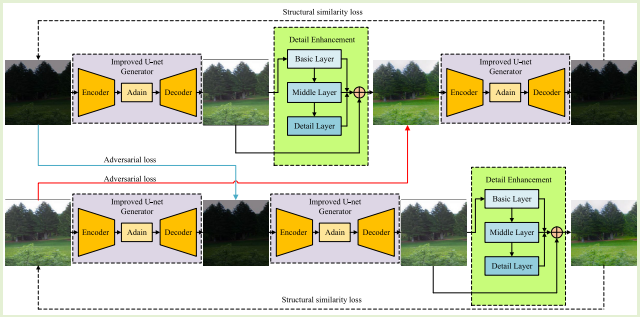
An Improved CycleGAN-Based Model for Low-Light Image Enhancement

Guangyi Tang, Jianjun Ni, *Senior Member, IEEE*, Yan Chen, Weidong Cao, *Member, IEEE*, and Simon X. Yang, *Senior Member, IEEE*

***Abstract*—The low-light image enhancement is a challenging and hot research issue in the image processing field. In order to enhance the quality of low-light images to obtain full structure and details, many low-light image enhancement algorithms have been proposed and deep learning-based methods have achieved great success in this field. However, most of the deep learning methods require paired training data, which is difficult to obtain. And the overall visual quality of the enhanced image is still not very satisfying. To deal with these problems, an unsupervised low-light image enhancement model based on an improved cycle-consistent generative adversarial network (CycleGAN) is proposed in this article. In the proposed model, a low-light enhancement generator of the CycleGAN network is constructed based on an improved U-Net structure, and the adaptive instance normalization (AdaIN) is designed to learn the style of the normal light image. In particular, a detail enhancement method based on multilayer guided filtering is added to the proposed model, which can improve the quality and visual pleasantness of image enhancement. In addition, a joint training strategy based on structural similarity is presented, to strengthen the constraints on generating more realistic and natural images. At last, extensive experiments are conducted and the results show that the proposed method can accomplish the task of transferring low-light images to normal light and outperform the state-of-the-art approaches in various metrics of visual quality.**



***Index Terms*—Cycle-consistent generative adversarial network (CycleGAN), image enhancement, low-light problem, unsupervised learning.**

I. Introduction

THE vast majority of information that humans acquire comes from vision. As the main carrier of visual information, image plays an important role in semantic segmentation, 3-D reconstruction, automatic driving, target detection, and so on [1], [2], [3], [4]. With the rapid development of optics and computer technology, image acquisition equipment is constantly updated. There is a lot of valuable information in the image waiting to be discovered and acquired by human beings [5], [6], [7], [8]. However, in the process of visible light imaging, the ambient light intensity often affects the quality of the image. The image will be low-light when the ambient light is low, such as at night or in a dark room. This kind of image has some disadvantages, such as low contrast, low brightness, lack of availability, and detail obscured [9], [10]. It will bring great difficulties to the following tasks such as target detection, image recognition, and segmentation. Moreover, the photoreceptor cells on the retina are less sensitive to color and can only perceive black and white gray levels. Therefore, low-light images will make the image seen by the human eye appear blurry, dim, lack detail, and color vibrancy.

In the past few decades, researchers have proposed several non-learning-based image enhancement methods, such as the adaptive histogram equalization method and the multiscale retinex theory [11], [12]. Most of them can significantly improve the contrast and brightness of an image. However, these methods are difficult to reduce or suppress noise directly, and may even amplify noise or cause color distortion in

Manuscript received 5 May 2023; revised 28 June 2023; accepted 13 July 2023. Date of publication 21 July 2023; date of current version 16 July 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 61873086 and in part by the Science and Technology Support Program of Changzhou under Grant CE20215022. The associate editor coordinating the review of this article and approving it for publication was Prof. Lan Lan. *(Corresponding author: Jianjun Ni.)*

Guangyi Tang and Weidong Cao are with the School of Artificial Intelligence and Automation, Hohai University, Changzhou, Jiangsu 213022, China (e-mail: tang_gy@hhu.edu.cn; cwd2018@hhu.edu.cn).

Jianjun Ni and Yan Chen are with the School of Artificial Intelligence and Automation and the College of Information Science and Engineering, Hohai University, Changzhou, Jiangsu 213022, China (e-mail: njjhhuc@gmail.com; stcy401_doc@hhu.edu.cn).

Simon X. Yang is with the Advanced Robotics and Intelligent Systems (ARIS) Laboratory, School of Engineering, University of Guelph, Guelph, ON N1G 2W1, Canada (e-mail: syang@uoguelph.ca).

Digital Object Identifier 10.1109/JSEN.2023.3296167

1558-1748 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.

the process of enhancement. The following noise removal operations often bring blur and details disappear.

Recently, with the development of deep learning technology [13], [14], many image enhancement methods based on supervised and unsupervised learning have been proposed. For example, Gharbi et al. [15] incorporated deep networks with the idea of a bilateral grid processing and local affine color transforms with pairwise supervision. Chen et al. [16] adopted a learning-based approach to estimate image blurriness. These methods have obtained a good effect, which proves that the deep learning methods can automatically learn features of images without human design feature extractors and can be applied to different image enhancement tasks. Therefore, we use the deep learning-based method for image enhancement in this article.

For most supervised work, there must be a hyperparameter during the training process connecting the input and reference images, which can be used to adjust the contrast and brightness of the entire image during the testing process. However, hyperparameters are difficult to obtain automatically, and even with careful adjustment of the hyperparameters, the brightness and contrast of some local areas may not be satisfactory. In addition, in the field of low-light image enhancement, paired data sets are difficult to collect. In other words, collecting two images of the same scene with the same content and details but with different lighting conditions is very difficult or even impractical. Therefore, it is necessary to consider how to make the model learn the features between the two image domains without the paired dataset.

To eliminate the reliance on paired data, the researchers developed several methods based on unsupervised deep learning [17], [18]. These methods encourage the distribution of the generated image to be close to the target image without pair supervision and have good generalization in real-world scenes. However, the contrast and lighting of the enhanced image may not be satisfactory and there is usually a problem of color distortion and inconsistency. Therefore, this article will focus on an unsupervised-based low-light image enhancement model.

Among the unsupervised-based methods, the generative adversarial networks (GANs) have achieved impressive results in the creation of unpaired maps between low-light and normal-light image spaces [19], [20]. There are some limitations of the general GANs, such as the non-convergence and collapse problems in the training and learning process of the model. To deal with these problems, some improved GAN-based models have been proposed, and cycle-consistent GAN (CycleGAN) is one of the most successful models [21], [22]. CycleGAN has two main improvements compared with the general GANs: 1) CycleGAN can transform between input images and target images and 2) CycleGAN allows the input image to be restored to the original image after being transformed by two generators. The CycleGAN-based methods have been used widely in the image processing field, such as image style transfer and image restoration [23], [24]. In this article, the CycleGAN is used for image enhancement, to take full advantage of its abilities of the cycle consistency to guarantee the content similarity between input and output images and the adversarial loss to guarantee the quality of output images. However, there is still some room for further improvement of CycleGAN-based models. For example, the generators of CycleGAN cannot explicitly distinguish multiple image domain information of an image, so the generated results of CycleGAN have obvious drawbacks of arbitrary variation of irrelevant domain features. Therefore, the main purpose of this article is to propose an improved low-light enhancement network based on CycleGAN, to further improve the enhancement performance of the model.

The main contributions of this article are as follows. 1) A CycleGAN-based low-light image enhancement framework is proposed. In this framework, the CycleGAN is improved, where the generator is constructed based on an U-Net structure. In addition, the U-Net structure is improved by an adaptive instance normalization (AdaIN) to learn the style of the normal light image. 2) A detail enhancement module based on multilayer guided filtering is presented, to improve local visibility by adding a high-frequency component to the generated image. 3) A joint training strategy based on structural similarity is used to strengthen the constraints on generating normal-light images. Finally, in order to verify the validity and feasibility of the proposed method, several experiments are carried out and the experimental results are analyzed in detail.

This article is organized as follows. Section II introduces the related work. Section III gives out the details of the proposed model. The low-light enhancement experiments under various scenes are given in Section IV. Section V carries out some discussions on the proposed model. The conclusions and possible future research are given out in Section VI.

II. Related Work

Low-light image enhancement approaches amplify illumination and improve the visibility of dark images, which can be divided into three main classes and will be introduced simply in this section.

A. Traditional Methods

Low-light image enhancement has been actively studied as an image-processing problem for a long time, some of the classical algorithms are adaptive histogram equalization [11], Retinex model [25], and multiscale Retinex model [26]. Specifically, Celik and Tjahjadi [27] constructed the 2-D histogram using the mutual relationship between each pixel and its neighboring ones and conducted mapping the diagonal elements of the input histogram to the target histogram for image enhancement. Guo et al. [28] proposed a simple yet effective low-light image enhancement (denoted as LIME), where the illumination of each pixel was first estimated by finding the maximum value in its RGB channels, then the illumination map was refined by imposing a structure prior.

Ren et al. [29] further proposed a joint microlight enhancement and denoising strategy that also considered noise mapping to suppress noise and improve contrast compared to the conventional Retinex model. Since the coherence with the perceptual quality is one of the important factors to evaluate

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.

the performance of low-light image enhancement, several studies have attempted to apply the image quality assessment to the enhancement process. For example, Gu et al. [30] established a robust image enhancement framework based on quality optimization to successively rectify image brightness and contrast to a proper level.

