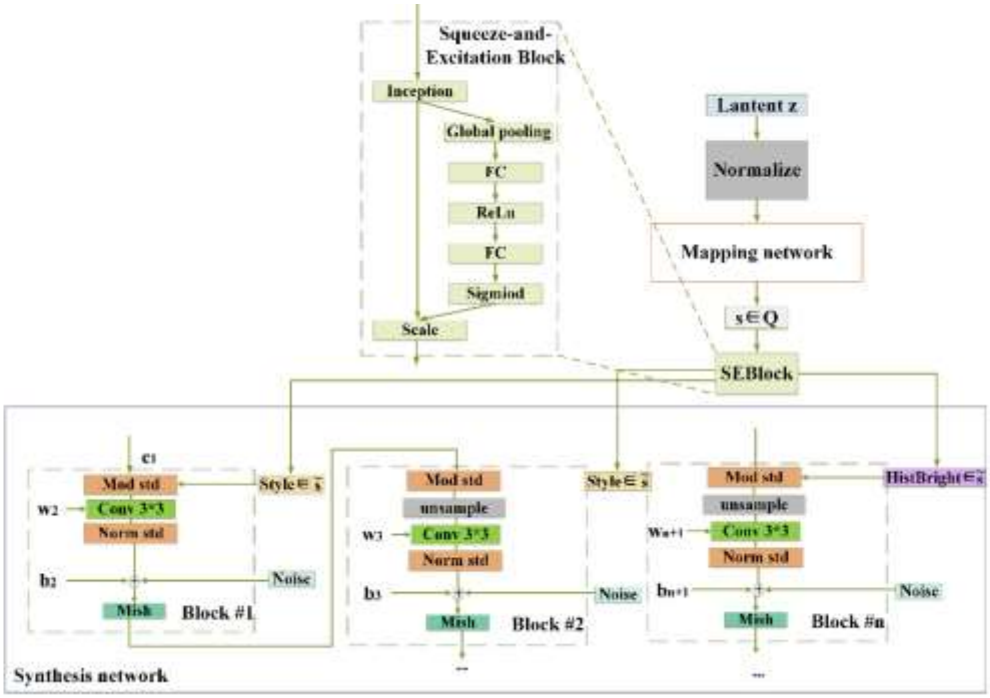


Fig. 2. Structure of SE-SGAN generator network

Here, we propose the SE-SGAN for NIR image colorization. The generator structure of the proposed method, including the noise (Noise), learned weight (w), bias (b), constant input (c), and number of blocks (n), is shown in Figure 2. The StyleGAN network comprises a mapping network and synthetic network [19]. The mapping network comprises eight fully-connected layers, and the synthetic network middle potential space controls the generator in each convolutional layer via adaptive instance normalization (AdaIN). Note that Gaussian noise is added after each convolution before evaluating the nonlinearity. In addition, the StyleGAN backbone network can control the generated images by style mixing, i.e., feeding different potential w to different layers during the inference process. Thus, we pass the style features (style), color histogram features, and brightness information (HistBright) of the target image into the synthesis network. However, the style modulation process can amplify some feature maps by an order of magnitude or more. For effective style modulation, we must remove as much of this amplification as possible on a per-sample basis; otherwise, subsequent layers will operate on the data in a meaningless manner. Thus, we use standard deviation adaptive normalization (SdaN) to improve and introduce the channel attention mechanism between the mapping and synthesis networks, and we replace the original activation function, i.e., Leaky ReLU, with the Mish activation function. The proposed SE-SGAN can help the overall model improve the quality and accuracy of the generated images.

