1233: ROBUST ENHANCEMENT, UNDERSTANDING AND ASSESSMENT OF LOW-QUALITY MULTIMEDIA DATA

Colorectal endoscopic image enhancement via unsupervised deep learning

Guanghui Yue[1] [ID] · **Jie Gao[1]** · **Lvyin Duan[1]** · **Jingfeng Du[2]** · **Weiqing Yan[3]** · **Shuigen Wang[4]** · **Tianfu Wang[1]**

Abstract

Currently, various deep learning methods have been developed to address the image enhancement tasks based on paired high-quality images as references. For the low-light endoscopic image enhancement task, it is difficult to obtain paired high-quality images and to extract features from dark areas. In addition, the enhanced images easily appear color distortions. In this study, we propose an unsupervised deep learning scheme based on the Cycle Generative Adversarial Network to enhance the endoscopic image. Because extracting features in the dark areas is important but challenging, we embedded an adaptive reverse attention module in generators to help the network focus on low-light areas and enhance these areas. We also introduce a color consistency constraint to maintain color constancy. To evaluate the performance of the proposed enhancement method, a blind evaluation methodology is proposed in view of no specific quality assessment metric specially designed on this field. Extensive subjective and objective experiment results demonstrate that the proposed method is competent for the colorectal endoscopic image enhancement task, and performs better than both conventional methods and popular deep learning-based methods on 200 real-captured colonoscopy images. In the objective experiment, the proposed method ranks first with a PIQE score of 11.1525 and an NIQE score of 11.1525, outperforming five competing methods. It also receives the best results from an average score of 1.455 over 200 test images of the subjective experiment.

✉ Guanghui Yue
   yueguanghui@szu.edu.cn

✉ Lvyin Duan
   duanlvyin@tju.edu.cn

[1] School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen University, Shenzhen 518060, China

[2] Department of Gastroenterology and Hepatology, Shenzhen University General Hospital, Shenzhen 518000, China

[3] School of Computer and Control Engineering, Yantai University, Yantai 261400, China

[4] Yantai IRay Technology Co., Ltd., Yantai 261400, China

🗠 Springer

1 Introduction

Nowadays, colorectal cancer is one of the leading killers, which threatens human life with the second-highest mortality rate and the third-highest incidence rate among all cancers [38]. Endoscopy is considered an effective way to screen colorectal diseases and to prevent early colorectal cancer in clinical practice [52]. Meanwhile, endoscopic images play a crucial role in effective diagnosis and treatment, which provide physicians with adequate visual information related to biological tissues [3, 6, 7, 33, 34, 50].

Clinically, endoscopic imaging unavoidably suffers from low quality due to the intestinal peristalsis when capturing images, and the poor clearing before endoscopy. Another factor that affects the quality of endoscopy images is weak illuminance, which is caused by the absence of extra light illumination inside the body except for the unidirectional light source emitted from the moving endoscope. Such a dynamic lighting process easily creates dark areas that affect the surgical environment (Fig. 1). Moreover, the low-light problem also weakens the performance of many subsequent image analysis tasks, such as polyp detection, polyp segmentation [11, 46, 47, 54, 56], and the computer-aided diagnosis [12, 44, 51]. Therefore, developing an image enhancement algorithm can enhance the visual effect and improve surgical accuracy for surgeons.

In the past few years, numerous methods have been proposed to enhance low-light images at the software end. Early works mainly designed conventional handcrafted feature-based methods, such as histogram-based algorithms [1, 13] and Retinex-based algorithms [8, 10, 36]. The former increases the image contrast by redistributing the intensity based on the histogram. The latter divides the image into two components, and processes them separately to generate the final enhanced image. For instance, Hiroyuki Okuhata et al. [33] presents a real-time image enhancement technique for gastric endoscopy based on the Retinex theory by introducing a variational model to minimize the computational cost. To improve surgical vision, Luo et al. [26] proposed a multi-scale bilateral-weighted retuned strategy which is capable of removing non-uniform and highly directional illumination. These methods, however, are difficult to reproduce high-quality images due to the complex image contents and limited representation of handcrafted features.
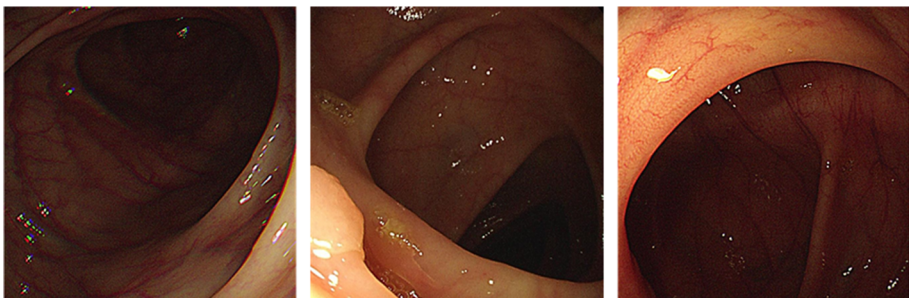