Even though the above methods are conceptually simple and easy to implement, these methods are highly dependent on the specific conditions of a given image and thus still have their limitations.

B. Supervised Methods

Inspired by the great success of deep learning for the task of image recognition, several researchers started to adopt the deep neural network (DNN) to infer the relationship between low-light inputs and normal outputs. For training parameters of the network, the simple way is to use image pairs of low-light input and the target output, with appropriate loss functions. For example, Lore et al. [31] proposed a deep autoencoder-based approach to identify signal features from low-light images and adaptively brighten images. Wei et al. [32] proposed an enhancement network which is learned with only key constraints (denoted as RetinexNet). Zhang et al. [33] proposed a network (denoted as KinD), which decomposed images into reflectance and illumination components, to effectively remove visual defects amplified through lightening dark regions. Yang et al. [34] proposed a network (denoted as SGM), where the authors jointly trained the subnetworks for image decomposition and illumination enhancement. Lu and Jung [35] combined coefficient estimations with a joint operation to solve the problem of joint illumination adjustment, color enhancement, and denoising.

In addition, some studies have tried to combine multiscale features in order to consider global and local properties separately in the enhancement process. For example, Li et al. [36] presented a lightweight and efficient luminance-aware pyramid network to reconstruct normal-light images in a coarse-to-fine strategy, which brightened up low-light images with rich details and high contrast. Lu et al. [37] proposed a multibranch topology residual block to alleviate training difficulties and grasp global and local features more efficiently.

Although supervised learning-based methods offer significant improvements in low-light image enhancement, there is an obvious limitation that the enhanced results are strongly influenced by the characteristics of the training data.

C. Unsupervised Methods

Collecting a large number of paired datasets from uncontrolled scenarios is very difficult. Therefore, some unsupervised learning methods that do not require paired training images have been proposed. For example, Zhu et al. [38] realized unsupervised end-to-end image translation using CycleGAN. Kosugi and Yamasaki [39] proposed a reinforcement learning framework where the generator works as the agent that selects parameters of Adobe Photoshop and is rewarded when it fools the discriminator. Liang et al. [40] presented a deep learning-based self-supervised low-light image enhancement method, where the reflectance and illumination of the input image were parameterized by two untrained neural networks. Kwon et al. [41] used synthetic images to identify dark areas and used dark visual perception attention to enhance the brightness of dark areas. The method proposed in [41] is denoted as DALE. Jiang et al. [42] proposed a highly effective unsupervised GAN (denoted as EnlightenGAN), to regularize the unpaired training using the information extracted from the input itself and prove to generalize very well on various real-world test images. Guo et al. [43] presented a novel method ZeroDCE to estimate globally adjusted mapping curves in a pixel-wise manner without any paired or unpaired data. To do this, they carefully designed a set of no-reference loss functions to evaluate the visual quality of the enhancement result. They further presented an accelerated and light version of ZeroDCE, called ZeroDCE++, which is optimized by iteratively minimizing a loss function without any prior image samples [44]. Ma et al. [45] developed a new self-calibrated illumination (denoted as SCI) learning framework for fast, flexible, and robust brightening images in real-world low-light scenarios.

Recently, other unsupervised learning methods have emerged to deal with the low-light image enhancement problem. For example, Zhang et al. [46] provided customized low-light image enhancement output by applying strategies flexibly at different times (denoted as ReLLIE). Zheng and Gupta [47] proposed a recurrent image enhancement network to progressively enhance the low-light image with affordable model size (denoted as SGZ). Fan et al. [48] proposed an image enhancement network (denoted as HWM-Net) based on an improved hierarchical model. To increase the brightness while preserving edge information, Lv et al. [49] designed an attention-guided illumination adjustment network (denoted as DPDBL).

The above models have achieved good results in the low-light enhancement. However, the contrast and illumination of enhanced images by the current unsupervised methods may not be good enough, which are often affected by color distortion and inconsistencies. Moreover, their performance in unknown scenes lacks stability due to the dependence on datasets. Thus, an improved CycleGAN-based unsupervised method is proposed in this article, which can be flexibly applied to enhance real low-light images from different domains.

III. Proposed Method

To further improve the performance of the low-light image enhancement method, an improved CycleGAN network is proposed to restore the texture and features of the low-light image. In the CycleGAN network, there are two generators G, F, and two discriminators D_A, D_B [21], [38]. For two sample spaces X (low light image sample) and Y (normal light image sample), there are two mapping relations: $X \rightarrow Y$ and $Y \rightarrow X$. The generator G transforms the image sample X into a $G(X)$ approximating the image sample Y, and uses discriminator D_A to judge whether $G(X)$ is a real image or not. The generated image $G(X)$ is transformed into $F(G(X))$

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.

similar to sample image by F generator. In order to ensure that the resulting $F(G(X))$ is as similar as possible to the original sample image, a loss of cycle consistency is introduced to ensure that the $F(G(X))$ is more and more similar to the sample image X during continuous training. This process is a forward-looping process. Since the input of generator F is transmitted by generator G, in order to ensure the learning effect of generator F, a reverse cycle process is set up to realize the cycle process of sample image Y.

Although the general CycleGAN can be trained with unpaired datasets to obtain the same results as paired datasets, the performance of the general CycleGAN needs further improvement. In the general CycleGAN, the ability to recover some details is not enough, and the brightness of the enhanced image may be unbalanced. At the same time, the traditional discriminator is not enough to judge the detail of the enhanced image, which makes the convergence of the network difficult. To deal with these problems, an improved CycleGAN-based model is presented in this article. The main differences between the proposed model and the general CycleGAN-based methods are as follows: 1) the structure of the CycleGAN generator is improved based on U-Net, where the U-Net is also improved; 2) a detail enhancement module is added into the CycleGAN-based model by using the multilayer guided filter, making it more suitable for low-light enhancement tasks; and 3) a joint training strategy is used for the CycleGAN-based model, where the structural similarity and cycle consistency are jointly trained as a loss function in this article. The structure of the proposed model is shown in Fig. 1, and will be introduced as follows.

A. Generator Structure Based on U-Net

In the CycleGAN-based low-light enhancement model, G and F represent the normal light generator and the low-light generator, respectively, which are composed of encoder, converter, and decoder. The general operations of encoding and decoding often lead to the loss of feature information, which makes the final generated image significantly different from the original image. The main reasons are analyzed as follows: the input of generator G contains not only the low-light images from the dataset but also the low-light images generated by generator F. Obviously, the distributions of these two low-light images are different, so using the same generator G to learn effective low-light information from the data with different distributions is not conducive. Similarly, generator F cannot learn the normal-light information effectively either.

In order to deal with the problem above, a generator model based on U-Net is proposed in this article [50], because the jump connection structure of U-Net can be used to avoid the loss of feature information. However, the general U-Net structure still has many shortcomings. First, using of a 2-D convolution pooling operation to extract the features of low-light images will lead to the loss of rich spatial information in enhanced images. Second, a large amount of context information is lost in the down-sampling process, and the target details and corresponding spatial dimensions are not fully recovered during the up-sampling process. To deal with these problems, the U-Net structure is improved in this article, where a multiscale feature extraction module is used, and some AdaIN transforms are introduced to transfer the autoencoder features [51]. The proposed generator structure is shown in Fig. 2.

The improved U-Net generator structure proposed in this article includes three parts: encoder, transcoder, and decoder. The encoder is a down-sampling process, which is used for feature information compression and extraction of low-light images.

First, a multiscale convolution module is used to extract the features of the original image. In general, low-light images will appear large areas of darkness, resulting in a local feature that is relatively simple. To deal with this problem, in this article, 1×1, 3×3, 5×5, and 7×7 convolution kernels are used in the generator to construct multiscale convolution modules, each with 16 channels. The main idea of the proposed multiscale convolution module is analyzed as follows: In the same layer of the network, the receptive field of the network obtained by using large convolution kernels can capture the spatial relationship with multiple surrounding pixels, but is poor for capturing detailed features. While the receptive field of the network using small convolution kernels can focus on the geometric details of the image, but the ability to characterize the image is poor. Therefore, the different receptive fields obtained by multiscale feature extraction can further improve the image characterization ability.

Subsequently, the encoder performs three down-sampling operations, compressing the feature map of 256×256 to the size of 32×32. Each down-sampling module consists of a 3×3 convolution core (the step size is 2), AdaIN, and the activation function ReLU. AdaIN aligns the normalized channel mean and variance of the content image with the style image so that the resulting image has the same feature distribution as the low-light image. Unlike other normalization methods, such as batch normalization (BN) and instance normalization (IN), AdaIN does not have learnable affine parameters, but adaptively generates a set of parameters based on the input style image. The AdaIN is described as follows:

$$\text{AdaIN}(x, y) = \sigma_y \left(\frac{x - \mu_x}{\sigma_x} \right) + \mu_y \tag{1}$$

where x and y are the characteristic images of low-illumination images and normal-light images, respectively; μ_x and μ_y denote the mean of the x and y images; σ_x and σ_y denote the standard deviation of the x and y images. The proposed CycleGAN generator with AdaIN layer can help identify whether the removed feature is noise or real structure between two target domains. Based on this structure, the feature map can be used effectively.

