2025 8th International Conference on Advanced Algorithms and Control Engineering (ICAACE) | 979-8-3315-3508-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICAACE65325.2025.11019998

Low-light Gastroscopic Image Enhancement Network Based on CycleGAN

Runjiang Mao
School of computer Science and Artificial Intelligence
Zhengzhou University
Zhengzhou, China
1756311295@qq.com

Rui Zhao
Endoscopy Center
Peking University People's Hospital
Beijing, China
zhaorui_001@126.com

Liming Zhang*
Endoscopy Center
Peking University People's Hospital
Beijing, China
* Corresponding author: zhlim315@163.com

***Abstract*—Aiming at problems such as uneven illumination and low image contrast in gastroscopic images, an unsupervised low-light gastroscopic image enhancement network based on CycleGAN is proposed. Firstly, a down-sampling module based on wavelet transform is designed to reduce the loss of detail information during the generator's down-sampling process. Secondly, a dynamic feature fusion module is used as a skip connection structure to enhance the fusion of multi-scale features. Finally, an illumination-aware attention module is introduced to improve the accuracy of illumination estimation and suppress local overexposure. This method was compared with RRDNet, EnlightenGAN, LE_GAN, and SCI. It showed excellent performance in both the Natural Image Quality Evaluator (NIQE) and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) metrics. Overall, this method can effectively enhance the brightness of gastroscopic images while preserving texture information, optimizing the visual recognition of mucosal vessel textures and early lesions, which helps improve the accuracy of clinical diagnoses and holds significant clinical application value.**

Keywords-Gastrointestinal endoscopy; Image Enhancement; Low light; CycleGAN

I. INTRODUCTION

Gastroscopy, as a core diagnostic technique for digestive tract diseases, holds irreplaceable clinical value in lesion screening, minimally invasive surgery navigation, and dynamic monitoring of malignant tumors. However, due to factors such as the complex structure of the gastrointestinal tract, the optical characteristics of endoscopic illumination systems, and the sensitivity of image sensors, gastroscopic images often suffer from uneven illumination and inadequate contrast. These issues result in blurred mucosal textures and weakened lesion boundaries. By employing illumination enhancement techniques to improve the quality of gastroscopic images, not only can the visual identification of fine mucosal structures and early lesions be enhanced, but the contrast between vascular textures and lesion boundaries can also be optimized. This improvement enhances the detection sensitivity of endoscopists to precancerous lesions. The enhanced high-quality image sequences can provide high signal-to-noise ratio input data for computer-aided diagnostic systems (such as deep learning-based lesion recognition and segmentation models), thereby promoting the development of digestive disease diagnosis and treatment towards more intelligent and precise directions.

Existing low-light enhancement algorithms are primarily designed for natural images and are not well-suited for medical images. Gastroscopic images have distinct characteristics compared to natural images, featuring rich mucosal textures and vascular distributions with fine structural details, but relatively sparse overall semantic information. This characteristic necessitates that enhancement algorithms focus particularly on the preservation and enhancement of local texture features. In this context, this paper proposes a low-light gastroscopic image enhancement network based on CycleGAN[1]. The proposed network includes improvements to the generator's downsampling module and skip connection module. Additionally, an illumination-aware attention module is introduced to reduce information loss during feature extraction and suppress overexposure in local regions. These enhancements effectively improve the visual quality of low-light gastroscopic images.

II. METHODS

The proposed model is an improvement on the traditional CycleGAN, employing a dual-generator and dual-discriminator structure to form a bidirectional cycle network where Generator G learns the enhancement transformation from low-light gastroscopic images to normal-light gastroscopic images and Generator F learns the inverse transformation, converting normal-light gastroscopic images back to low-light images, while Discriminator D_L is responsible for distinguishing between real and generated low-light gastroscopic images and Discriminator D_N is responsible for distinguishing between real and generated normal-light gastroscopic images, and through the adversarial process between the generators and discriminators, the model learns the true data distribution and generates new data, with the introduction of cycle consistency ensuring the stability of the transformations, and the model only requires an unpaired dataset of low-light and normal-light gastroscopic images, leveraging adversarial learning to

automatically establish cross-domain mappings, effectively overcoming the challenge of scarce paired medical image data.

A. Generator Network Architecture

The generators are designed based on an improved U-Net[2] architecture, with the overall framework shown in Figure 1, where both the encoder and decoder use three Conv blocks, and the encoder incorporates wavelet transform downsampling between Conv blocks instead of conventional pooling to reduce feature map resolution while retaining high-frequency texture information; the skip connection structure between the encoder and decoder utilizes a Dynamic Feature Fusion (DFF) module to achieve effective propagation and aggregation of multi-scale features through cross-level feature interaction, and an illumination-aware module is introduced, employing an attention mechanism to dynamically enhance feature responses in low-light regions.