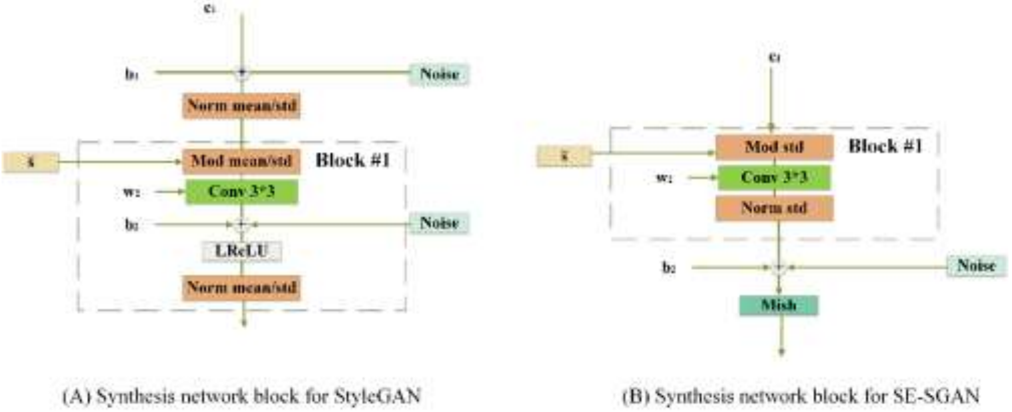
The images are processed by the SE-SGAN network according to the following three steps.

(1) Composition of the input code z. The network transmits the latent code z through the input layer. Here, z comprises two parts. (1) To ensure that fine details are preserved in the image after colorization, the early latent features generated by the first two U-Net blocks are connected via skip as input. (2) The target color information and luminance information (Hb) extracted via Equation (7) are also passed to the mapping network of the proposed SE-SGAN. z is mapped to the intermediate potential space Q by normalization and a mapping network comprising eight fully-connected layers to obtain s($s \in Q$) [19].

(2) The SE attention mechanism can inscribe image details. This module comprises a squeeze block (squeeze) and excitation block (excitation) [20], and it is an adaptive attention mechanism that recalibrates the channel feature responses adaptively by explicitly modeling the interdependencies between the channels. The style, color, and brightness styles of the mapping network output are manipulated by the introduced squeeze and stimulus blocks. The squeeze block collects the global spatial information through global average pooling, and the stimulus block captures channel relationships and outputs an attention vector using the fully-connected and nonlinear layers (ReLU and sigmoid). Each input feature is then scaled by multiplying the corresponding elements of the attention vector by the channels. The compression and excitation block F_{se} (parameter θ) with $\tilde{s}$ as the output are expressed as follows:

$$\tilde{s} = F_{se}(s, \theta) \cdot s \, , \tag{8}$$

where $(\text{Style}, \text{HistBright}) \in \tilde{s}$ and the SE module suppresses noise while emphasizing important channels.



(A) Synthesis network block for StyleGAN

(B) Synthesis network block for SE-SGAN

Fig. 3. Redesigned synthetic network block

(3) SdaN enhances the image quality. Step (2) generated is input to the synthesis network. The adaptive normalization [21] involves normalizing the mean and standard deviation, as well as adjusting the mean and standard deviation, as shown in Figure 3(A). However, the relative effect between the deviation and noise is inversely proportional to the effect of the image style; thus, the normalization and adjustment of the mean is not required. In this paper, only the standard deviation needs to be normalized and adjusted, as shown in Figure 3(B).

We assume that the input activation is a random variable with unit standard deviation, and after modulation and convolution, the standard deviation of the output activation is expressed as follows:

$$\sigma_j = \sqrt{\sum_{i,k} {w'_{ijk}}^2} \,. \tag{9}$$

In other words, the output is scaled by the L_2 parametric of the corresponding weights. The subsequent normalization process is performed to restore the output to unit standard deviation based on Equation (9). Here, if we scale each output feature map *j* by $\frac{1}{\sigma_j}$, the effect of *d* can essentially be removed. Note that we place the entire style block into a single convolutional layer whose weights are scaled according to the input mapping scale.