Fig. 1 Low-light endoscopic images

In recent years, deep learning (DL) based approaches had been widely adopted in computer vision[17–20, 42, 45, 48], medical image processing and analysis [15, 43, 49, 53], and have gradually evolved as an alternative of image enhancement [2, 9, 14, 27]. Contrary to conventional methods that change the intensity distribution or that relies on potentially physical models, DL-based methods can enhance low-light images automatically. For instance, the pioneering DL-based LLNet [25] employs a variant of stacked-sparse encoder to brighten low-light images. Later, Ren et al. [35] designed a more complex end-to-end network, which includes an encoder-decoder sub-network and a recurrent neural network. The encoder-decoder sub-network is used for image content enhancement, and the recurrent neural network is used for image edge enhancement. To improve the ability of integrating feature representations, pyramid network [22], residual network [41] and the Laplacian pyramid [23] are applied to low-light image enhancement. Among these methods, deep Retinex-based methods exhibit better performance in most cases. The key points of such methods lie in dividing low-light images into illuminance and reflectance components, and enhancing these two components separately. For instance, Retinex-Net [5], the first deep retinex-based method, includes a Decom-Net and an Enhance-Net. The Decom-Net is used for splitting the input image into illuminance and reflectance, and the Enhance-Net is used for enhancing the illumination map according to the reference image. To estimate the illumination map, Wang et al. [42] propose a DeepUPE network which is capable of learning an image to illumination mapping by extracting local and global features. Zhang et al. [55] develop a KinD network consisting of three sub-networks: layer decomposition, reflectance restoration and illumination adjustment.

Due to the lack of paired training data, numerous unsupervised learning schemes have been proposed to address the issue of image enhancement. For instance, Li et al.[16], propose a robust Retinex model that predicts the noise map, estimates the structural and reflectance maps, and segmentes the illumination map to better describe images captured in low-light conditions. Zhu et al.[58] introduce a new underexposed image restoration method called RRDNet, which uses a three-branch convolutional neural network framework to internally optimize the input image's lighting, reflection, and noise for better generalization under various lighting conditions. In addition, Li et al.[21] propose a new method called Zero-DCE, which describes image enhancement as a task of estimating image-specific curves using a deep network.

Although aforementioned DL-based methods have shown impressive performance on natural image enhancement, they are unsuitable for endoscopic image enhancement. First, low-light natural images in existing public datasets are generally with globally dark appearance, which is contrary to endoscopic images that contain both dark regions and bright regions. Due to the scene difference between these two kinds of images, most of these existing enhancement methods designed for natural images are prone to over-enhance the bright regions, resulting in poor visual experience. It is worth noting that, high-quality endoscopic images should generally have uniform illumination. Second, most of these existing enhancement methods usually require a large number of paired images to supervise the network during training. Unfortunately, because of the particularity of endoscopic imaging environment, it is very difficult or even impractical to simultaneously obtain the paired endoscopic images. Therefore, it is necessary to develop effective endoscopic image enhancement methods without paired images. Recently, the rapid development of unsupervised learning gives us a new inspiration for addressing this problem. Among numbers of unsupervised learning methods, CycleGAN [57] is a popular learning framework for mapping image in one domain to another domain, and matches the requirement of unpaired image enhancement task to some extent. However, most existing CycleGAN-based works are generally unrestricted and

have limited ability in effectively capturing color and detail information because they mainly learn the global appearance in the domain and the cycle consistency between domains.

Given these aspects above, we in this paper propose a novel unsupervised low-light endoscopic image enhancement method, namely Color Constrained GAN (CCGAN). Specifically, it bridges the mapping between low-light and normal-light endoscopic images without any paired information. Considering that the low-light areas are usually with dark appearance, we introduce an adaptive reverse attention module (ARAM) in generators to help the network focus on local features in these areas. Moreover, a novel color consistency loss is proposed to relieve the problem of color distortion. Since existing literature lacks quality evaluation metrics specifically designed for endoscopic images, a blind evaluation method is developed. Additionally, we collect a clinical real-world dataset with unpaired low/normal light endoscopic images to train the network. Experimental results on the collected dataset show that the proposed CCGAN is competent to the endoscopic image enhancement task, and outperforms four mainstream competing methods in terms of objective and subjective evaluation. The four main contributions of this paper are as follows:

- A CCGAN is proposed to address the low-light endoscopic image enhancement problem without any pair information. The proposed method pays emphasis on the dark region enhancement and color details preservation.
- Considering the intensity distribution of low-light images, we propose an ARAM to focus on dark areas. This attention module can adaptively determine the weight values of dark regions and extract local features.
- A novel loss function, named color consistency loss, is proposed to preserve color information and relieve the color distortion for enhanced images.
- A blind quality evaluation methodology is proposed to evaluate the endoscopic image quality.

The remainder of the paper is organized as follows. The proposed image enhancement network is introduced in Section 2. In Section 3, the proposed blind quality evaluation method is described in detail. Experimental results are shown and analyzed in Section 4, followed by the discussions in Section 5 and conclusions in Section 6.

2 Method

2.1 Network architecture

In this work, an unpaired learning framework, CCGAN, is introduced to enhance low-light endoscopic images. The network is responsible for learning a suitable mapping from domain A to domain B without the requirement of paired images in training phase, as shown in Fig. 2. The proposed network mainly includes four essential parts: two generators ($G_{A2B} : A \rightarrow B$ and $F_{B2A} : B \rightarrow A$) and two discriminators (D_B and D_A). The generator is capable of generating fake data to fool the discriminator. The discriminator tries to distinguish the fake data from the total data, including the real data and fake data. The network stops the training procedure until the discriminator can not distinguish the difference between the generated fake data and the real data. To make the network focus more on low-light areas, an ARAM is embedded in each generator. Furthermore, we also introduce a novel color consistency loss function apart from the basic loss functions of the CycleGAN network to alleviate the color distortion caused by luminance changes in generators.