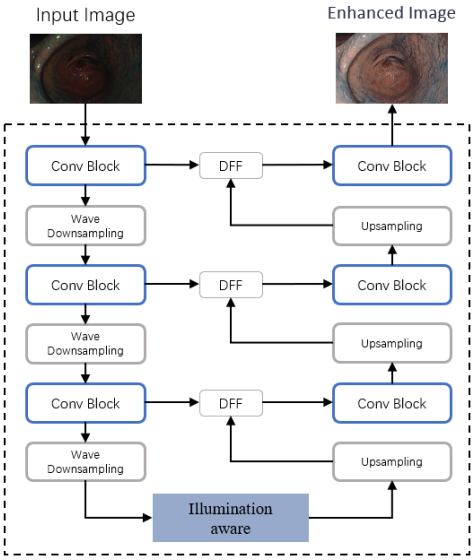


Figure 1. Generator Network Architecture

B. Wavelet Transform Downsanpling

Wavelet transform can perform multi-resolution decomposition on image signals and capture both the low-frequency and high-frequency signals of the image[3].The wavelet transform downsampling module structure, as shown in Figure 2, performs convolution operations on each row of the input feature (H×W×C) with both a low-pass filter H_0 and a high-pass filter H_1, followed by stride-2 downsampling to reduce redundancy, generating intermediate low-frequency matrix D and high-frequency matrix A. Each column of D and A is again subjected to the same low-pass filtering, high-pass filtering, and downsampling, ultimately producing four sub-bands: LL encoding the overall contour and smooth regions of the image, LH capturing horizontal directional detail features, HL extracting vertical directional high-frequency information, and HH capturing diagonal texture changes; the size of these four sub-bands is H/2×W/2×C, and they are concatenated along the channel dimension, expanding the number of channels in the output feature map to four times that of the input feature. The concatenated features are then fed into an ECANet[4] module to adaptively learn the weights of each channel, suppressing noise components in the high-frequency sub-bands (LH, HL, HH) while retaining the image's detail features. Following this, the features pass through a convolution block to adjust the number of channels in the feature map, thereby filtering out redundant information and enabling subsequent modules to learn more representative features.

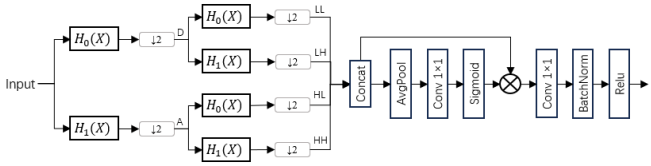


Figure 2. Wavelet Transform Downsampling Model

C. Dynamic Feature Fusion

The dynamic feature fusion module, as shown in Figure 3, processes features F_1 from the encoder and F_2 from the decoder by concatenating them along the channel dimension to generate a composite feature F. This composite feature F is then fed into an ECANet[4] module to generate channel-wise weights, which help to emphasize informative feature channels while suppressing redundant ones. Separately, F1 and F2 are each passed through a 1×1 convolution layer, and the resulting outputs are element-wise added together, followed by applying a Sigmoid activation function to generate a spatial attention weight map. The channel attention weights and spatial attention weights are then jointly applied to the composite feature F, achieving multi-dimensional adaptive fusion of cross-scale features. This process enhances the completeness and semantic consistency of feature representation while maintaining computational efficiency.

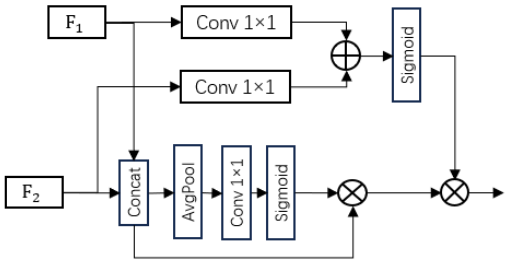


Figure 3. Dynamic Feature Fusion Model

D. Illumination Aware

The illumination-aware attention module structure, as shown in Figure 4, includes both local and global illumination attention branches; the local illumination attention branch ensures more accurate brightness estimation for adjacent pixels, where the input feature is fed in parallel into three 1×1 convolution layers to generate three intermediate features f, g, and h, the dot product of f and g is computed and then passed through a SoftMax function to generate weights for h, these weights are applied to h to produce a local illumination attention feature map, which is element-wise added to the input feature to obtain the output of the local illumination attention branch; in the global illumination attention branch, an adaptive channel attention mechanism is employed to help the network

selectively enhance informative semantic responses while suppressing irrelevant ones, processing the input feature with global average pooling along the channel dimension, followed by two fully connected layers for feature dimensionality reduction and expansion, generating weights for each channel, these weights are applied to the input feature to obtain the output of the global illumination attention branch, finally, the output features from the local illumination attention branch and the global illumination attention branch are element-wise added together and fed into the decoder part, enhancing the accuracy and realism of the generated images while maintaining computational efficiency.

