Research on Low-Light Image Enhancement Algorithm Based on Generative Adversarial Networks

Chengzhong Kuang
Nanning Power Supply Bureau of Guangxi Power Grid Co., Ltd
Nanning, China
290660483@qq.com

Shifeng Ou*
Electric Power Science Research Institute of Guangxi Power Grid Co., Ltd.
Nanning, China
84899941@qq.com
*Corresponding author

Xiaoxuan Lin
Nanning Power Supply Bureau of Guangxi Power Grid Co., Ltd
Nanning, China
lxx704335435@qq.com

Chunyu Zhong
Nanning Power Supply Bureau of Guangxi Power Grid Co., Ltd
Nanning, China
455427540@qq.com

2025 10th International Conference on Computer and Communication System (ICCCS) | 979-8-3315-2314-5/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICCCS65393.2025.11069934

***Abstract*—Low-light image enhancement has long been a hot research topic in the field of image processing. Under low-light conditions, issues such as poor image quality and loss of details are prominent. To improve the quality of low-light images, an improved weak-light image enhancement method based on CycleGAN is proposed. In this method, the EMA attention mechanism is integrated into the residual network to optimize the generator, enhancing the model's feature extraction capability and thus improving the enhancement effect. Additionally, the ASPP module is employed to strengthen the network's multi-scale feature extraction ability, expand the receptive field, and improve the model's generalization performance. Experimental results show that the proposed method achieves good performance on low-light image datasets.**

Keywords—ASPP module, EMA attention mechanism, low-light image enhancement, CycleGAN

I. INTRODUCTION

Images from low-light images is of great research value. With the development of deep learning, methods based on deep learning have become the mainstream approach for low-light enhancement. Long et al.[1] used a Generative Adversarial Network (GAN) approach for training, focusing on color information and brightness adjustment in images, aiming to improve low-light image quality by enhancing image details and color information. The method effectively improves the quality of low-light images by recovering details through adversarial training and residual modules, and ensures natural image colors by adding color loss. Guo et al.[2] used deep neural networks to learn the brightness curve of low-light images and enhance their quality through brightness curve estimation. The method does not require paired data for training and uses contrastive loss to optimize the network, effectively improving the brightness, details, and color of low-light images. Jiang et al.[3] used an unpaired supervised method for low-light image enhancement, effectively improving the brightness, details, and color of low-light images through adversarial and self-supervised learning. Although the above methods address the problem of low-light image enhancement to varying degrees, they still have their respective limitations.

To address issues such as detail blurring, color distortion, and insufficient contrast that may occur in traditional CycleGAN [9] for low-light image enhancement tasks, this paper proposes an improved CycleGAN network based on feature fusion for low-light image enhancement. The network can effectively extract and fuse multi-scale feature information from low-light images and introduces an efficient attention mechanism in the residual network to enhance the quality of the generated images. This section first introduces the proposed improved CycleGAN network architecture, followed by a detailed description of the Atrous Spatial Pyramid Pooling (ASPP)[4] module and the improved attention mechanism (EMA)[5] module in the generator network, and concludes with an introduction to the loss functions used in this paper.

II. BASIC THEORY

A. Cycle Generative Adversarial Network

The CycleGAN architecture proposed in this paper is shown in Figure 1, where the generator is used to generate the corresponding normal-light image from a low-light image. This architecture is designed with two mutually symmetric generative adversarial networks.

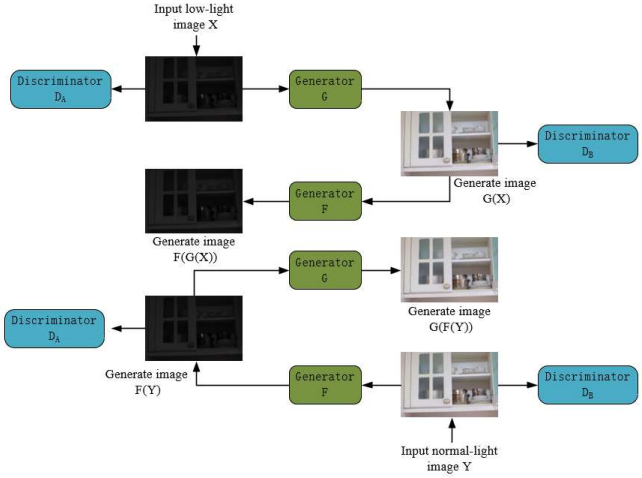

Fig. 1. CycleGAN network architecture.