2.2 Generators

The task of endoscopic image enhancement is treated as a translation from one domain to another domain. The generate can create new data under certain constraints. In the proposed network, there are two generators: G_{A2B} and F_{B2A}. The former are used to translate the image from domain A (low-light images) to B (normal-light images), and the latter is used to translate the image from domain B to A. The generators adopt an encoder-decoder structure based on U-Net [37]. Specifically, for each low-light image a from domain A, it is firstly forwarded into the generator G_{A2B}, which generates a new normal-light image $\tilde{a} = G_{A2B}(a)$ based on the image style of domain B. Secondly, D_B distinguishes whether the generated image $\tilde{a}$ is real or fake. Thirdly, $G_{A2B}(a)$ is transmitted to the generator F_{B2A} to generate a low-light image $F_{B2A}(G_{A2B}(a))$. Finally, a forward cycle-consistency loss is applied for constraining the input image a and the generated image $F_{B2A}(G_{A2B}(a))$. The above process can be described as a forward cycle consistency: $a \rightarrow G_{A2B}(a) \rightarrow F_{B2A}(G_{A2B}(a)) \approx a$. Similarly, a backward cycle consistency is formed as: $b \rightarrow F_{B2A}(b) \rightarrow G_{A2B}(F_{B2A}(b)) \approx b$, where b represents one normal-light image from the domain B, $F_{B2A}(b)$ is a image generated by the image b through the generator F_{B2A}, $G_{A2B}(F_{B2A}(b))$ is a image generated by the image $F_{B2A}(b)$ through the generator F_{B2A}. In the image generation process, the ARAM is embedded to help the network extract local features and focus on low-light areas of images.

More concretely, generators G_{A2B} and F_{B2A} have three encoder layers and three decoder layers, as shown in Fig. 2. The outputs of the second and third layer in the encoder are multiplied by the outputs of the second and first layers in the decoder, respectively. The residual blocks consist of two stacked 3×3 Convolution-BatchNorm-ReLU units and use a shortcut to connect the input and output. The ARAM consists of an adaptive reverse channel attention module (ARCAM) and a reverse spatial attention module (RSAM). For convenient understanding, the architecture details of generators are presented in Table 1.
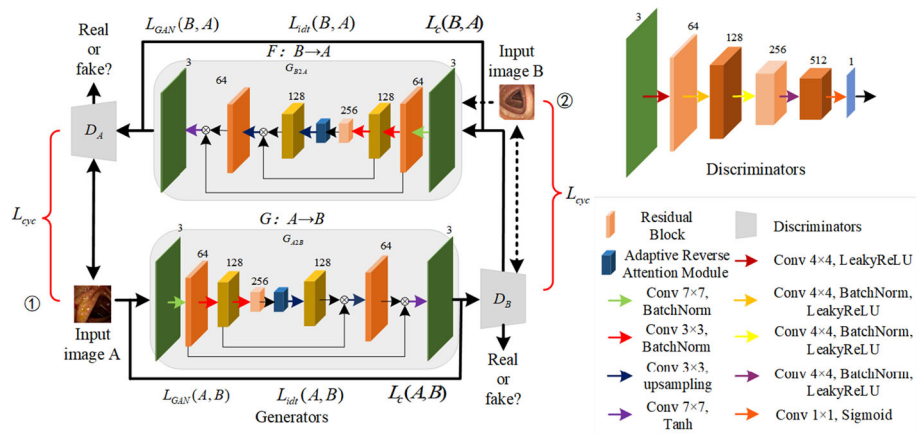

Fig. 2 The overall structure diagram of CCGAN. It comprises two generators (G_{A2B}) and (F_{B2A}), two discriminators (D_B) and (D_A), and two types of cycle consistency: ① forward cycle consistency: $a \rightarrow G_{A2B}(a) \rightarrow F_{B2A}(G_{A2B}(a)) \approx a$; ② backward cycle consistency: $b \rightarrow F_{B2A}(b) \rightarrow G_{A2B}(F_{B2A}(b)) \approx b$. a and b represent low-light and normal-light endoscopic images, respectively. L_{GAN}, L_{cyc} and L_c denote the transfer loss, cycle consistency loss and color consistency loss, respectively.