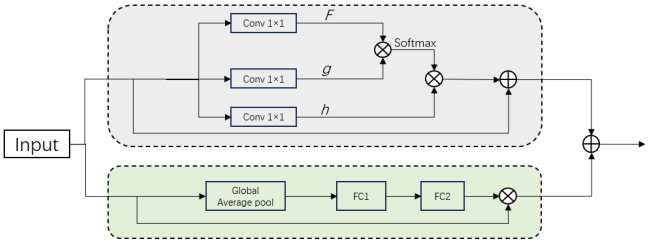


Figure 4. Illumination Aware model

E. Discriminators

The role of the discriminators is to distinguish between the data samples generated by the generators and the real samples from the training set, outputting the probability that a given sample is a real sample. In this work, both discriminators utilize a PatchGAN[5] network. For an input image of size 256×256, the discriminators output a 30×30 matrix, where each element represents the probability that a fixed-size local region of the input image is a real image. The overall authenticity score is then calculated based on the probabilities of these local regions.

F. Loss function

The loss function of the model in this paper includes adversarial loss, cycle consistency loss, and feature Identity loss invariant[6].

$$Loss = L_{adv} + \lambda_1 L_{cyc} + \lambda_2 L_{Identity} \quad (1)$$

III. Expersement Results

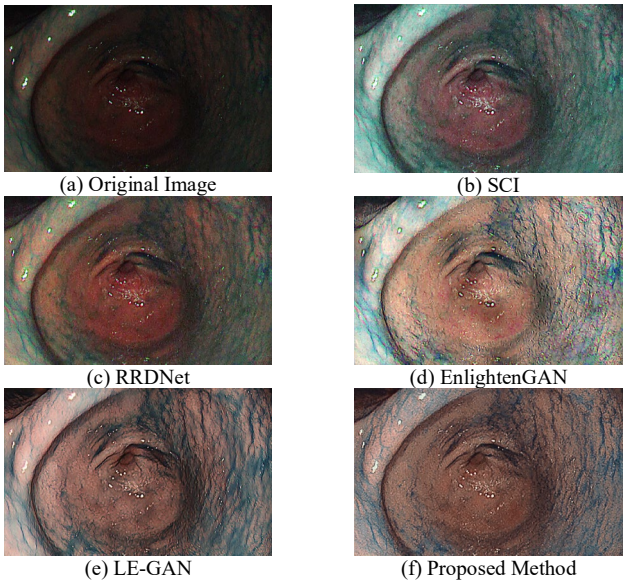
To validate the effectiveness of the proposed algorithm, we compared it with existing low-light image enhancement algorithms. Since our method is a reference-free low-light enhancement algorithm, we selected four other reference-free algorithms for comparison: RRDNet[7], SCI[8], EnlightenGAN[9], and LE-GAN. Among these, RRDNet is based on Retinex theory, while EnlightenGAN and LE-GAN are end-to-end image generation networks. To ensure the fairness of the comparative experiments, all compared algorithms used their default parameter settings and the model weights provided in their original papers. This rigorous comparison allows us to objectively assess the performance and advantages of our proposed method in enhancing low-light gastroscopic images.

A. Datasets

The dataset used in this study was sourced from the clinical database of the Endoscopy Center at Peking University People's Hospital. Low-quality images caused by factors such as gastric fluid obstruction, motion artifacts, or defocusing were excluded. A dataset comprising 600 low-light gastroscopic images and 600 standard-light gastroscopic images was constructed. All images were cropped and resized to 600×400 pixels.

B. Experimental results

The enhancement effects of gastroscopic images are shown in Figure 5.



(a) Original Image (b) SCI (c) RRDNet (d) EnlightenGAN (e) LE-GAN (f) Proposed Method

Figure 5. comparison of enhancement effects

It can be observed that methods such as SCI, RRDNet, EnlightenGAN, and LE-GAN have varying degrees of shortcomings. The images generated by the SCI algorithm exhibit significant color deviations, with an overall grayish hue and a substantial amount of noise. The images produced by the RRDNet algorithm appear relatively natural but suffer from insufficient brightness enhancement. Additionally, the texture details are weakened, resulting in lower image contrast. The enhancement results of EnlightenGAN show severe local overexposure, with numerous white patches in the images, which negatively impact visual quality. The images enhanced by the LE-GAN suffer from excessive color saturation, leading to an uneven surface appearance on the stomach walls. Comparing the top-left corners of the images, it is evident that the other four algorithms result in varying degrees of loss of edge detail information. In contrast, the proposed method retains more of the fine vascular textures on the stomach wall surface. Compared to these methods, the proposed approach performs well in enhancing the illumination of gastroscopic images. The enhanced images have moderate brightness, natural colors, and no noticeable local overexposure. While maintaining a natural appearance, the method effectively preserves detailed image information, demonstrating its effectiveness.