In addition, style modulation is achieved by scaling the convolution weights to modulate each input feature mapping based on the incoming style scaling convolution, as shown in Equation (10):

$$w'_{ijk} = d_i \cdot w_{ijk} \,, \tag{10}$$

where w and w' are the original and modulation weights, respectively, d_i is the scale corresponding to the i-th input feature map, and j and k are the spatial quantities of the output feature map and the convolution, respectively. We use standard deviation instance normalization, which can essentially remove the effect of *d* from the statistics of the convolutional output feature map.

The Mish activation function [22], which is improved from the leaky ReLU activation function, is used in the proposed method. The leaky ReLU activation function is used in existing StyleGAN networks; however, it suffers from an inability to hold negative values when used. In contrast, the Mish function is nonmonotonic, which allows the function to retain a small negative value and stabilize the network gradient flow. As a result, most neurons cannot be updated effectively. The Mish activation function calculates the following equation:

$$f(\tilde{c}_i) = \tilde{c}_i \tanh(\ln(1 + e^{\tilde{c}_i})) \cdot \tag{11}$$

Based on the StyleGAN architecture, the proposed SE-SGAN method can control the color of the generated images using the color histogram features and luminance features. As shown in Figure 1, the output of the last two blocks is adjusted to represent the information conveyed by the target histogram and brightness, thereby resulting in a colorized NIR image.

3 Experimental results and analysis

3.1 Experimental configuration and parameter selection

In our experiments, the training parameters were set as follows. The input image size was 256×256, the batch size was set to 2, the number of epochs was set to 150,000, and the learning rate was set to $2e^{-4}$. The experimental configuration is shown in Table 1.

Table 1. Experimental software and hardware configuration

CPU	Intel Core i9-9900K
RAM	32G
GPU	NVIDIA GTX 1080 Ti
Operating system	Windows 10
Deep learning framework	Pytorch

3.2 Experimental dataset

The proposed network was validated on the publicly available KAIST [23] dataset, which contains Korean campus, road, and city street scenes captured under both day and night conditions. Each image in the KAIST dataset consists of a visible image and its corresponding long-wave infrared image. It was captured simultaneously by a global shutter color camera (PointGrey Flea3) and a thermal camera (FLIR-A35). The color camera has a spatial resolution of 640 × 480 pixels, and the thermal camera has a spatial resolution of 320 × 256 pixels. The color camera has a larger field of view than the thermal camera; thus, both the NIR and RGB images are cropped during preprocessing, and the final image size is a 256 × 256 pixel image pair. In this experiment, the training set contained 50,172 pairs of visible and LWIR paired images acquired in all weather conditions. The test set contained 2,252 pairs of visible and LWIR paired images, of which 1,455 pairs were acquired during the day, and the remaining 797 pairs were acquired at night.

3.3 Evaluation metrics

In this evaluation, the SSIM and the PSNR were used to evaluate the quality of the generated RGB images.

The PSNR is used to evaluate the quality of two images compared to each other, i.e., the images are evaluated in terms of their distortion. A higher PSNR value indicates a lower level of image distortion. The PSNR [24] is calculated as follows:

$$PSNR = 10 \times \log_{10}(\frac{MAX_I^2}{MSE}), \tag{12}$$

where MAX is the maximum pixel value of the image.

The SSIM measures the structural similarity between two images near ground truth image x and the generated RGB color image y in terms of three aspects, i.e., brightness, contrast, and structure. The SSIM [25] is calculated as follows:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x{}^2 + \mu_y{}^2 + C_1)(\sigma_x{}^2 + \sigma_y{}^2 + C_2)}, \tag{13}$$

where μ_x and μ_y are the mean values of x and y for the brightness measurement, σ_x and σ_y are the standard deviations of x and y for the contrast measurement, respectively, σ_{xy} is the covariance of x and y for the structure measurement, and C is a positive constant to prevent division by 0 anomalies.

3.4 Performance Comparison and Analysis

3.4.1 Comparison and analysis based on KAIST dataset

We randomly selected several sets of corresponding NIR and ground truth color images in the KAIST dataset, as well as colorized images obtained by algorithmic models of SNet [8] and Deep CNN [26]. The selected images were compared with the colorized images obtained by our model and used to verify that our model outperforms other methods for colorizing NIR images.