Table 1 The detailed architecture of each generator for the proposed CCGAN

Block	Layer	Type[1]	Input	Params[2]
Conv block1	conv1	conv+IN+ReLU	image	$(7 \times 7 \times 64, 1, 3)$
Conv block2	conv2	conv+IN+ReLU	conv1	$(3 \times 3 \times 128, 2, 1)$
Conv block3	conv3	conv+IN+ReLU	conv2	$(3 \times 3 \times 256, 2, 1)$
Residual Block	Resblock	(conv+IN+ReLU)×2 ⊕ conv3	conv3	$(3 \times 3 \times 256, 1, 1) \times 2$
ARAM	-	-	Resblock	-
Upblock1	upconv1	tconv+IN+ReLU	ARAM	$(3 \times 3 \times 128, 2, 1)$
	multiply1	multiply	upcovn1,conv2	-
Upblock2	upconv2	tconv+IN+ReLU	multiply1	$(3 \times 3 \times 64, 2, 1)$
	multiply2	multiply	upcovn2,conv1	-
Upblock3	upconv3	tconv+IN+ReLU	multiply2	$(7 \times 7 \times 3, 1, 3)$

1: This column lists the type of layers, including, 'conv': down-sampling with padding and matching convolutional layer; 'tconv': transposed convolutional layer;'IN': instance normalization layer; 'ReLU': ReLU layer

2: 'Params' is formatted in kernel size, stride and padding for convolution layers, and kernel size, stride, and padding for transposed convolution layers

2.2.1 Adaptive Reverse Attention Module (ARAM)

In the proposed network, we introduce an ARAM to enforce the network to pay more attention on low-light areas and to help extract meaningful information from these areas. Specifically, ARAM is composed of ARCAM and RSAM, as shown in Fig. 3.

Adaptive Reverse Channel Attention Module (ARCAM): In general, the average pooling operation can describe global information of features. However, it is insufficient to reflect the significance of salient objects. In our proposed ARCAM, we use the max pooling operation to compensate the average pooling operation and combine the results of these two operations together to express more high-level features, as shown in the upper part of the Fig. 3. Specifically, when the feature $I_F \in \mathbb{R}^{C \times H \times W}$ is forwarded into the ARCAM, two features ($I_{F_c,A} \in \mathbb{R}^{C \times 1 \times 1}$ and $I_{F_c,M} \in \mathbb{R}^{C \times 1 \times 1}$) are obtained based on average pooling and max pooling operations, where C, H, and W denote the channel number, height, and width of the input feature I_F, respectively. Subsequently, a shared multi-layer perception (MLP) is used to refine the obtained features, and an adaptive parameter γ is introduced to change weights of the refined features. The refined maximal result with an adaptive parameter γ and the refined average result with an adaptive parameter $1 - \gamma$ is added to generate one adaptive weights map. Then, this map is activated by a Sigmoid function to produce the attention weight $W_B(I_{F_c})$ Finally, $W_B(I_{F_c})$ is multiplied by the input feature map to obtain the feature $M_B(I_{F_c})$. Finally, we use the reverse operation to change the obtained attention weights and get the reverse channel attention map $M_D(I_{F_c})$. In short, the proposed ARCAM mechanism can be described as follows:

$$M_D(I_{F_c}) = 1 - I_{F_c} \times [\sigma(\gamma \times M_1(ReLU(M_0 I_{F_c,A})) \\ + (1 - \gamma) \times M_1(ReLU(M_0 I_{F_c,M}))] \tag{1}$$

where $\sigma(\cdot)$ is the Sigmoid function. M_0 and M_1 are the weights of MLP. ReLU is the rectified linear unit activation function. γ is a learnable parameter to adapt the weight values of the input feature.

Reverse Spatial Attention Module (RSAM) To restrain the interference of irrelevant areas and focus on salient areas, we utilize the RSAM to enable the network to focus on low-light areas in spatial space. As shown in Fig. 3, when the feature $I_F \in \mathbb{R}^{C \times H \times W}$ is fed into the RSAM, two features $I_{F_s,A}$ and $I_{F_s,M}$ are obtained based on global average
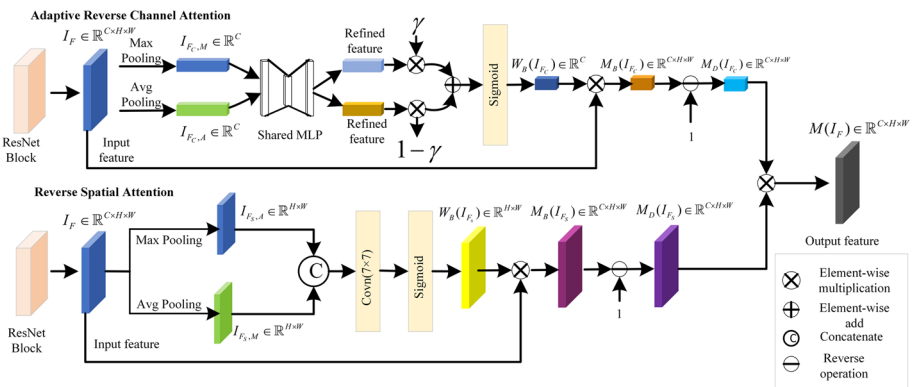

Fig. 3 The Adaptive Reverse Attention Module

pooling and global max pooling, respectively. Subsequently, the features ($I_{F_s,M} \in \mathbb{R}^{1 \times H \times W}$ and $I_{F_s,A} \in \mathbb{R}^{1 \times H \times W}$) are concatenated, followed by a 7×7 convolution operation and a Sigmoid function to obtain the attention weights $W_B(I_{F_s})$. Then, the attention weights $W_B(I_{F_s})$ is multiplied by the input feature map to obtain the spatial attention map $M_B(I_{F_s})$. Finally, we use the reverse operation to change the obtained attention weights and get the reverse spatial attention map $M_D(I_{F_s})$. In short, the proposed RSAM mechanism can be described as follows:

$$M_D(I_{F_s}) = 1 - [\sigma(Conv^{7 \times 7}(I_{F_s} \times [I_{F_s,A}; I_{F_s,M}]))], \qquad (2)$$

where $Conv^{7 \times 7}$ denotes a convolution operation with the kernel size of 7×7.