To further objectively evaluate the enhancement effects of different algorithms, we selected several images from the low-light gastroscopic image dataset for testing. The effectiveness of the algorithms was assessed by comparing the average NIQE[10] and BRISQUE[11] scores. The comparison results are shown in Table 1.

TABLE I. COMPARISON OF IMAGE ENHANCEMENT RESULTS

Algorithm	*NIQE*	*BRISQUE*
SCI	5.255	59.515
RRDNet	5.316	56.341
EnlightenGAN	5.853	55.972
LE-GAN	4.819	52.662
Proposed Method	4.832	51.243

From the table, it can be seen that: LE-GAN performed best in terms of the NIQE metric, with an average NIQE score of 4.819. The proposed method second in the NIQE metric, with an average NIQE score of 4.832. In terms of the BRISQUE metric, the proposed method performed the best, achieving an average BRISQUE score of 51.243. LE-GAN followed closely behind in the BRISQUE metric. Upon comprehensive comparison and analysis, the enhanced gastroscopic images produced by the proposed algorithm exhibit less distortion and higher naturalness, making them closer to real gastroscopic images. Both in terms of visual effect comparisons and objective metric analyses, the proposed method demonstrates significant advantages.

IV. CONCLUSIONS

Aiming at the issues of local overexposure and loss of detailed texture information during the enhancement of gastroscopic images, this paper proposes a low-light gastroscopic image enhancement network based on an improved CycleGAN. The proposed method replaces traditional pooling methods with a wavelet transform down-sampling module to reduce the loss of image information. It employs dynamic feature fusion in skip connections to enhance the effective integration of shallow and deep network features, promoting feature propagation throughout the network. Additionally, an illumination-aware attention module is introduced to generate illumination weights for features, suppressing local overexposure and making the overall image illumination more natural. Experimental results show that the proposed algorithm effectively enhances the brightness of gastroscopic images while ensuring clearer details and excellent visual effects, performing well in metrics such as NIQE and BRISQUE .

The proposed method can effectively enhance the brightness of gastroscopic images even in the absence of paired datasets, demonstrating broad application prospects in clinical practice. In the future, this technology is expected to be applied to the illumination enhancement of other endoscopic images, such as colonoscopy and laparoscopy.

FUNDING INFORMATION

This work was supported by the Peking university people's hospital scientific research development fund (RDL2024-04) (RDL2023-13)

REFERENCES

[1] ZHU J, 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. Venice, Italy: IEEE, 2017: 2223-2232.

[2] Chen X ,Tang X ,Xiong J , et al.Pore characterization was achieved based on the improved U-net deep learning network model and scanning electron microscope images[J].Petroleum Science and Technology,2025,43(7):715-729.

[3] Xu Guoping ,Liao Wentao ,Zhane Xuan ,et al. Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation[J].Pattern Recognition,2023,143.

[4] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle,USA:IEEE, 2020: 11534 11542.

[5] HEPBURN A, LAPARRA V, MCCONVILLE R, et al. Enforcing perceptual consistency on generative adversarial networks by using the normalised laplacian pyramid distance[EB/OL].(2020-06)[2024-11-26].

[6] Fu Ying, Hong Yang, Chen Linwei, et al. LE-GAN: unsupervised low light image enhancement network using attention module and identity invariant loss [J]. Knowledge-Based Systems, 2022, 240: 108010.

[7] Zhu Anqi, Zhang Lin, Shen Ying, et al. Zero-shot restoration of underexposed images via robust retinex decomposition [C]// Proc of IEEE International Conference on Multimedia and Expo. Piscataway, NJ: IEEE Press, 2020: 1-6.

[8] Ma Long, Ma Tengyu, Liu Risheng, et al. Toward fast, flexible, and robust low-light image enhancement [C]// Proc of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2022: 5637-5646.

[9] Jiang Yifan, Gong Xinyu, Liu Ding, et al. Enlightengan: deep light enhancement without paired supervision [J]. IEEE Trans on Image Processing, 2021, 30: 2340-2349.

[10] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a “Completely Blind” Image Quality Analyzer," IEEE Signal Processing Letters, vol. 20, no. 3, pp. 209-212, March 2013.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[11] Anish M ,Krishna A M ,Conrad A B .No-reference image quality assessment in the spatial domain.[J].IEEE transactions on image processing : a publication of the IEEE Signal Processing Society,2012,21(12):4695-708.