The transcoder is mainly used to integrate the image features extracted by the encoder. The general U-Net only relies on the down-sampling module, which leads to domain shift problems inevitably in generated images, due to the lack of effective supervised signals for low-light images. To deal with these problems and speed up the convergence of the network, six residual blocks are used in this article to convert the features

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.

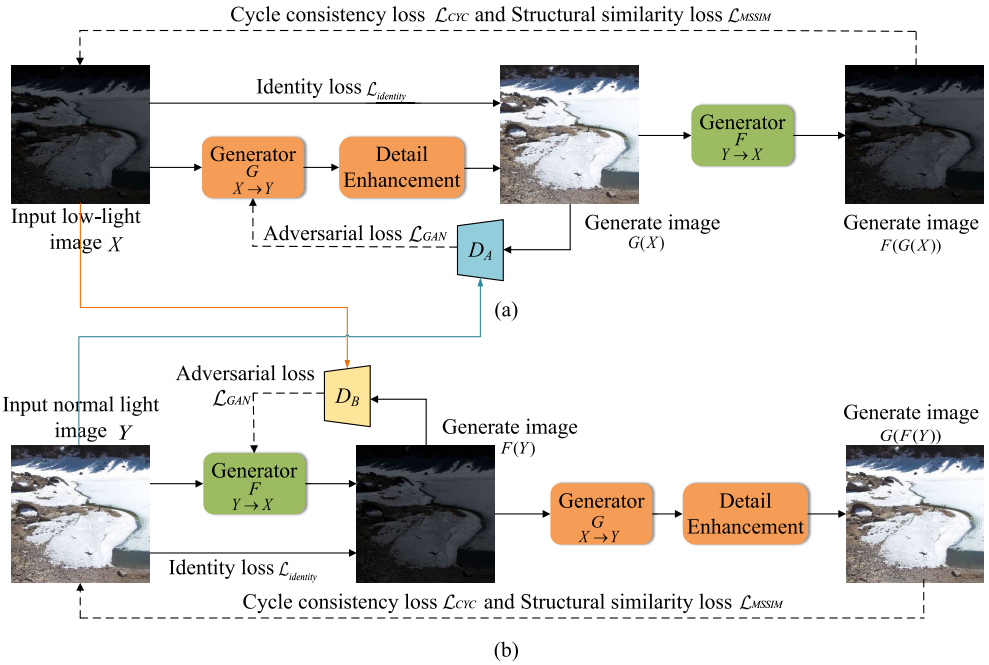


Fig. 1. Proposed low-light image enhancement network based on CycleGAN. (a) Branch one, which first performs the low-light image enhancement process and then converts the enhanced image into a low-light image. (b) Branch two, which first converts the normal light image into a low-light image and then performs the low-light enhancement process. Both generators need to update their parameters simultaneously, so that they can maintain consistency in both forward and backward conversions. All the networks are trained using the adversarial loss $\mathcal{L}_{\text{GAN}}$, cycle consistency loss $\mathcal{L}_{\text{CYC}}$, identity loss $\mathcal{L}_{\text{identity}}$, SSIM loss $\mathcal{L}_{\text{MSSIM}}$.

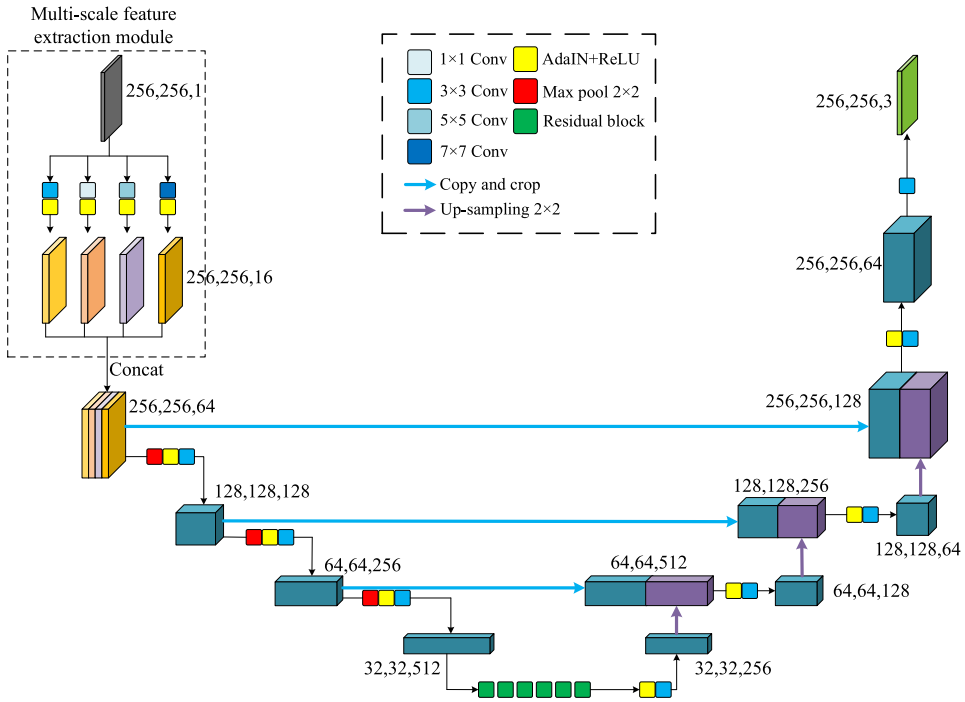


Fig. 2. Improved U-Net generator structure proposed in this article. Here, blue cubes represent feature maps generated by encoder down-sampling, and purple cubes represent feature maps generated by decoder up-sampling. They have the same size and number of channels, and are connected by jump structures. In this proposed structure, multiscale feature extraction module can effectively capture different levels of details and structures. AdaIN layer is used to speed up model convergence and reduce memory requirements. The transcoder uses six residual blocks (the green squares), with 3 $\times$ 3 Conv and AdaIN + ReLU, to integrate the image features extracted by the encoder.

of low-light images into normal-light images. This is because the residual blocks can add the learned residual information to the output. In addition, the residual blocks can solve the problem of gradient disappearance and degradation during the training of DNNs [52], [53].

The decoder is used to recover the original resolution of the feature map and consists of two parts: skip connection and up-sampling module. Skip connection fuses the underlying position information with the deeper feature information by concatenating. The up-sampling module consists of a

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.

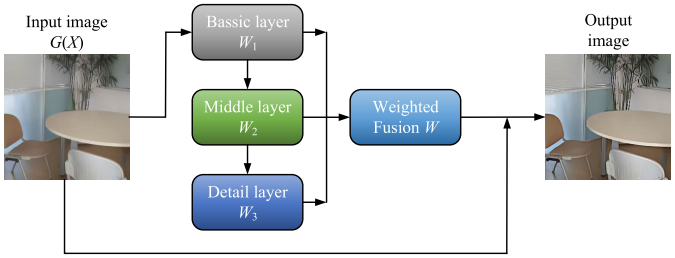


Fig. 3. Work flow of the detail enhancement based on multilayer guided filtering.

combination of 3 × 3 convolution core (the step size is 2), AdaIN, and the activation function ReLU for restoring the image size to the same size as the input.

Remark 1: In the proposed generator structure based on U-Net, each layer of convolution is replaced by a multiscale feature extraction module, which can make the generated images closer to the original images. In addition, with the repeated use of residual blocks and AdaIN, the generator network is able to extract more low-light image detail information and thus has better differentiation effect for low-light edge regions.

B. Detail Enhancement Based on Multilayer Guided Filtering

The amount of image detail information is determined by the image exposure. The more appropriate the image exposure is, the richer the image detail information will be. As we know, the low-light image has almost invisible detail information. Although the low-light images can be enhanced by the proposed generators (see Section III-A), some semantic information that should be retained is lost. In addition, the over-learning of the model for different parts of the image brightness information will bring about unnatural voids.

To deal with these problems, a detail enhancement module is presented in this article, to enhance image details and edge contours. The basic idea of the proposed detail enhancement module is as follows: considering that the image details are concentrated in the high-frequency part, the generated image $G(X)$ is decomposed into multiple layers. Then the different layers are enhanced to different degrees. Finally, the enhanced layers are fused to get the final output image. This method can improve the contrast and sharpness, while preserving the edges and details of the output image. The work flow of the proposed detail enhancement module is shown in Fig. 3.

As shown in Fig. 3, the generated image is decomposed into the basic layer, middle layer, and detail layer first. For the kth image to be fused, the input image is smoothed by circular mean filtering in the base layer, and the calculation is shown as follows:

$$W_{1_k} = G(X)_k \times dr \tag{2}$$

where W_{1_k} is the basic layer of the kth image; $G(X)_k$ is the input image; and dr is the circular mean filter with radius r. Then, by subtracting the base layer from the image to be fused, the second layer in the three-scale fusion frame, i.e., the middle layer is obtained, which is calculated by

$$W_{2_k} = G(X)_k - G(X)_k \times dr \tag{3}$$

where W_{2_k} is the middle layer of the kth image, and then subtracts the smoothed middle layer based on the circular mean filter to get the third layer, the detail layer, which is shown as follows:

$$W_{3_k} = W_{2_k} - W_{2_k} \times dr \tag{4}$$

where W_{3_k} is the detail layer of the kth image.