After obtaining the $M_D(I_{F_c})$ and $M_D(I_{F_s})$, the reverse attention map $M(I_F)$ can be computed as:

$$M(I_F) = M_D(I_{F_c}) \times M_D(I_{F_s}). \qquad (3)$$

2.3 Discriminators

In discriminators D_A and D_B, the patchGAN [24] is used to classify the fake data and real data based on image patches rather than the whole image. In patchGAN, there are five convolution operations with the kernel size of 4×4, a stride of 2 in the first three layers, and a stride of 1 in the last layers, and their channel numbers are 3, 64, 128, 256, and 512, respectively. The middle three convolution layers adopt the Instance Normalization (IN) layers, followed by a LeakyReLU with a scope of 0.2 [39]. Finally, the Sigmoid activation function is utilized to produce a 1-dimensional output. The details of discriminators are shown in Table 2.

2.4 Objective function

The proposed CCGAN framework has two kinds of loss functions, including a transfer loss L_t and a color consistency loss L_c. L_t is responsible for constraining the generated image and the original image. L_c is used to keep the color consistency further. The total objective loss L_{total} can be described as:

$$L_{total} = L_t + L_c \qquad (4)$$

In the following sections, we will introduce the transfer loss and the color consistency loss in detail.

Table 2 The detailed architecture of each discriminator for the proposed CCGAN

Block	Layer	Type[1]	Input	Params[2]
Conv block1	conv1	conv+LeakyReLU	image	$(4 \times 4 \times 64, 2, 1)$
Conv block2	conv2	conv+IN+LeakyReLU	conv1	$(4 \times 4 \times 128, 2, 1)$
Conv block3	conv3	conv+IN+LeakyReLU	conv2	$(4 \times 4 \times 256, 2, 1)$
Conv block4	conv4	conv+IN+LeakyReLU	conv3	$(4 \times 4 \times 512, 1, 1)$
Conv block5	conv5	conv+LeakyReLU	conv4	$(4 \times 4 \times 1, 1, 1)$

[1]: 'Type' lists the type of layers, including, 'conv': down-sampling with padding and matching convolutional layer; 'IN': instance normalization layer; 'LeakyReLU': LeakyReLU layer
[2]: 'Params' is formatted in kernel size, stride, and padding for convolution layers

2.4.1 Transfer loss

Transfer loss is one kind of basis objective functions in CCGAN framework, including two adversarial losses, two cycle consistency losses and one identity mapping loss. The adversarial loss L_{GAN} is applied to both the generator/ discriminator pairs $(G_{A2B}/D_B, F_{B2A}/D_A)$. Formally, it can be expressed as:

$$L_{GAN}(G_{A2B}, D_B, A, B) = \mathbb{E}_{b \in B}[log(D_B(a))]$$
$$+ \mathbb{E}_{a \in A}[log(1 - D_B(G_{A2B}(a)))] \tag{5}$$

$$L_{GAN}(F_{B2A}, D_A, B, A) = \mathbb{E}_{a \in A}[log(D_A(b))]$$
$$+ \mathbb{E}_{b \in B}[log(1 - D_A(F_{B2A}(b)))] \tag{6}$$

where a and b are samples from domains A and B, respectively. $G_{A2B}(a)$ converts the image a from domain A to domain B based on the image style of domain B. In contrast, $F_{A2B}(b)$ converts an image b from domain B to domain A based on the image style of domain A. D_B (or D_A) identifies the difference between real samples from domain B (or A) and the generated ones from domain A (or B).

The proposed CCGAN framework contains two consistencies: 1) forward cycle consistency: $a \rightarrow G_{A2B}(a) \rightarrow F_{B2A}(G_{A2B}(a)) \approx a$; 2) backward cycle consistency: $b \rightarrow F_{B2A}(b) \rightarrow G_{A2B}(F_{B2A}(b)) \approx b$. With such consistencies, the output image retains the same content as the input image, but has different image styles. The total cycle consistency loss is defined as:

$$L_{cyc}(G_{A2B}, F_{B2A}) = \mathbb{E}_{a \in A}[\|F_{B2A}(G_{A2B}(a)) - a\|_{SmoothL1}]$$
$$+ \mathbb{E}_{b \in B}[\|G_{A2B}(F_{B2A}(b)) - b\|_{SmoothL1}] \tag{7}$$

where $\|\cdot\|_{SmoothL1}$ denotes smooth L loss, which is used to help the network converge.