In the diagram, the generators G and F serve to convert low-light images into normal-light images and convert normal-light images into low-light images, respectively. The role of the discriminators is to determine whether the input image is a real image or an image generated by the generator. As shown in the diagram, the generators G and F convert the low-light image X and normal-light image Y into the normal-light image G(X) and low-light image F(Y), respectively. The generated images G(X) and F(Y) are then fed into discriminators D_A and D_B to determine whether the input to the generator is a real image or a generated image. To ensure that the generated images are as close as possible to the distribution of real images, adversarial loss is computed to optimize the generators. Subsequently, the generated images G(X) and F(Y) are re-input into generators F and G, respectively, to generate the corresponding low-light image F(G(X)) and normal-light image G(F(Y)). This achieves a bi-directional mapping from low-light to normal-light and from normal-light to visible light. The cyclic consistency loss between the input image and the final generated image is calculated to constrain the network, ensuring that the image retains information from its original state after two conversions.

B. EMA Residual Block Network

In CycleGAN [6], the task of the generator is to convert one type of image into another, such as converting low-light images into normal-light images. To extract and transform the features of the input image, the generator typically uses a Residual Network (ResNet) [7]. As the number of layers in the network increases, the computational demand of the model also increases, and the network is prone to overfitting. Residual blocks effectively address the vanishing gradient problem in deep networks through skip connections, preventing the network from overfitting. However, in the task of low-light image enhancement, low-light images often suffer from issues like detail loss and increased noise. To focus more on the important features in the image and improve the enhancement effect, this paper introduces the EMA attention mechanism into the residual module. The EMA attention mechanism dynamically assigns weights based on the importance of the targets, allowing the network to pay more attention to key regions and details in the image during the enhancement process. The network is shown in Figure 2.

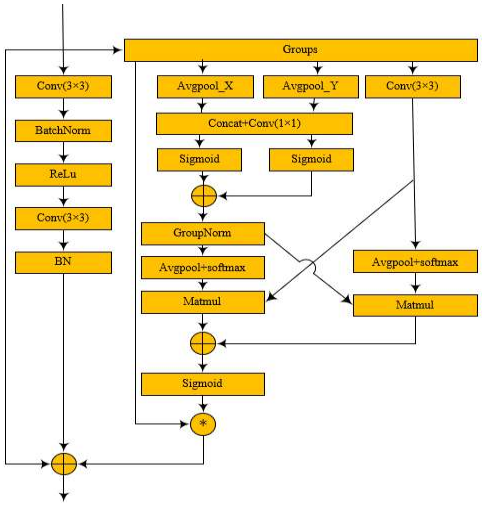

Fig. 2. EMA_block Module.

The Residual Block module consists of convolutional layers, normalization layers, and activation layers (ReLU). The EMA attention mechanism performs adaptive channel feature weighting through multi-stage operations (grouping, pooling, convolution, weighting calculation, Softmax, etc.). By adding the feature map output from the residual network with the EMA (Efficient Multi-head Attention) attention mechanism, this design retains the original information while enhancing the network's sensitivity to specific regions or details, thus effectively improving the accuracy of image generation.

C. ASPP Module

To address the limitation of the encoder, which can only use regular convolution operations, restricting the receptive field, especially when processing large-scale image features, it is unable to comprehensively analyze the image from a multi-scale perspective. This may cause the loss of important details and prevent the effective capture of global information, resulting in limited feature representation ability, which in turn affects the quality of the final generated image. This paper uses the ASPP (Atrous Spatial Pyramid Pooling) [8]module to solve this problem by extracting features at multiple scales through different dilated convolutions, effectively expanding the receptive field of the network. The network can not only focus on local detail information but also capture global information from a broader region. The network is shown in Figure 3.