Then, we get the multiscale guided filter detail image W_k by weighted fusion of three detail images W_{1_k}, W_{2_k}, and W_{3_k}, namely

$$W_k = \left(1 - \omega_1 \times \operatorname{sgn}\left(W_{1_k}\right)\right) \times W_{1_k} + \omega_2 \times W_{2_k} + \omega_3 \times W_{3_k} \tag{5}$$

where ω_1, ω_2, and ω_3 are the regulation parameters. For the sake of easy realization, they are set as 0.5, 0.5, and 0.25, respectively in this study. The basic layer W_{1_k} expands the gray level differences near the edge but may cause the enhanced image to be oversaturated. The part of $(1 - \omega_1 \times \operatorname{sgn}(W_{1_k}))$ in (5) is used to avoid this problem by reducing the positive component and amplifying negative components of basic layer W_{1_k}. Finally, we add the overall detail W_k to the global enhancement image $G(X)_k$.

Remark 2: Based on the proposed detail enhancement module, a three-scale image decomposition is used to separate the generated image into multiscale images with different frequency intensities. Then they are fused considering the detailed information of different layers. Thus, the generated image based on the proposed model can be clearer and have more details.

C. Joint Training Strategy Based on Structural Similarity

CycleGAN consists of two mirror-symmetric GAN networks. During the joint training process of the two networks, the generator and discriminator play against each other, and the data distribution generated by the generator will be closer to the real data distribution until the discriminator is successfully fooled. Because the traditional CycleGAN network can only distinguish the parts with the most significant differences between the two domains during the training process, it will ignore important features or regions in the image. Moreover, the generator will introduce artifacts and noise when generating texture structures that do not exist or are uncommon in the target domain. To deal with these problems, a joint training strategy based on structural similarity is proposed in this article.

In the joint training of the CycleGAN network, the loss function plays a vital role. The traditional CycleGAN mainly includes the adversarial loss, cycle consistency loss, and identity loss. The idea of CycleGAN network is adversarial training, using adversarial loss to optimize the generator and discriminator continuously and scoring the cost and real data by the discriminator. The generator needs the discriminator to judge the generated image as a real image, and the

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.

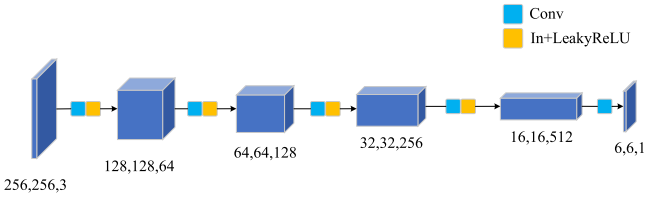


Fig. 4. Structure of the proposed discriminator.

discriminator needs to judge whether it is real data or generated samples.

In the proposed model, the structure of the generator is introduced in Section III-A, and the discriminator has a similar structure to the discriminator introduced in Patch GAN (Patchgan) [54]. The structure of the proposed discriminator is shown in Fig. 4.

As shown in Fig. 4, the first four convolution blocks contain a convolution layer, an IN layer and a nonlinear activation function LeakyReLU, with a step size of 2. At the last level, the output is obtained by reducing the number of channels to one. The loss functions for the mapping $X \rightarrow Y$ and its corresponding discriminator D_A are as follows:

$$\mathcal{L}_{\mathrm{GAN}}(G, D_A, X, Y) = \mathbb{E}y \sim p_{\mathrm{data}}(y)\left[\log D_A(y)\right] + \mathbb{E}y \sim p_{\mathrm{data}}(y)\left[\log\left(1 - D_A(G(x))\right)\right] \tag{6}$$

where x represents the sample of the normal photo data set, and y represents the sample of the low-light data set. $p_{\mathrm{data}}(x)$ and $p_{\mathrm{data}}(y)$ represent the sample distribution for the image domain X and Y, respectively. $G(x)$ represents the generated image, and D_A is a $0-1$ classifier used to distinguish the generated image from the normal-light image. G tries to minimize this loss function, and D_A tries to maximize it.

The loss of cycle consistency is a new loss function introduced by CycleGAN. When the image of the source domain passes through the generator to the target domain and then back to the source domain through the generator, the cycle consistency loss is used to constrain the similarity between the original image of the source domain and the reconstructed image, so that the reconstructed image is close to the original input image, that is to say, $F(G(x)) \approx x$. CycleGAN uses the $L1$ norm to calculate this loss as follows:

$$\mathcal{L}_{\mathrm{CYC}}(G, F) = \mathbb{E}x \sim p_{\mathrm{data}}(x)\left[\|F(G(x)) - x\|_1\right] + \mathbb{E}y \sim p_{\mathrm{data}}(y)\left[\|G(F(y)) - y\|_1\right] \tag{7}$$

where $F(G(x))$ and $G(F(x))$ are the reconstructed images; and $\|\cdot\|_1$ is the $L1$ norm function. The $L1$ norm function has good robustness and a stable gradient to the input value, which can avoid the problem of gradient explosion.

In order to avoid the image distortion caused by GAN loss in the target domain, in the CycleGAN network, the function of generator G is to generate a Y-style domain image. So, when inputting a Y domain image, the G should also generate a Y-style domain image. In this case, $G(y)$ is required to be as close as possible to y, and $F(x)$ is required to be as close as possible to x, so that the enhancement image and the input image have the same color. After adding the identity loss function, the generator G adds the input of image y to the original input, and the generator F adds the input of image x to the original input, which means that the generator compares the difference between the generated image and the original image for feedback optimization. If there is no identity loss, the generator may modify the color of the image, making the color of the image distorted. The identity loss is defined as follows:

$$\mathcal{L}_{\mathrm{identity}}(G) = \mathbb{E}y \sim p_{\mathrm{data}}(y)\left[\|G(y) - y\|_1\right] + \mathbb{E}x \sim p_{\mathrm{data}}(x)\left[\|F(x) - x\|_1\right]. \tag{8}$$

In the traditional CycleGAN network, the generators G and F cannot clearly distinguish different kinds of low-light information. So the image will randomly change in the low-light area during the enhancement process, which will produce noise and distortion. Although a detail enhancement module is used in the proposed model (see Section III-B), it mainly focuses on improving image details and edge contours, which does not consider the structural information. In order to solve this problem, SSIM loss is introduced in this article, to distinguish low-light regions with different luminance and preserve the characteristics of low-light regions as much as possible.

SSIM is a perception-based model, which treats image degradation as a perceptual change of structural information, and combines important perception phenomena, including luminance masking and contrast masking [55]. It is an image quality assessment method that reflects the structural similarity between two images. The main reason we introduce SSIM loss into the CycleGAN network is that SSIM loss can provide more flexibility for detail structure modeling and improve the performance of image restoration [56], [57].

We use SSIM loss in the final loss function to generate an image closer to the target image, taking into account the details, texture, and color information. The SSIM between two images can be defined as

$$\mathrm{SSIM}(x, y) = \frac{\left(2\mu_x\mu_y + C1\right)\left(2\sigma_{xy} + C2\right)}{\left(\mu_x^2 + \mu_y^2 + C1\right)\left(\sigma_x^2 + \sigma_y^2 + C2\right)} \tag{9}$$

where μ_x and μ_y denote the mean value of the x and y images; σ_x and σ_y denote the standard deviation of the x and y images; σ_x^2 and σ_y^2 denote the variance images of the two images; σ_{xy} denotes the covariance of the x and y images; and $C1$ and $C2$ are constants set to avoid the denominator being zero.

In this article, considering that there is a strong correspondence between the enhanced image $F(G(x))$ and the original image x, the SSIM loss is used to preserve their content and structure. In addition, the SSIM loss can preserve the contrast of the high-frequency regions (edges and details) effectively. The SSIM loss in this study is defined as

$$\mathcal{L}_{\mathrm{MSSIM}}(G) = 1 - \mathrm{MSSIM}(x, F(G(x))) \tag{10}$$

where MSSIM is the average of the SSIM of each local image block in images x and y

$$\mathrm{MSSIM}(x, y) = \frac{1}{N}\sum_{j=1}^{N}\mathrm{SSIM}\left(x_j, y_j\right) \tag{11}$$

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.

where x_j and y_j are the j block of image x and y, respectively; and N is the number of local blocks.

After the above improvements, the total target loss used in the improved CycleGAN network can be determined, which is as follows:

$$\mathcal{L}(G, F, D_A) = \mathcal{L}_{\text{GAN}}(G, D_A, X, Y) + \mathcal{L}_{\text{CYC}}(G, F) + \lambda \mathcal{L}_{\text{identity}}(G) + \tau \mathcal{L}_{\text{MSSIM}}(G) \quad (12)$$

where λ and τ are weighted parameters that represent the loss of identity and loss of structural consistency, which are set as 0.6 and 0.3, respectively, in this study based on experiments.

CycleGAN's training can be accomplished by solving the following min-max problems:

$$\min_{G,F} \max_{D_A} \mathcal{L}(G, F, D_A) \quad (13)$$

where the discriminator is trained to maximize the total loss and the generator is trained to minimize it. The generator and discriminator are updated alternately for counter training.

Remark 3: In the loss of the joint training strategy, the adversarial loss promotes sharper images, but ignores detailed textures. SSIM loss captures more detailed textures, but usually results in blurred images. The combination of these two losses preserves texture and generates clear boundaries. The cycle consistency loss and the identity loss make the generated images more realistic. Because the joint training strategy considers the problems in low-light image enhancement comprehensively, the low-light image can be enhanced efficiently.

IV. Experiments

In this section, quantitative and qualitative experiments are conducted to evaluate the performance of our model.