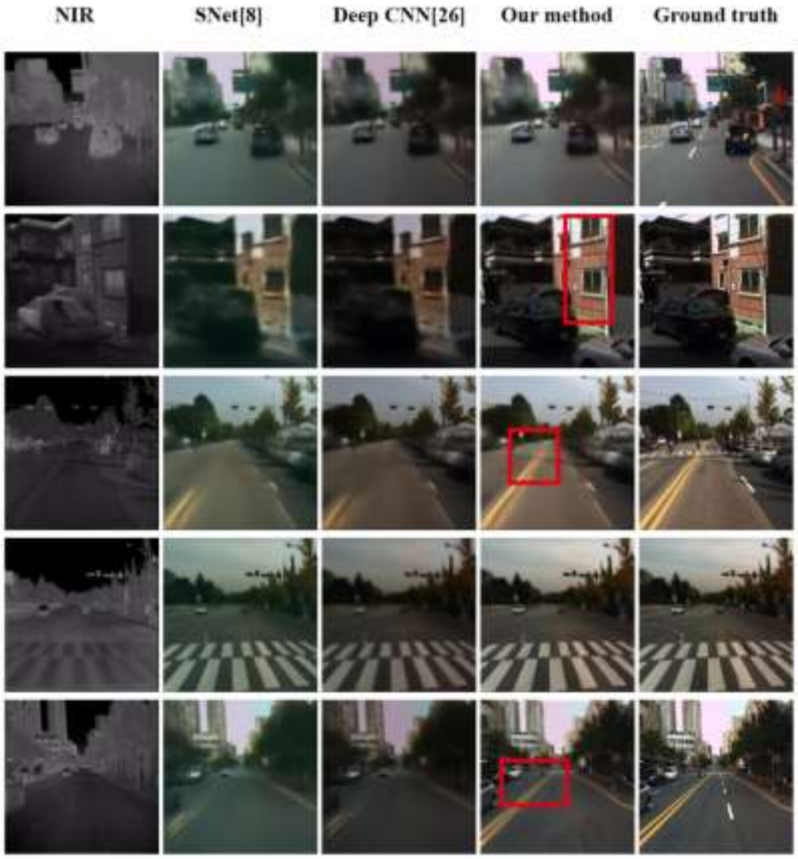


Fig. 4. The results of NIR image color visualization SNet, DeepCNN and our method are compared.

The image colorization results of five sets of randomly selected NIR images in the KAIST dataset were compared. We found that the colorization results obtained by the proposed method are more similar to the ground truth in terms of both color and detail (Figure 4). In particular, the area marked by the red box in the figure 4 is better reproduced by the proposed method compared with the SNet and Deep CNN algorithms, the edge details of the buildings are well reproduced, and the overall image color and brightness are closest to the ground truth images. In addition, the colorized images obtained by the proposed method are better than those acquired by the compared networks due to the added color and brightness features.

Table 2. Performance comparison of SNet, DeepCNN, and the proposed method on the KAIST dataset

Method	PSNR/dB↑	SSIM/dB↑
SNet[8]	23.11	0.77
Deep CNN[26]	25.97	0.71
Our method	27.15	0.83

The experimental results are shown in Table 2. As can be seen, the PSNR value of the proposed method is 27.15, and the SSIM value is 0.83. In contrast, the PSNR and SSIM values of the SNet algorithm are 23.11 and 0.77, and those of the Deep CNN algorithm are 25.97 and 0.71, respectively. Thus, the proposed method outperformed the compared algorithms. These results demonstrate that the quality of the colorized images obtained by the proposed method is closer to the ground truth image. Compared with the SNet algorithm [8], the PSNR value of the proposed method is improved by 17.48%, and the SSIM value is improved by 7.79%. Compared with the Deep CNN algorithm [26], the PSNR value of the proposed method is improved by 4.54%, and the SSIM value is improved by 16.90%. Thus, the proposed method exhibits state-of-the-art NIR image colorization accuracy because the proposed method learns and extracts color and brightness information by training on RGB color images. The encoding and decoding processing of the NIR images based on the attention mechanism also emphasizes the significant feature of passing through skip connections. Finally, the acquired color and luminance information, as well as the processed NIR images, are further processed by the proposed SE-SGAN, which improves the image quality while restoring the color and luminance of the images.