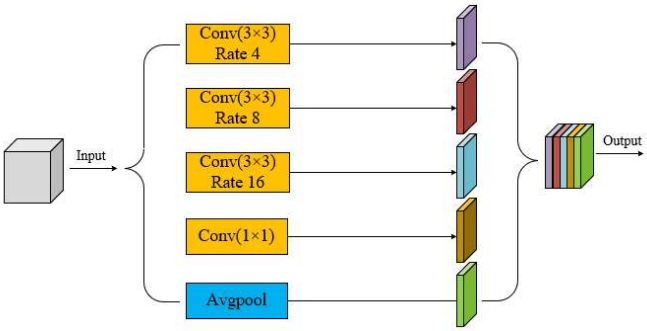

Fig. 3. ASPP Module.

The core idea of the Atrous Spatial Pyramid Pooling (ASPP) module is to expand the receptive field through atrous convolution. Its basic structure includes a 1×1 convolution branch, a global average pooling branch, and three parallel atrous convolution branches. The dilation rates of the three convolution branches are 4, 8, and 16, respectively.

III. Network Structure

A. Generator Network

The low-light image enhancement model consists of two generators and two discriminators. Generator G generates the corresponding normal-light image from the input low-light image, while generator F is responsible for converting the normal-light image back to the low-light image to ensure cycle consistency. Discriminators DA and DB are used to distinguish between real and generated low-light images and normal-light images, respectively, thereby enhancing the realism of the generated images. The generator consists of an encoder, transformer, and decoder. The encoder is composed of multiple convolutional sequences and downsampling layers, responsible

for extracting high-level features from the input image and reducing the spatial resolution. At the end of the encoder, ASPP is introduced. The addition of ASPP allows multi-scale feature extraction at various receptive field scales, enhancing the global receptive field and improving the network's ability to capture multi-scale features, thus expanding the receptive field and enhancing the model's generalization ability. The transformer consists of multiple residual blocks that perform mapping between different lighting conditions, learning feature transformations from low-light to normal-light images. Based on the original residual block, the EMA attention mechanism is introduced to form the residual attention module. The addition of the EMA attention mechanism dynamically adjusts the importance of different features, enhancing the model's focus on key information and making CycleGAN more accurate when learning style transformations. The decoder uses multiple upsampling convolutional sequences and employs transposed convolutions to upsample the image and generate the target domain image. The generator structure is shown in Figure 4.

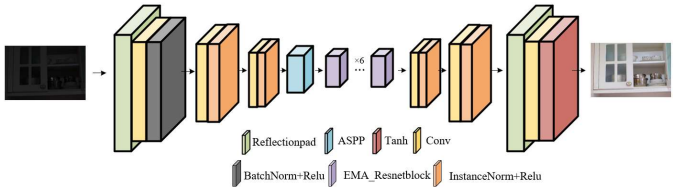

Fig. 4. Generator Network.

B. Discriminator network

The discriminator is used to distinguish whether the image is a real image or a fake image generated by the generator. During the training process, the discriminator receives real images and fake images generated by the generator, and determines whether the input image is real data. It continuously updates its parameters through training to improve its ability to classify real and fake images. The discriminator uses a fully convolutional neural network (PatchGAN). Unlike traditional binary classifiers that output a single probability, PatchGAN's core idea is to perform classification based on local regions of the image rather than the entire image. The output is a decision matrix, which is an N×N matrix where each element represents the classification result of the corresponding local region.

C. Loss Function

In CycleGAN, several different loss functions are used to train the generator and discriminator, ensuring that the generated images both realistically mimic the target images and preserve the structure and content of the input images. The loss functions in CycleGAN can be generally divided into three parts: adversarial loss L_{GAN} , identity loss L_{identity} , and cycle consistency loss L_{cyc} .

The total loss function of the network can be expressed as:

$$L_{\text{total}} = \lambda_{\text{identity}} L_{\text{identity}} + L_{\text{GAN}} + \lambda_{\text{cycle}} L_{\text{cycle}} \tag{1}$$

The adversarial loss can be expressed as:

$$L_{\text{GAN}}(G, D_Y, X, Y) = \mathrm{E}_{y \sim P_{\text{dan}}(y)} \ln D_B(y) + \mathrm{E}_{x \sim P_{\text{dan}}(x)} [\ln(1 - D_B(x))] \tag{2}$$

The cycle consistency loss can be expressed as:

$$L_{\text{cyc}}(G, F) = \mathrm{E}_{x \to P_{\text{dan}}(x)} [\| F(G(x)) - x \|_1] + \mathrm{E}_{y \to P_{\text{dan}}(y)} [\| G(F(y)) - y \|_1] \tag{3}$$