In addition, when the real sample from A (or B) are applied to G_{A2B} (or F_{B2A}), the generated sample and the real sample should be similar. They follow identity mappings as below: $\tilde{a} = F_{B2A}(a) \approx a$ and $\tilde{b} = G_{A2B}(b) \approx b$, where $\tilde{a}$ and $\tilde{b}$ are the generated samples by the generator G_{A2B} and F_{B2A}. The identity mapping loss L_{idt} is defined as:

$$L_{idt}(G_{A2B}, F_{B2A}) = \mathbb{E}_{b \in B}[\|G_{A2B}(b) - b\|_1] + \mathbb{E}_{a \in A}[\|F_{B2A}(a) - a\|_1] \tag{8}$$

The transfer loss is finally defined as:

$$L_t = \lambda_1(L_{GAN}(G_{A2B}, D_B, A, B) + L_{GAN}(F_{B2A}, D_A, B, A))$$
$$+ \lambda_2 L_{cyc}(G_{A2B}, F_{B2A}) + \lambda_3 L_{idt}(G_{A2B}, F_{B2A}) \tag{9}$$

where $\|\cdot\|_1$ denotes L1 loss. λ_1, λ_2, and λ_3 are weight parameters. In this study, we set λ_1, λ_2, and λ_3 to 0.5, 5.0, and 10.0, respectively.

2.4.2 Color consistency loss

Although the transfer loss can achieve inter-domain image translation, it is difficult to preserve color consistency due to the under-constrains in the adversarial training process. For medical image enhancement, the luminance change affects the color expression, which may lead to misdiagnosis. To keep color consistency, we propose a color consistency loss function L_c. In the proposed CCGAN, there are two generators (G_{A2B} and F_{B2A}). For generator G_{A2B},

the color consistency loss can be computed as:

$$L_c(A, B)_{A \to B} = (H_{oriA2B} - H_{genA2B})^2 + (\frac{S_{oriA2B}}{S_{genA2B}} - \frac{V_{oriA2B}}{V_{genA2B}})^2 \tag{10}$$

For generator F_{B2A}, the color consistency loss can be computed as:

$$L_c(B, A)_{B \to A} = (H_{oriB2A} - H_{genB2A})^2 + (\frac{S_{oriB2A}}{S_{genB2A}} - \frac{V_{oriB2A}}{V_{genB2A}})^2 \tag{11}$$

where $L_c(A, B)$ is the color consistency loss of G_{A2B}. Notably, we transfer the image from RGB to HSV color space since it is convenient to process the color and luminance information separately. $H_{oriA2B}(H_{oriB2A})$ and $H_{genA2B}(H_{oriB2A})$ are the hue values of original and output images of the generator $G_{A2B}(F_{B2A})$. $S_{oriA2B}(S_{oriB2A})$ and $S_{genA2B}(S_{genB2A})$ are the saturation values of original and output images of the generator $G_{A2B}(F_{B2A})$. $V_{oriA2B}(V_{oriB2A})$ and $V_{genA2B}(V_{genB2A})$ are the luminance values of original and output images of the generator $G_{A2B}(F_{B2A})$. $L_c(B, A)$ is the color consistency loss of F_{B2A}. In (11), $(H_{oriA2B} - H_{genA2B})^2$ is capable of preserving the hue consistency between low-light and enhanced images, and $(\frac{S_{oriA2B}}{S_{genA2B}} - \frac{V_{oriA2B}}{V_{genA2B}})^2$ is used for keeping the color saturation change with the luminance increasing.

Finally, the color consistency loss L_c of the proposed CCGAN can be expressed as:

$$L_c = L_c(A, B)_{A \to B} + L_c(B, A)_{B \to A} \tag{12}$$

3 Proposed blind quality evaluation method

In general, the distance between the reference and distorted image is a direct measurement to reveal the quality of the distorted image [40]. However, it is not suitable for distorted images without paired images. For the enhancement task in this study, there are no perfect-quality images as reference images. Consequently, one no-reference quality evaluation method
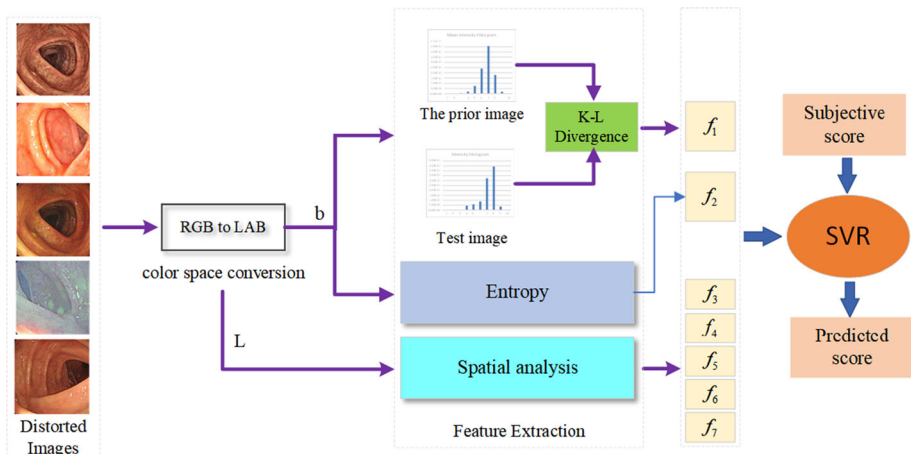

Fig. 4 Framework of the proposed blind endoscopic image quality evaluation method

should be considered. However, existing literature lacks of methods designed for endoscopic image evaluation. To solve this problem, a Blind Endoscopic Image Quality Evaluation (BEIQE) method is proposed, as shown in Fig. 4. First, the endoscopic image is converted to LAB color space from RGB color space. Second, Kullback-Leibler (K-L) divergence distance f_1 between the test image histogram and the prior one is calculated. Specifically, f_1 is extracted by analyzing the b-chromaticity channel of normal-light and low-light endoscopic images. Third, the entropy and spatial feature are extracted separately. The entropy value f_2 of each image reflects the information amount. Spatial features (f_3, f_4, f_5, f_6, f_7) are extracted from the generalized Gaussian distribution (GGD) and asymmetric generalized Gaussian distribution (AGGD) fittings of the mean subtracted contrast normalized (MSCN) [29]. Then, seven features mentioned above are combined into a feature vector. Finally, a quality assessment model is built through support vector regression (SVR) to connect the relationship between features and subjective ratings. The quality score of a test image can be estimated by feeding its feature vector into the quality assessment model.