3.4.2 Ablation experiments

Table 3. Results of ablation experiments

RGB-uv	Lab-uv	SE	AdaIN	SdaN	ReLU	Mish	PSNR↑	SSIM↑
√		√		√		√	26.84	0.79
	√	√		√		√	27.14	0.81
	√			√		√	26.83	0.81
	√	√	√			√	27.09	0.77
	√	√		√	√		26.90	0.82
	√	√		√		√	27.15	0.83

Ablation experiments were conducted to compare the image color feature extraction methods based on the RGB and LAB color spaces. The results are shown in Table 3, rows 1 and 5. As can be seen, the PSNR value of the image color information extracted by the Lab-uv method is 27.15 and the SSIM value is 0.83, which is 1.15% higher than the PSNR and 5.06% higher than the SSIM of the color histogram information extracted using the RGB-uv method. A comparison of the results obtained with and without the SE attention mechanism in the proposed SE-SGAN is also shown in Table 3, rows 2 and 5. As shown, the PSNR and SSIM values of the image generated without SE are 26.83 and 0.81, respectively. When SE was used, the PSNR value is 1.19% higher, and the SSIM value is 2.41% higher. Rows 3 and 5 in Table 3, show the results of method-to-table experiments for adaptive normalization. As can be seen, the PSNR value when using AdaIN is 27.09, and the SSIM value is 0.77. When using SdaN, the PSNR value is improved by 0.22%, and the SSIM value is improved by 6.49%. The comparative tests of the activation function selection are shown in rows 4 and 5 in Table 3. Here, when using the ReLU activation function, the PSNR and SSIM values are 26.90 and 0.82, respectively. When using the Mish activation function, these values are improved by 0.93% and 1.22%, respectively. We found that when using the Lab-uv method to extract the color information of the images, the inclusion of both the attention mechanism and the Mish activation function improved the color and quality of the generated images. They are embodied in the results of PSNR and SSIM.

The results of the ablation experiments verified that all components implemented in the proposed CF-StyleGAN help realize improved image colorization.

4 Conclusion

In this paper, we have proposed the CF-StyleGAN, which is an unsupervised method for NIR image colorization that realizes color feature control based on the SE attention mechanism of the conventional StyleGAN. To solve the problems related to color difference and low image quality in unsupervised NIR image colorization methods, the proposed model is based on histogram LAB color features and luminance extraction, which solves the problem whereby the color and luminance of the model's colorization results do not match with the actual situation. In addition, the proposed SE-SGAN method, which introduces the channel attention mechanism based on StyleGAN and adds SdaN and the Mish activation function to the synthesis network, improves the quality of the output image. Compared with several advanced NIR image colorization methods, the proposed CF-StyleGAN demonstrates superior results. However, similar to other GAN networks, the proposed model requires a large number of samples and high training time. Thus, in future work, we plan to introduce other techniques, e.g., fewer training samples and lightweight models, to further improve and optimize the proposed CF-StyleGAN to make it more widely applicable in other fields, e.g., image generation.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical Approval Not applicable.

Funding Not applicable.