The identity loss can be expressed as:

$$L_{\text{identity}}(G, F) = \mathrm{E}_{y \to P_{\text{dan}}(y)} [\| G(y) - y \|_1] + \mathrm{E}_{y \to P_{\text{dan}}(y)} [\| F(x) - x \|_1] \tag{4}$$

In the formula, G and F are the generators, D_A and D_B are the discriminators. $\mathrm{E}_{y \sim P_{\text{dan}}(y)}$ and $\mathrm{E}_{x \sim P_{\text{data}}(x)}$ represent the mathematical expectations of the real data for the input images x and y in the data space, respectively. $\lambda_{\text{identity}}$ and λ_{cycle} are the weight coefficients for the cycle consistency loss and identity loss, respectively.

IV. Experimental Results

To verify the performance of the proposed network model, experiments are conducted in this section. The experiments are carried out on an Ubuntu 20.04 operating system for training and testing, using the PyTorch deep learning framework. The specific experimental environment configuration is shown in Table 1. The batch size is set to 8, the number of epochs is set to 200, and the initial learning rate is set to 0.0001.

Table 1. Experimental Environment Configuration

Configuration Name	Configuration Parameters
Python	python3.9
GPU	NVIDIA GeForce RTX4090
GPU Memory	24G
Deep Learning Framework	PyTorch1.13.0
CUDA	NVIDIA CUDA 11.6

This section selects the LOL dataset for training, which contains 485 pairs of low-light images and their corresponding normal-light images. The proposed method is tested on three test sets: 15 pairs from the LOL dataset, 100 pairs from the LOLV2_real dataset, and 30 randomly selected pairs from the LSRW dataset. Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Learned Perceptual Image Patch Similarity (LPIPS), and No-Reference Image Quality Evaluation (NIQE) are chosen as evaluation metrics. PSNR and SSIM are commonly used objective metrics for image enhancement. PSNR measures the signal-to-noise ratio between the enhanced image and the real image. A higher PSNR value indicates better image clarity, detail preservation, and higher image quality. SSIM evaluates the similarity between images in terms of brightness, contrast, and structure. A higher SSIM value means the image is more similar to the real image and better matches human visual perception. LPIPS

is a deep learning-based no-reference image quality assessment metric that better reflects human visual perception of image quality. A lower LPIPS value indicates higher perceptual similarity between the images. NIQE is a no-reference image quality assessment metric that evaluates image quality by simulating human visual perception. A lower NIQE value indicates higher image quality and more natural perception.

A. Comparative Experiments

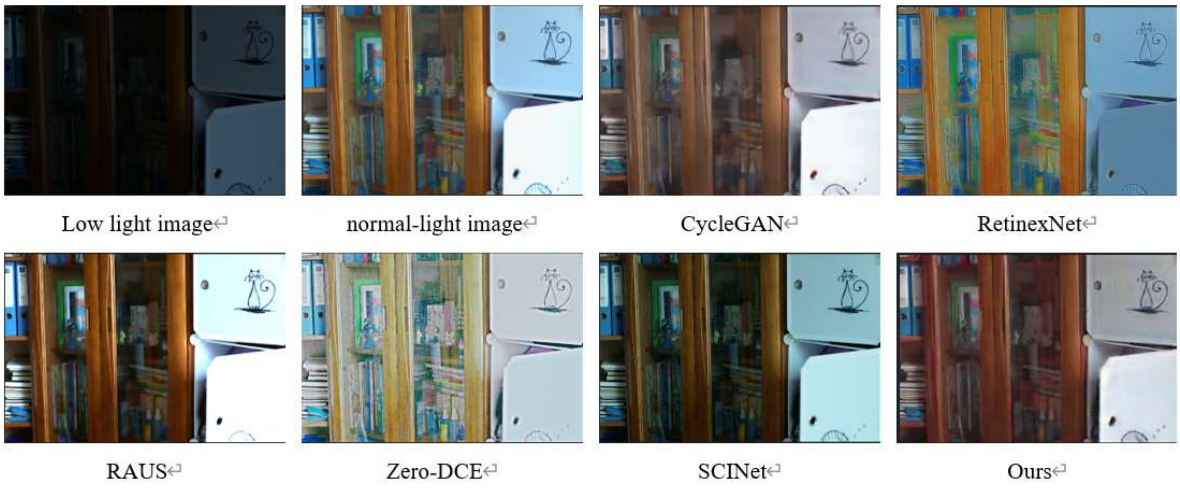

Fig. 5. Test Results of the LOL Dataset.