3.1 Chroma feature extraction

Generally, endoscopic images usually suffer from color distortions during image enhancement. To illustrate this, we convert 200 normal-light and 200 low-light endoscopic images from RGB color space to LAB color space (some examples are shown in Figs. 5 (a) and 5 (b)), and analyze the statistical properties of each corresponding channel between them. As shown in Fig. 5 (c), the b-chromaticity histogram distribution of normal-light is more concentrated than that of low-light images. Thus, quantifying this statistical regularity can provide us one way to measure the color difference between normal-light and low-light endoscopic images.

In the experiment, the dataset consists of 1,000 normal-light and 1,000 low-light endoscopic images. Normal-light endoscopic images include diverse categories of normal-light endoscopic scenes (such as polyps, bubbles, reflective. etc.), it is assumed that the mean b-chromaticity histogram on this dataset can approximately characterize the b-chromaticity distribution of normal-light endoscopic scenes. For a query image, we can measure the chro-
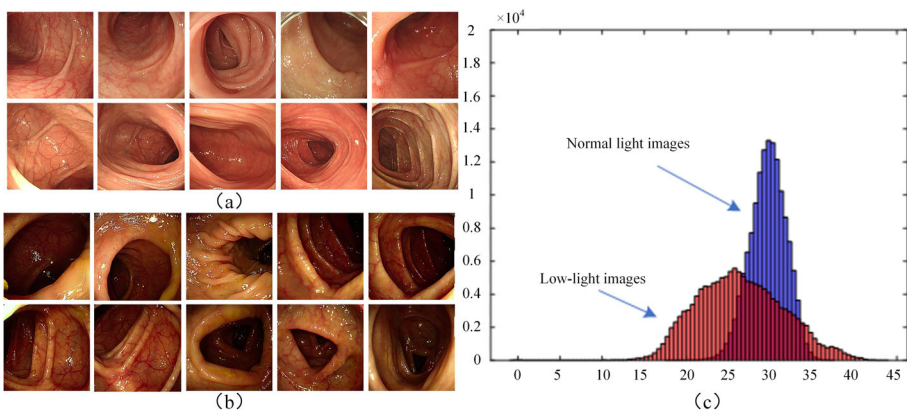

Fig. 5 (a) Examples of normal light endoscopic images (b) Examples of low-light endoscopic images (c) Mean histograms of b-chromaticity of normal-light endoscopic images and low-light endoscopic images
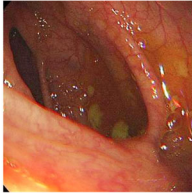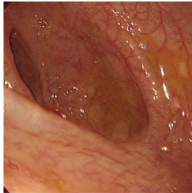
| Retinex | RRDNet | CCGAN |

$f_1 = 0.0901$ $f_1 = 0.2153$ $f_1 = 0.2879$

$f_2 = 6.4738$ $f_2 = 6.1564$ $f_2 = 5.3482$

$MOS = 3$ $MOS = 2$ $MOS = 1$

Fig. 6 Comparisons between the extracted features and MOSs (mean opinion score)

maticity distribution change via the K-L divergence DL_{KL} which can be expressed as:

$$DL_{KL}(p\|q) = \sum_{i=1}^{N} p(x_i) \cdot (log\ p(x_i)) - log\ q(x_i)) \tag{13}$$

where p and q represent b-chromaticity histograms of the prior image and query image, respectively. x_i is the probability of the ith bin value of the b-chromaticity histogram. N denotes the total bin number of the b-chromaticity histogram. In this proposed method, the average b-chromaticity histogram is used as prior knowledge for normal light endoscopic images, serving as the reference distribution. The b-chromaticity histogram of the query image is used as the comparison distribution to calculate the K-L divergence, in order to evaluate the level of distortion.

3.2 Entropy

For a high-quality endoscopic image, it contains plenty of details in textures, structures and colors, information change, and so on. Here, we use image entropy to characterize the aggregation properties of the b-chromaticity distribution and to reflect the information amount. The entropy E as the image quality feature f_2, which can be computed as:

$$E = -\sum_{i=1}^{255} \phi_i \cdot log\phi_i \tag{14}$$

where ϕ_i is the probability of the ith b-chromaticity value.

3.3 Spatial features

Spatial features, extracted form the empirical distribution under a spatial scene statistic model, can exhibit distortions, e.g., blur or noise and so on [29]. Based on this fact, we first compute locally normalized luminances via local MSCN for the distorted image [37]. Then, the first spatial feature f_3, one shape parameter, is obtained by fitting the MSCN using the GGD. We also explore the statistical relationships between neighboring pixels and extract other four shape spatial features f_4, f_5, f_6 and f_7 from four orientations - horizontal, vertical, main-diagonal and secondary-diagonal by fitting the MSCN using the AGGD.