A. Implementation Settings

During the training process, the batchsize is set to 1 and the number of epochs is set to 200. The learning rate is set to 0.0001 in the first 100 epochs, and will be linearly decayed to 0 in the next 100 epochs. The network is optimized by using Adam optimizer, and the hyperparameters of optimizer are set as (0.9, 0.999, and $10e^{-8}$). The training and testing of the network are completed on Nvidia GTX 2080Ti GPU and Inter Core E5-2620 CPU, and the code is based on the PyTorch framework. In terms of parameter size, our network's stored PyTorch model size is 39.4 MB, which is 26.77% (relative value) less than the baseline model CycleGAN (53.8 MB). Thus, our model is more lightweight than the general CycleGAN.

In this study, our model is compared with some classic and state-of-the-art methods including LIME [28], KinD [33], RetinexNet [32], CycleGAN [38], ZeroDCE++ [44], EnlightenGAN [42], DALE [41], SGM [34], DPDBL [49], ReLLIE [46], SGZ [47], HWM-Net [48], and SCI [45]. The principle of selecting these models for the comparison experiments in this study is that the code of these models can be obtained or the experiments of these models are also conducted on the same public dataset used in this article.

TABLE I
Qualitative Results Among Different Methods on the Indoor Scenes

Model	PSNR	SSIM	LPIPS
LIME [28]	12.283	0.810	0.181
CycleGAN [38]	13.887	0.788	0.198
ZeroDCE++ [44]	14.862	0.890	0.155
ReLLIE [46]	17.470	0.8691	0.130
DALE [41]	17.615	0.859	0.160
EnlightenGAN [42]	17.744	0.930	0.125
SCI [45]	18.098	0.911	0.122
RetinexNet [32]	18.857	0.876	0.282
SGM [34]	19.108	0.884	0.117
DPDBL* [49]	20.230	0.840	–
SGZ* [47]	20.600	0.793	–
KinD [33]	22.252	0.965	0.049
HWM-Net [48]	24.227	0.928	0.068
Ours	**27.783**	**0.972**	**0.029**

Note: The results of the model marked with * is obtained from the original paper directly; '-' means that the original paper didn't provide the value on this metric. The best results are highlighted in bold.

Three metrics are adopted for quantitative comparison including peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and learned perceptual image patch similarity (LPIPS) [58]. PSNR and SSIM are reference image quality assessment methods, which indicate the noise level and the structural similarity between the result and the reference, respectively. LPIPS is designed for human perception. It should be noted that these indexes can only reflect the image quality in some aspects, which are not completely consistent with the evaluation results given by the human visual system.

B. Evaluation on the Indoor Scenes

We first evaluate our model on the indoor scenes using the LOL dataset. The LOL dataset is a large-scale dataset that includes 500 indoor paired images [32]. The low-light images in the LOL dataset contain noise generated during the shooting process. In the comparison experiments, 485 low-light images are used for training, and the left image pairs are used for testing. The comparison results of different methods on the LOL dataset are shown in Table I. Some visual results of different models are shown in Fig. 5, where the results of DPDBL are not given out because its code cannot be obtained.

The results in Fig. 5 show that LIME improves the brightness of images but meanwhile introduces blur and noise. Although the image processed by RetinexNet has a little more information, the overall image is too bright, the contrast is low, and the distortion is obvious. The CycleGAN network has serious color distortion and a lot of details are lost. The ZeroDCE++ algorithm can better preserve image details, but the image brightness is too high and the visual performance is poor. The color of the image generated by DALE is not consistent with the reference image. SGM produces good

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.

Fig. 5. Visual comparison results on the indoor scenes: (a) low-light image; (b) based on LIME; (c) based on CycleGAN; (d) based on ZeroDCE++; (e) based on ReLLIE; (f) based on DALE; (g) based on EnlightenGAN; (h) based on SCI; (i) based on RetinexNet; (j) based on SGM; (k) based on SGZ; (l) based on KinD; (m) based on HWM-Net; (n) reference; and (o) based on our model.

image enhancement results. However, some details are lost. The EnlightenGAN model has less color discrimination in dark places and loses some details. The whole image processed by the SCI algorithm is dark. HWM-Net works well but the color tone of enhanced images is higher than the original image. The images generated by ReLLIE and SGZ are dark. Our model outputs a slightly brighter image, which is not over-brightened and washed-out like ZeroDCE++ output. In addition, our model can better maintain the structure information and color information.

From Table I, we can see that the proposed method outperforms all the other methods. The PSNR of the proposed model are 14.68% higher than HWM-Net (the second best model). Higher PSNR means our model is less affected by noise, less distortion, and better image enhancement. The SSIM of the proposed model are 0.73% higher than KinD (the second best model). A better SSIM value indicates that our model better preserves the structural information of the image and improves the overall quality of the image. Considering LPIPS is designed for human perception, the LPIPS of the proposed model reduces by 40.82% compared with KinD (the second best model), which means not only the brightness of the enhanced image based on our model is very close to the brightness of the normal illumination image, but also the enhanced image is more natural and the visual performance is better.

C. Evaluation on the Outdoor Scenes

In order to further evaluate the performance of the method proposed in this article, we perform the evaluation on outdoor scenes. In this study, we manually selected 976 low-light outdoor images from the Ve-LOL-L dataset, where 876 images are used for training and 100 images for testing. The Ve-LOL-L dataset is a subset of the large-scale dataset Ve-LOL, which includes synthetic images with diverse backgrounds and various objects [59]. Fig. 6 shows some enhancement results by different methods, and Table II displays the results of different methods. Here, only the models with code are compared with the proposed model.

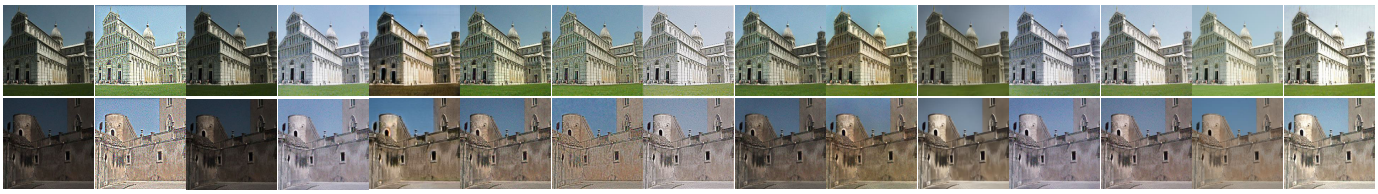
In Fig. 6, we can see that the low-light image enhancement performance of all the models decrease obviously. The main reason is that there are much noise in the images of the outdoor scenes from the Ve-LOL-L dataset. The generated images of LIME, SCI, and SGZ are dark, but SCI and SGZ are able to suppress noise well and retain image details. The RetinexNet model greatly improves the contrast of the image, and it corrupts the output with large noise content. CycleGAN generates images that are gray and have severe color bias. ZeroDCE++ is able to improve contrast without significant display halo, but it tends to amplify the noise in the image. EnlightenGAN also has severe color bias. DALE has low visibility and under-exposure. SGM and ReLLIE over-enhance the local illumination effect and increase the noise. The color of the images generated by HWM-Net is not nature. The proposed model performs well with its capability to suppress noise in most of the images while improving local contrast.

TABLE II
QUALITATIVE RESULTS AMONG DIFFERENT METHODS ON THE OUTDOOR SCENES

Model	PSNR	SSIM	LPIPS
SGM [34]	10.800	0.741	0.210
KinD [33]	10.929	0.642	0.190
SGZ [47]	12.101	0.847	0.135
CycleGAN [38]	12.114	0.739	0.198
LIME [28]	13.674	0.781	0.149
RetinexNet [32]	14.217	0.823	0.251
SCI [45]	14.254	0.737	0.129
DALE [41]	14.544	0.773	0.162
EnlightenGAN [42]	14.622	0.867	0.130
HWM-Net [48]	14.903	0.868	0.149
ZeroDCE++ [44]	15.758	0.866	0.167
ReLLIE [46]	17.386	0.895	0.134
Ours	**18.763**	**0.903**	**0.115**

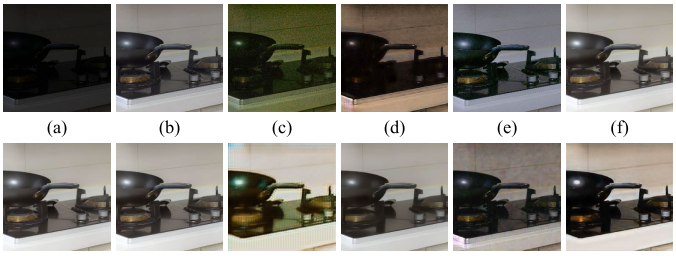
As shown in Table II, compared with the ReLLIE model with the second best results in these experiments, our model outperforms it by 7.92% and 0.89% in PSNR and SSIM, respectively, and by 14.18% lower in LPIPS. Compared with the SGM model with the worst PSNR in Table II, our model outperforms it by 73.73% and 21.86% in PSNR and SSIM,

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l) (m) (n) (o)

Fig. 6. Visual comparison results on the outdoor scenes: (a) low-light image; (b) based on SGM; (c) based on KinD; (d) based on SGZ; (e) based on CycleGAN; (f) based on LIME; (g) based on RetinexNet; (h) based on SCI; (i) based on DALE; (j) based on EnlightenGAN; (k) based on HWM-Net; (l) based on ZeroDCE++; (m) based on ReLLIE; (n) reference; and (o) based on our model.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l)

Fig. 7. Some visual experimental results of different weight parameters λ and τ: (a) low-light image; (b) reference; (c) $\lambda = 0$, $\tau = 0$; (d) $\lambda = 0$, $\tau = 0.5$; (e) $\lambda = 0.4$, $\tau = 0.3$; (f) $\lambda = 0.6$, $\tau = 0.2$; (g) $\lambda = 0.6$, $\tau = 0.3$; (h) $\lambda = 0.6$, $\tau = 0.4$; (i) $\lambda = 0.6$, $\tau = 0.5$; (j) $\lambda = 0.8$, $\tau = 0.3$; (k) $\lambda = 1$, $\tau = 0$; and (l) $\lambda = 1$, $\tau = 0.2$.

respectively, and by 45.24% lower in LPIPS. The quantitative results show that the images generated by our model are closer to the real images and have the best color enhancement and edge detail retention, which can enhance outdoor low-light images well.