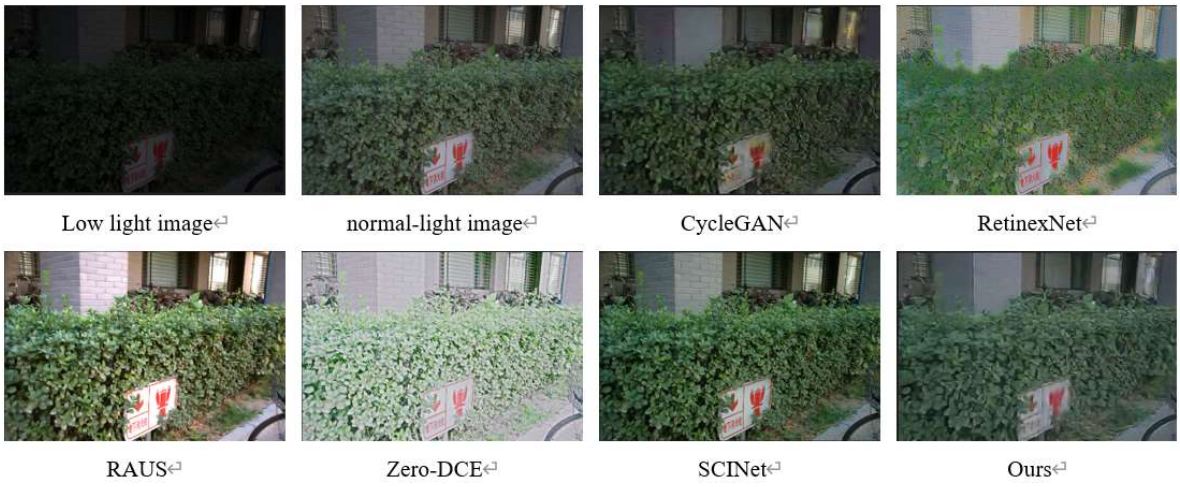

Fig. 6. Test Results of the LOLV2_real Dataset.

Fig. 7. Test Results of the LSRW Dataset.

The low-light image test results on the LOL, LSRW, and LOLV2_real datasets are shown in Figures 5-7. According to the test results, Zero-DCE produces severe color distortion. SCI demonstrates good enhancement effects, but the enhanced images are overall too dark. RetinexNet exhibits significant noise and color distortion, with insufficient handling of textures, edges, and other details. In contrast, the proposed algorithm effectively enhances low-light images and preserves the original image's color and details better.

Table 2. PSNR, SSIM, LPIPS, and NIQE Results of Different Methods on the LOL Test Set

Method	PSNR(db)↑	SSIM↑	LPIPS↓	NIQE↓
Zero-DCE[2]	14.56	0.497	0.452	7.907
RetinexNet[10]	17.72	0.671	0.359	7.1279
RUAS[11]	16.34	0.503	**0.270**	6.2557
SCINet[1]	13.77	0.511	0.346	7.7239
CycleGAN[6]	18.93	0.696	0.387	7.2935
Ours	**20.26**	**0.712**	0.352	**7.0241**

Table 2 presents the results of the proposed method and five other methods on the LOL test set, where "↑" indicates that a higher value is better, and "↓" indicates that a lower value is better. From the table, it can be seen that, compared to the original algorithm, the PSNR has increased by 1.33 dB, SSIM has improved by 1.6%, and the proposed method also outperforms the original method in the NIQE metric.