Data availability The datasets analysed during the current study are available in the KAIST, http://rcv.kaist.ac.kr/multispectral-pedestrian/

References

1. S. Anwar, M. Tahir, C. Li, A. Mian, F. S. Khan, and A. W. J. a. p. a. Muzaffar, "Image colorization: A survey and dataset," (2020).
2. R. Zhang, J.-Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, and A. A. Efros, "Real-time user-guided image colorization with learned deep priors," ACM Transactions on Graphics **36**, 1-11 (2017).
3. H. Kim, H. Y. Jhoo, E. Park, and S. Yoo, "Tag2Pix: Line Art Colorization Using Text Tag With SECat and Changing Loss," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV),* (2019), pp. 9055-9064.
4. J. W. Su, H. K. Chu, and J. B. Huang, "Instance-aware Image Colorization," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020),
5. S. C. Deshpande, M. M. Pawer, D. V. Atkale, and D. M. Yadav, "Fusion of handcrafted edge and residual learning features for image colorization," Signal, Image and Video Processing **16**, 291-299 (2021).
6. M. He, D. Chen, J. Liao, P. V. Sander, and L. Yuan, "Deep exemplar-based colorization," ACM Transactions on Graphics **37**, 1-16 (2018).
7. P. Lu, J. Yu, X. Peng, Z. Zhao, and X. Wang, "Gray2ColorNet: Transfer More Colors from Reference Image," in *Proceedings of the 28th ACM International Conference on Multimedia,* (2020), pp. 3210-3218.
8. Z. Dong, S. I. Kamata, and T. P. Breckon, "Infrared Image Colorization Using a S-Shape Network," (2018).
9. P. L. Suarez, A. D. Sappa, B. X. Vintimilla, and R. I. Hammoud, "Near InfraRed Imagery Colorization," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018),
10. S. Sekiguchi and M. Yamamoto, "Near-Infrared Image Colorization by Convolutional Neural Network with Perceptual Loss," in *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, 2020),
11. D. Valsesia, G. Fracastoro, and E. Magli, "NIR image colorization with graph-convolutional neural networks," in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 2020),
12. Z. Yang and Z. Chen, "Learning From Paired and Unpaired Data: Alternately Trained CycleGAN for Near Infrared Image Colorization," in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 2020),
13. M. Afifi, M. A. Brubaker, and M. S. Brown, "HistoGAN: Controlling Colors of GAN-Generated and Real Images via Color Histograms," in *Computer Vision and Pattern Recognition*, 2021),
14. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *18th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015, October 5, 2015 - October 9, 2015*, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Springer Verlag, 2015), 234-241.
15. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. Mcdonagh, N. Y. Hammerla, and B. Kainz, "Attention U-Net: Learning Where to Look for the Pancreas," (2018).
16. M. Afifi, B. Price, S. Cohen, and M. S. Brown, "When color constancy goes wrong: Correcting improperly white-balanced images," in *32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019, June 16, 2019 - June 20, 2019*, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (IEEE Computer Society, 2019), 1535-1544.
17. E. Eibenberger and E. Angelopoulou, "The importance of the normalizing channel in log-chromaticity space," in *2012 19th IEEE International Conference on Image Processing, ICIP 2012, September 30, 2012 - October 3, 2012*, Proceedings - International Conference on Image Processing, ICIP (IEEE Computer Society, 2012), 825-828.
18. M. Afifi and M. S. Brown, "Sensor-independent illumination estimation for DNN models," in *30th British Machine Vision Conference, BMVC 2019, September 9, 2019 - September 12, 2019*, 30th British Machine Vision Conference 2019, BMVC 2019 (BMVA Press, 2020), Amazon; Apple; et al.; facebook; Intel; Microsoft.

19. T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019),
20. J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018),
21. T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, June 14, 2020 - June 19, 2020*, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (IEEE Computer Society, 2020), 8107-8116.
22. D. Misra, "Mish: A Self Regularized Non-Monotonic Neural Activation Function," (2019).
23. S. Hwang, J. Park, N. Kim, Y. Choi, and I. S. Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, June 7, 2015 - June 12, 2015*, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (IEEE Computer Society, 2015), 1037-1045.
24. Q. Huynh-Thu and M. J. E. L. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," **44**, 800-801 (2008).
25. W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. J. I. T. I. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," **13**(2004).
26. M. Limmer and H. Lensch, "Infrared Colorization Using Deep Convolutional Neural Networks," (2016).