V. Discussions

In Section IV, the performances of the proposed model have been proved by some comparison experiments on two public datasets. In this section, some discussions are given out on the setting of the weight parameters in the loss function, the generalization performance, the effects of the key improvements, and the computation complexity of the proposed model.

A. Setting of the Weight Parameters in the Loss Function

In the proposed model, a joint training strategy is presented and two important parameters are introduced into the CycleGAN-based model, namely λ and τ in (12). In order to choose suitable values of λ and τ, an experiment was conducted under the same conditions introduced in Section IV-B, except choosing different λ and τ.

In the process of joint training, the identity loss is used to maintain the color of the enhanced image and the SSIM loss is used to increase image details. For low-light images, the network needs to change images and restore color information. However, the excessive SSIM loss will increase the risk of creating image voids. Therefore, the range of the weight for the identity loss λ is 0–1, and the range of the SSIM loss τ is 0 to 0.5. Some quantitative results of this experiment are listed in Table III. Some visual experimental results are shown in Fig. 7.

TABLE III
Some Quantitative Results of Different Weight Parameters λ and τ

λ	τ	PSNR	SSIM	LPIPS
0	0	7.359	0.590	0.245
0	0.5	8.391	0.571	0.286
0.4	0.3	19.376	0.894	0.088
⋮	⋮	⋮	⋮	⋮
0.6	0.2	20.471	0.913	0.070
0.6	0.3	**27.783**	**0.972**	**0.029**
0.6	0.4	26.757	0.951	0.035
0.6	0.5	15.738	0.811	0.173
0.8	0.3	24.420	0.920	0.076
1	0	12.049	0.628	0.305
1	0.2	11.855	0.747	0.214

The results in Fig. 7 show that the generated image is severely distorted and details are lost, when the identity loss and the SSIM loss are not used (namely $\lambda = 0$, $\tau = 0$, see Fig. 7(c). When only using the SSIM loss, there is also a distortion problem [namely $\lambda = 0$, $\tau = 0.5$, see Fig. 7(d)]. When only using the identity loss, a large amount of noise occurs [namely $\lambda = 1$, $\tau = 0$, see Fig. 7(k)]. The results in Table III show that the comprehensive enhancement effect is best at $\lambda = 0.6$ and $\tau = 0.3$ [see Fig. 7(g)]. Therefore, the weight parameters λ and τ are set as these values above in the proposed model.

B. About the Generalization

We first discuss the generalization of the proposed model in a cross-dataset manner. Namely, all the models are trained on one dataset, but tested on a different dataset. In this study, the training dataset is LOL, and the testing dataset is VITA-Dataset, which is a large-scale multiexposure image dataset with high-resolution image sequences of multiple scenes. The reason of using VITA-Dataset as testing dataset is to ensure that there are no identical images between testing and training datasets. The comparison experimental results are shown in Table IV and Fig. 8. Because LIME is a non-learning-based low-light image enhancement model, which is not compared in the generalization experiments.

From Table IV, we can see that in this cross-dataset evaluation experiment, our model generates smoother images and eliminates most of the noise compared to the general

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.

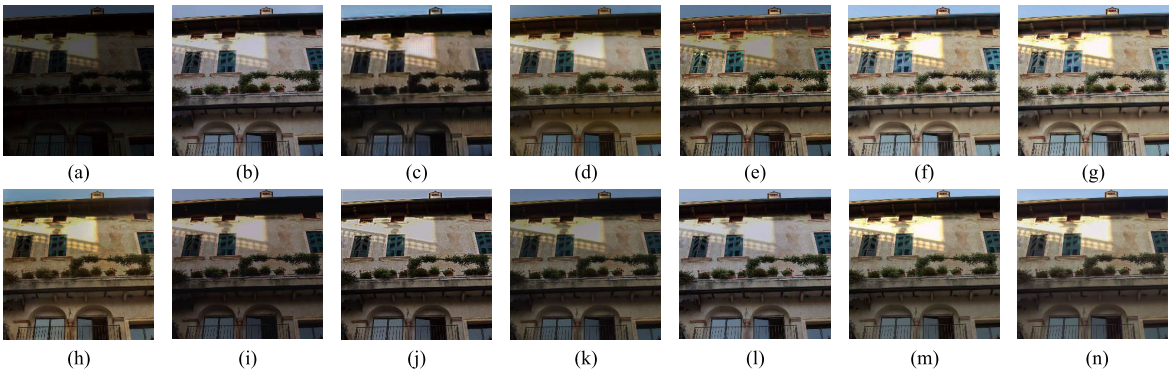


Fig. 8. Results of cross-dataset evaluation for the generalization: (a) low-light image; (b) based on SGM; (c) based on CycleGAN; (d) based on SGZ; (e) based on RetinexNet; (f) based on KinD; (g) based on SCI; (h) based on EnlightenGAN; (i) based on ReLLIE; (j) based on ZeroDCE++; (k) based on DALE; (l) based on HWM-Net; (m) reference; and (n) based on our model.

TABLE IV
RESULTS OF CROSS-DATASET EVALUATION FOR THE GENERALIZATION

Model	PSNR	SSIM	LPIPS
SGM [34]	14.335	0.700	0.226
CycleGAN [38]	14.605	0.716	0.216
SGZ [47]	14.765	0.789	0.184
RetinexNet [32]	14.809	0.694	0.317
KinD [33]	16.801	0.780	0.182
SCI [45]	17.381	0.848	0.174
EnlightenGAN [42]	17.740	0.851	0.129
ReLLIE [46]	17.963	0.845	0.171
ZeroDCE++ [44]	18.112	0.884	0.115
DALE [41]	18.353	0.811	0.133
HWM-Net [48]	**19.677**	0.836	0.144
Ours	19.590	**0.907**	**0.095**

CycleGAN. The PSNR and SSIM of the proposed model are 34.13% and 26.68% higher than the baseline model CycleGAN, respectively. In this experiment, our model is lower than HWM-Net in PSNR by 0.44%. The main reason is that our model does not consider the reference image in training, so it is difficult to ensure the similarity between the enhanced image and the reference image in terms of luminance, which has some impact on PSNR. Nevertheless, in this experiment, the proposed model has a higher SSIM (by 8.49%) and a lower LPIPS (by 34.03%) compared to the HWM-Net model. Besides, our model can obtain the best performance over other methods, which means that our improvements are very effective. Thus, the results of this experiment show that our model has good generalization performance.

As shown in Fig. 8, the results of KinD, RetinexNet, and EnlightenGAN suffer from severe noise amplification and overexposure artifacts, while ReLLIE and ZeroDCE++ do not sufficiently improve brightness and have noise. The unsupervised method CycleGAN produces very low quality due to its instability. DALE and SGM enhance the contrast of the images but lose details. HWM-Net does not fully recover the sunlit images. Although the SCI method also performs well, its results are too bright and some details

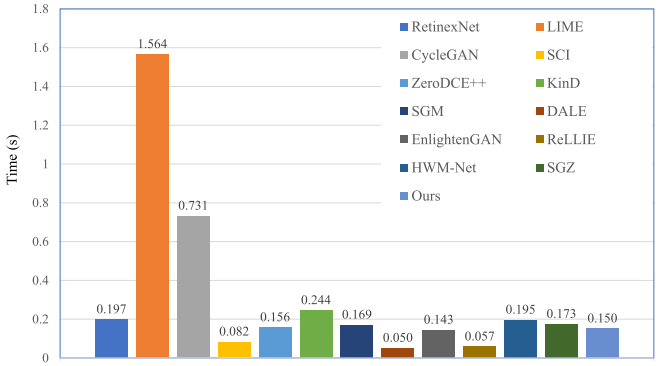


Fig. 9. Computation time for one image of each model.

are lost. In contrast, the proposed model produces the most satisfying visual enhancement results with an impressive balance between brightness and artifact/noise suppression. Due to the unpaired training, the proposed method can be easily adapted without requiring any supervised/paired data in the new environment, which greatly contributes to its usable range.

C. About the Computation Time

The computation time of each method is also an important indicator of the efficiency of a model, and the shorter the running time is, the less complex the algorithm of the model will be. In this section, to further verify the performance of the proposed model, we used the LOL dataset to conduct enhancement experiments, based on different models, which are the same as those in Section IV-B. The average computation time for enhancing one low-light image of each model is shown in Fig. 9.