Table 3. PSNR, SSIM, LPIPS, and NIQE Results of Different Methods on the LSRW Test Set

Method	PSNR(db)↑	SSIM↑	LPIPS↓	NIQE↓
Zero-DCE[2]	13.59	0.380	0.491	3.6176
RetinexNet[10]	17.29	0.485	0.422	5.0383
RUAS[11]	15.68	0.491	0.525	5.983
SCINet[1]	14.77	0.405	0.384	3.8947
CycleGAN[6]	15.78	0.519	0.365	4.7158
Ours	17.71	0.544	0.338	4.6469

Table 4. PSNR, SSIM, LPIPS, and NIQE Results of Different Methods on the LOLV2_real Test Set

Method	PSNR(db)↑	SSIM↑	LPIPS↓	NIQE↓
Zero-DCE[2]	12.71	0.446	0.497	8.3085
RetinexNet[10]	17.84	0.684	0.39	7.9897
RUAS[11]	15.34	0.494	0.307	6.4745
SCINet[1]	16.73	0.532	0.309	7.9281
CycleGAN[6]	21.8	0.683	0.463	7.5253
Ours	**23.04**	**0.717**	0.429	7.3664

Tables 3 and 4 present the results of the proposed method and five other methods on the LSRW test set and the LOLV2_real test set, respectively. The proposed method outperforms the other methods in both PSNR and SSIM metrics on these test sets, demonstrating its excellent performance in terms of contrast and structural information.

B. Ablation Experiments

To better explain the effectiveness of the network improvement scheme, this section also analyzes the results of the improvements for different modules and modifications. Ablation experiments for different modules are conducted, as shown in Table 5. The modules involved in the experiment include the EMA_block and ASPP modules. CycleGAN-EMA refers to the model with the EMA_block module added, while CycleGAN-ASPP refers to the model with the ASPP module added. Ablation experiments are performed on the LOL test set, and the results are evaluated using PSNR and SSIM.

Table 5. Ablation Experiment Results

Model	EMA	ASPP	PSNR(db)	SSIM
CycleGAN[6]			18.93	0.696
CycleGAN-EMA	√		19.82	0.707
CycleGAN-ASPP		√	19.96	0.695
Ours	√	√	20.26	0.712

As shown in the table, compared to the proposed method, the PSNR and SSIM metrics decrease after removing the ASPP module. Similarly, when the EMA_block module is removed, both the PSNR and SSIM metrics are lower than those of the proposed method. In conclusion, this demonstrates the effectiveness of the proposed improvement scheme, which enhances the low-light image enhancement performance.

V. CONCLUSION

This paper uses a CycleGAN to enhance low-light images. The method introduces the EMA attention mechanism into the generator, allowing the network to adaptively focus on the key features of low-light images, improving the restoration of image details. By adding the ASPP module at the end of the encoder in the generator, the network's multi-scale feature extraction ability is enhanced, expanding the receptive field and improving the model's generalization performance. Finally, in the experiments, the improved model is compared with other models, and the results show that the improved model performs better in terms of detail recovery, color restoration, and brightness enhancement.

REFERENCES

[1] Ma L, Ma T, Liu R, et al. Toward fast, flexible, and robust low-light image enhancement[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 5637-5646.

[2] Guo C, Li C, Guo J, et al. Zero-reference deep curve estimation for low-light image enhancement[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1780-1789.

[3] Jiang Y, Gong X, Liu D, et al. Enlightengan: Deep light enhancement without paired supervision[J]. IEEE transactions on image processing, 2021, 30: 2340-2349.

[4] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation. arXiv[J]. arXiv preprint arXiv:1706.05587, 2017, 5.

[5] Ouyang D, He S, Zhang G, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023-2023 IEEE International

Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[6] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.

[7] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[8] Wang P, Chen P, Yuan Y, et al. Understanding convolution for semantic segmentation[C]//2018 IEEE winter conference on applications of computer vision (WACV). Ieee, 2018: 1451-1460.

[9] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.

[10] Wei C, Wang W, Yang W, et al. Deep retinex decomposition for low-light enhancement. arXiv 2018[J]. arXiv preprint arXiv:1808.04560.

[11] Liu R, Ma L, Zhang J, et al. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 10561-10570.

[12] Fan C, Wang P, Ma H, et al. Performance degradation assessment of rolling bearing cage failure based on enhanced CycleGAN[J]. Expert Systems with Applications, 2024, 255: 124697.

[13] Mahmud S, Chowdhury M E H, Kiranyaz S, et al. Restoration of motion-corrupted EEG signals using attention-guided operational CycleGAN[J]. Engineering Applications of Artificial Intelligence, 2024, 128: 107514.

[14] Huang F, Tang X, Li C, et al. Cyclic style generative adversarial network for near infrared and visible light face recognition[J]. Applied Soft Computing, 2024, 150: 111096.