The results in Fig. 9 show that the LIME algorithm and the CycleGAN algorithm take the longest time, and the computation time of the former has exceeded 1 s. The computation of ReLLIE is optimal except for the DALE model. It is important to note that ReLLIE uses a deep reinforcement learning-based lightweight framework which reduces the complexity of the model. The DALE model is the most lightweight compared to other models, and is significantly superior to others in computation time, but it cannot light up some parts of the input images, and its results also contain noticeable noise. Our model does not have

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.

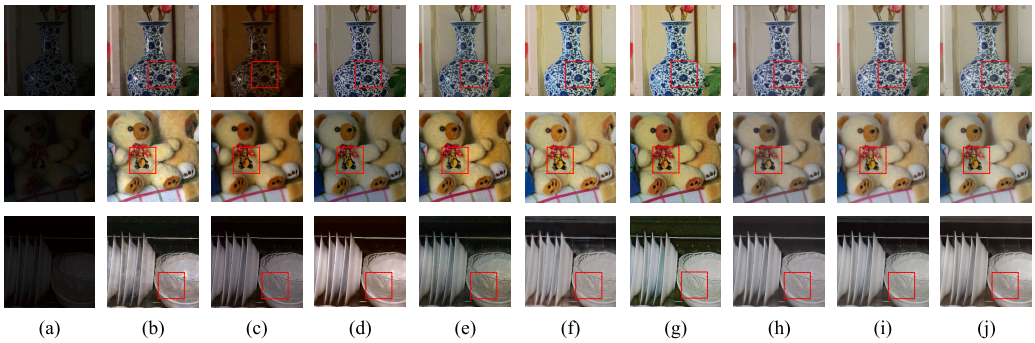


Fig. 10. Results of ablation experiments: (a) low-light image; (b) based on CycleGAN; (c) based on Model1; (d) based on Model2; (e) based on Model3; (f) based on Model4; (g) based on Model5; (h) based on Model6; (i) reference; and (j) based on our model.

TABLE V
RESULTS OF ABLATION EXPERIMENTS

Model	Generator	Joint	Detail	PSNR	SSIM	LPIPS
CycleGAN	×	×	×	13.887	0.788	0.198
Modle1	×	×	✓	14.146	0.804	0.179
Modle2	×	✓	×	14.779	0.873	0.144
Modle3	✓	×	×	17.513	0.817	0.092
Modle4	×	✓	✓	16.987	0.916	0.167
Modle5	✓	×	✓	17.906	0.862	0.174
Modle6	✓	✓	×	20.416	0.915	0.074
Ours	✓	✓	✓	**27.783**	**0.972**	**0.029**

much advantage in single image enhancement time due to the introduction of the detail enhancement module. However the enhancement of the results of each algorithm shows that our proposed model can greatly improve image quality in a relatively short period of time.

D. Ablation Experiments

To evaluate the effectiveness of the three improvements of the proposed model [the improved generator structure (Generator), the detail enhancement module (Detail), and the joint training strategy (Joint)], an ablation study of the proposed model is conducted on the LOL dataset. For the sake of description, different models used in these ablation experiments are named as "Model1," "Model2," "Model3," "Model4," "Model5," and "Model6," respectively. The details of these models can be seen in Table V. For example, only the proposed detail enhancement module is added into the baseline model CycleGAN in Model1, both the proposed joint strategy and detail enhancement module are added into CycleGAN in Model4, and so on. The results of the ablation experiments are shown in Table V. Some of the visual performance of each model for the ablation experiments is shown in Fig. 10.

The results in Table V show that each part of the three improvements in the proposed model can increase the low-light enhancement performance compared with the baseline model CycleGAN, which shows that the improvements of the proposed model are all very effective. Among these three improvements of the proposed model, the proposed generator structure based on the improved U-Net is more important than the other two parts. In addition, the results of the ablation experiments show that our model combined all the three improvements can increase the accuracy in the enhancement of images obviously compared to other models. The reasons are as follows: First, the proposed model can extract features and spatial correlation efficiently. Second, the proposed joint strategy with the SSIM loss function can learn the time dependence of the feature series. In addition, the detail enhancement can well preserve the color of the enhanced image and reduce noise.

As shown in Fig. 10, the images generated by the general CycleGAN are easily distorted, the main reason is that the generators G and F in general CycleGAN cannot clearly distinguish multiple image domain information of an image. But all the images generated by the models Modle1, Modle2, and Modle3 are better than those of the general CycleGAN, and some problems are still unsolved. For example, the images generated by Modle1 (CycleGAN + Detail) have poor color restoration, but retain the detailed information. The images generated by Modle2 (CycleGAN + Joint) are darker in color but remove noise. From the results in Fig. 10, we also can see that the results of the models with the combination of the two improvement parts can further improve the performance compared with the models with only one improvement part. For example, Model5 (CycleGAN + Generator + Detail) uses an improved U-Net generator structure, which makes the generated image closer to the original image by extracting multiscale features, but still suffers from color distortion and noise. The results of Model6 (CycleGAN + Generator + Joint) can contain true color. In contrast, the results of our model contain realistic color and are visually more satisfying, which validates the effectiveness of the three improvements of the proposed model [see Fig. 10(j)].

VI. CONCLUSION

In this article, we proposed a novel unsupervised learning method, where an improved U-Net structure is used for the generator of the CycleGAN-based model for low-light enhancement. In addition, a detail enhancement module is added into the CycleGAN-based model. Furthermore, a joint training strategy is proposed for the training of the CycleGAN-based model. Specifically, four convolutional blocks of different sizes are used in the proposed generator to extract the multidimensional features of low-light images, and the cycle consistency and structural similarity loss functions

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.

are used in the joint training strategy to make the enhanced images closer to the realistic and natural images. Furthermore, the generated images are enhanced with multiscale texture details. Experimental results on various low-light datasets show that our model is superior to many advanced methods in both subjective and objective indexes and produces visually pleasing enhanced images. In the future, we will explore how to use a more lightweight model to enhance image quality.

REFERENCES

[1] Z. Liu, P. Shi, H. Qi, and A. Yang, "D-S augmentation: Density-semantics augmentation for 3-D object detection," *IEEE Sensors J.*, vol. 23, no. 3, pp. 2760–2772, Feb. 2023.

[2] W. Zhou, J. Liu, J. Lei, L. Yu, and J.-N. Hwang, "GMNet: Graded-feature multilabel-learning network for RGB-thermal urban scene semantic segmentation," *IEEE Trans. Image Process.*, vol. 30, pp. 7790–7802, 2021.

[3] Q. Yu, C. Yang, and H. Wei, "Part-wise AtlasNet for 3D point cloud reconstruction from a single image," *Knowl.-Based Syst.*, vol. 242, Apr. 2022, Art. no. 108395.

[4] J. Ni, K. Shen, Y. Chen, W. Cao, and S. X. Yang, "An improved deep network-based scene classification method for self-driving cars," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.

[5] Z. Hao, Z. Wang, D. Bai, B. Tao, X. Tong, and B. Chen, "Intelligent detection of steel defects based on improved split attention networks," *Frontiers Bioeng. Biotechnol.*, vol. 9, Jan. 2022, Art. no. 810876.

[6] Y. Djenouri, A. Belhadi, A. Yazidi, G. Srivastava, P. Chatterjee, and J. C. Lin, "An intelligent collaborative image-sensing system for disease detection," *IEEE Sensors J.*, vol. 23, no. 2, pp. 947–954, Jan. 2023.

[7] V. Raman, K. ELKarazle, and P. Then, "Artificially generated facial images for gender classification using deep learning," *Comput. Syst. Sci. Eng.*, vol. 44, no. 2, pp. 1341–1355, 2023.

[8] Y. Yang, Z. Liu, M. Huang, Q. Zhu, and X. Zhao, "Automatic detection of multi-type defects on potatoes using multispectral imaging combined with a deep learning model," *J. Food Eng.*, vol. 336, Jan. 2023, Art. no. 111213.

[9] Y.-F. Wang, H.-M. Liu, and Z.-W. Fu, "Low-light image enhancement via the absorption light scattering model," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5679–5690, Nov. 2019.

[10] M. Purohit, A. Chakraborty, A. Kumar, and B. K. Kaushik, "Image processing framework for performance enhancement of low-light image sensors," *IEEE Sensors J.*, vol. 21, no. 6, pp. 8530–8542, Mar. 2021.

[11] S. M. Pizer et al., "Adaptive histogram equalization and its variations," *Comput. Vis., Graph., Image Process.*, vol. 39, no. 3, pp. 355–368, Sep. 1987.

[12] Z. Mahmood, T. Ali, S. Khattak, M. Aslam, and H. Mehmood, "A color image enhancement technique using multiscale retinex," in *Proc. 11th Int. Conf. Frontiers Inf. Technol.*, Islamabad, Pakistan, Dec. 2013, pp. 119–124.

[13] J. Ni, Y. Chen, Y. Chen, J. Zhu, D. Ali, and W. Cao, "A survey on theories and applications for self-driving cars based on deep learning methods," *Appl. Sci.*, vol. 10, no. 8, p. 2749, Apr. 2020.

[14] H. Zhang, Y. Zhang, L. Zhu, and W. Lin, "Deep learning-based perceptual video quality enhancement for 3D synthesized view," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 8, pp. 5080–5094, Aug. 2022.

[15] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, "Deep bilateral learning for real-time image enhancement," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–12, Aug. 2017.

[16] J. Chen, H.-T. Wu, L. Lu, X. Luo, and J. Hu, "Single underwater image haze removal with a learning-based approach to blurriness estimation," *J. Vis. Commun. Image Represent.*, vol. 89, Nov. 2022, Art. no. 103656.

[17] Y. Qu, Y. Ou, and R. Xiong, "Low light enhancement by unsupervised network," in *Proc. IEEE Int. Conf. Real-Time Comput. Robot. (RCAR)*, Asahikawa, Japan, Sep. 2020, pp. 404–409.

[18] Y. Shi, B. Wang, X. Wu, and M. Zhu, "Unsupervised low-light image enhancement by extracting structural similarity and color consistency," *IEEE Signal Process. Lett.*, vol. 29, pp. 997–1001, 2022.

[19] P. Wang, H. Zhu, H. Huang, H. Zhang, and N. Wang, "TMS-GAN: A twofold multi-scale generative adversarial network for single image dehazing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2760–2772, May 2022.

[20] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–9.

[21] D.-M. Tsai, S. S. Fan, and Y.-H. Chou, "Auto-annotated deep segmentation for surface defect detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.

[22] Y. Gao et al., "Wallpaper texture generation and style transfer based on multi-label semantics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1552–1563, Mar. 2022.

[23] X. Liu, H. Huang, and M. Qiu, "Intelligent algorithm for ceramic decorative pattern style transfer based on CycleGAN," in *Proc. 7th IEEE Int. Conf. Cyber Secur. Cloud Comput. (CSCloud), 6th IEEE Int. Conf. Edge Comput. Scalable Cloud (EdgeCom)*, New York, NY, USA, Aug. 2020, pp. 151–156.

[24] Y. Li, X. Nie, W. Diao, and S. Zheng, "Lifelong CycleGAN for continual multi-task image restoration," *Pattern Recognit. Lett.*, vol. 153, pp. 183–189, Jan. 2022.

[25] E. H. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, Dec. 1977.

[26] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.

[27] T. Celik and T. Tjahjadi, "Contextual and variational contrast enhancement," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3431–3441, Dec. 2011.

[28] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.

[29] X. Ren, M. Li, W.-H. Cheng, and J. Liu, "Joint enhancement and denoising method via sequential decomposition," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Florence, Italy, May 2018, pp. 1–5.

[30] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1301–1313, Apr. 2018.

[31] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.

[32] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Assoc. (BMVC)*, Newcastle, U.K., Sep. 2019, pp. 1–12.

[33] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. 27th ACM Int. Conf. Multimedia*, Nice, France, Oct. 2019, pp. 1632–1640.

[34] W. Yang, W. Wang, H. Huang, S. Wang, and J. Liu, "Sparse gradient regularized deep retinex network for robust low-light image enhancement," *IEEE Trans. Image Process.*, vol. 30, pp. 2072–2086, 2021.

[35] Y. Lu and S.-W. Jung, "Progressive joint low-light enhancement and noise removal for raw images," *IEEE Trans. Image Process.*, vol. 31, pp. 2390–2404, 2022.

[36] J. Li, J. Li, F. Fang, F. Li, and G. Zhang, "Luminance-aware pyramid network for low-light image enhancement," *IEEE Trans. Multimedia*, vol. 23, pp. 3153–3165, 2021.

[37] Y. Lu, Y. Guo, R. W. Liu, and W. Ren, "MTRBNet: Multi-branch topology residual block-based network for low-light enhancement," *IEEE Signal Process. Lett.*, vol. 29, pp. 1127–1131, 2022.

[38] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2242–2251.

[39] S. Kosugi and T. Yamasaki, "Unpaired image enhancement featuring reinforcement-learning-controlled image editing software," in *Proc. 34th AAAI Conf. Artif. Intell.*, New York, NY, USA, 2020, pp. 11296–11303.

[40] J. Liang, Y. Xu, Y. Quan, B. Shi, and H. Ji, "Self-supervised low-light image enhancement using discrepant untrained network priors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 7332–7345, Nov. 2022.

[41] D. Kwon, G. Kim, and J. Kwon, "DALE: Dark region-aware low-light image enhancement," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Manchester, U.K., 2020, pp. 1–12.

[42] Y. Jiang et al., "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.

[43] C. Guo et al., "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 1777–1786.
[44] C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4225–4238, Aug. 2022.
[45] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 5627–5636.
[46] R. Zhang, L. Guo, S. Huang, and B. Wen, "ReLLIE: Deep reinforcement learning for customized low-light image enhancement," in *Proc. 29th ACM Int. Conf. Multimedia*, China, Oct. 2021, pp. 2429–2437.
[47] S. Zheng and G. Gupta, "Semantic-guided zero-shot learning for low-light image/video enhancement," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Waikoloa, HI, USA, Jan. 2022, pp. 581–590.
[48] C.-M. Fan, T.-J. Liu, and K.-H. Liu, "Half wavelet attention on M-Net+ for low-light image enhancement," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Bordeaux, France, Oct. 2022, pp. 3878–3882.
[49] X. Lv, Y. Sun, J. Zhang, F. Jiang, and S. Zhang, "Low-light image enhancement via deep retinex decomposition and bilateral learning," *Signal Process., Image Commun.*, vol. 99, Nov. 2021, Art. no. 116466.
[50] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany, 2015, pp. 234–241.
[51] D. Jung, S. Yang, J. Choi, and C. Kim, "Arbitrary style transfer using graph instance normalization," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Abu Dhabi, United Arab Emirates, 2020, pp. 1596–1600.
[52] L. Xiao, B. Wu, and Y. Hu, "Surface defect detection using image pyramid," *IEEE Sensors J.*, vol. 20, no. 13, pp. 7181–7188, Jul. 2020.
[53] B. Mahaur, K. K. Mishra, and N. Singh, "Improved residual network based on norm-preservation for visual recognition," *Neural Netw.*, vol. 157, pp. 305–322, Jan. 2023.
[54] S. A. Rammy, W. Abbas, N.-U. Hassan, A. Raza, and W. Zhang, "CPGAN: Conditional patch-based generative adversarial network for retinal vessel segmentation," *IET Image Processing*, vol. 14, no. 6, pp. 1081–1090, 2020.
[55] 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.
[56] J. Lu, N. Li, S. Zhang, Z. Yu, H. Zheng, and B. Zheng, "Multi-scale adversarial network for underwater image restoration," *Opt. Laser Technol.*, vol. 110, pp. 105–113, Feb. 2019.
[57] B. Xu, X. Li, W. Hou, Y. Wang, and Y. Wei, "A similarity-based ranking method for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9585–9599, Nov. 2021.
[58] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.
[59] J. Liu, D. Xu, W. Yang, M. Fan, and H. Huang, "Benchmarking low-light image enhancement and beyond," *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 1153–1184, Apr. 2021.

Guangyi Tang received the B.S. degree from Hohai University, Changzhou, China, in 2019, where he is currently pursuing the Ph.D. degree in artificial intelligence with the School of Artificial Intelligence and Automation.

His research interests include artificial intelligence, robot control, and machine learning.

Jianjun Ni (Senior Member, IEEE) received the Ph.D. degree from the School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, China, in 2005.

He was a Visiting Professor with the Advanced Robotics and Intelligent Systems (ARIS) Laboratory, University of Guelph, Guelph, ON, Canada, from November 2009 to October 2010. He is currently a Professor at the School of Artificial Intelligence and Automation, Hohai University, Changzhou, China. He has published over 100 papers in related international conferences and journals. He serves as an Associate Editor and reviewer of a number of international journals. His research interests include control systems, neural networks, robotics, machine intelligence, and multi-agent system.

Yan Chen received the B.S. degree from Hohai University, Changzhou, China, in 2017, where he is currently pursuing the Ph.D. degree in IoT technology and application with the College of Information Science and Engineering.

His research interests include simultaneous localization and mapping, robot control, and machine learning.

Weidong Cao (Member, IEEE) received the Ph.D. degree in mechanical engineering from Chongqing University, Chongqing, China, in 2018.

He is currently working as a Lecturer at the School of Artificial Intelligence and Automation, Hohai University, Changzhou, China. His research interests include Swarm intelligence optimization algorithm, machine learning, and data-driven modeling.

Simon X. Yang (Senior Member, IEEE) received the B.Sc. degree in engineering physics from Beijing University, Beijing, China, in 1987, the first of two M.Sc. degrees in biophysics from the Chinese Academy of Sciences, Beijing, in 1990, the second M.Sc. degree in electrical engineering from the University of Houston, Houston, TX, USA, in 1996, and the Ph.D. degree in electrical and computer engineering from the University of Alberta, Edmonton, AB, Canada, in 1999.

He is currently a Professor and the Head of the Advanced Robotics and Intelligent Systems Laboratory, University of Guelph, Guelph, ON, Canada. His research interests include robotics, intelligent systems, sensors and multi-sensor fusion, wireless sensor networks, control systems, transportation, and computational neuroscience.

Dr. Yang has been very active in professional activities. He was the General Chair of the 2011 IEEE International Conference on Logistics and Automation, and the Program Chair of the 2015 IEEE International Conference on Information and Automation. He serves as the Editor-in-Chief of the *International Journal of Robotics and Automation*, and an Associate Editor of the IEEE TRANSACTIONS ON CYBERNETICS, and several other journals.

Authorized licensed use limited to: Khon Kaen University provided by UniNet. Downloaded on July 22,2025 at 10:42:29 UTC from IEEE Xplore. Restrictions apply.