To better understand these extracted features, we present three images processed by Retinex, RRDNet, and CCGAN in Fig. 6. It is clear that, the MOS value and feature values (f_1 and f_2) have monotonic relationships. Figure 7 shows the histogram of MSCN coefficients and the histogram of MSCN coefficients of four orientations, respectively. As seen, the image, processed by CCGAN, shows a narrowest shape followed by images processed by RRDNet and Retinex. These figures indicate that the extracted features are quality-aware.
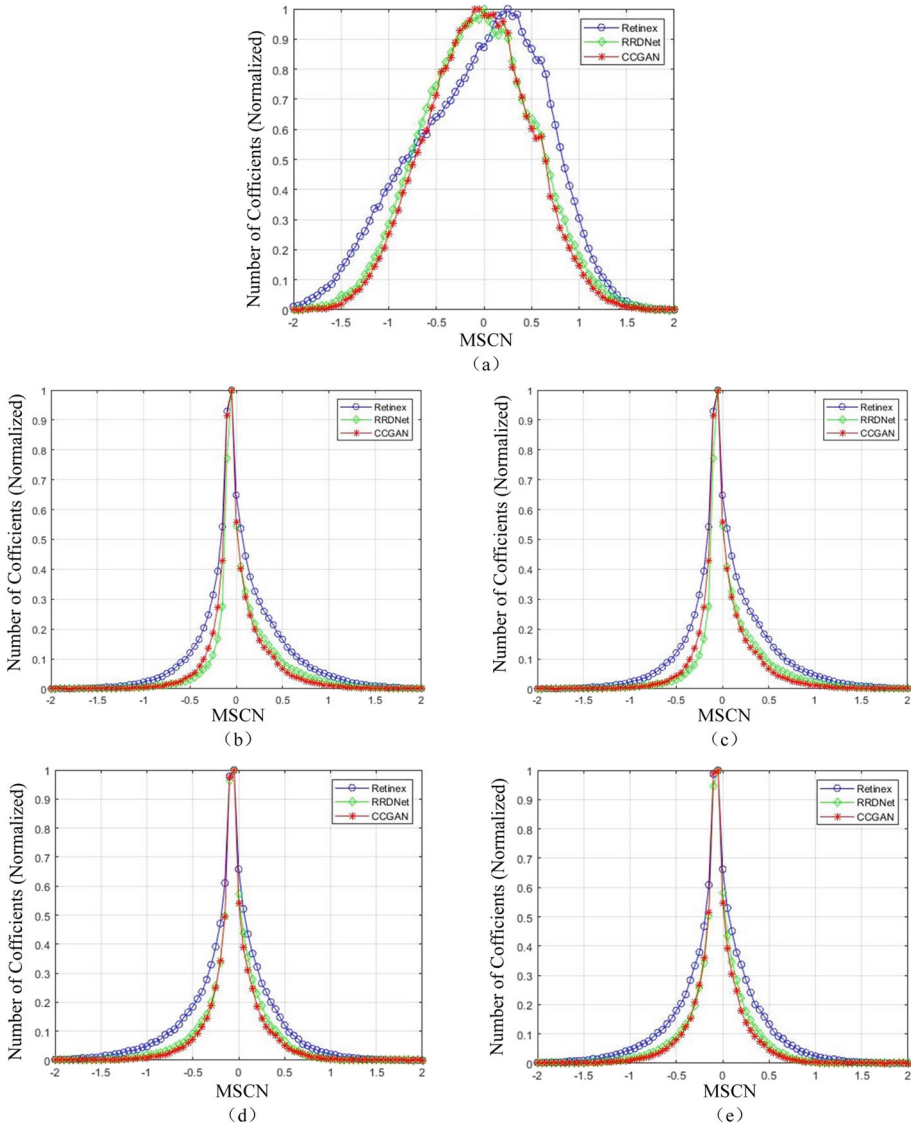

Fig. 7 Histogram of MSCN coefficients for images processed by Retinex, RRDNet, and CCGAN

3.4 Quality prediction

After feature extraction, we use SVR to train the extracted features as their corresponding subjective quality scores by employing the LIBSVM package [4]. In the experiment, a quality prediction dataset, which includes 1,000 images with MOS scores, is used. In addition, three commonly used evaluation criteria are adopted. Specially, Kendall's rank orders correlation coefficient (KRCC) [28] and Spearman's rank-order correlation coefficient (SRCC) [54] are two criteria for evaluating the prediction monotonicity, whereas the Pearson linear correlation coefficient (PLCC) [38] is a criteria for evaluating the prediction accuracy. For an excellent method, the values of PLCC, SRCC and KRCC are close to one. To ensure a fair evaluation, we randomly divide the dataset into training and testing subsets 1000 times, with 80% of the data for training SVR and the rest for testing. The median of the 1,000 results is reported as the overall performance, as shown in Table 3. As seen, the proposed method obtains a performance of PLCC=0.8701, indicating a strong correlation between the perceptual quality assessment score and subjective results. Moreover, it also performs a relatively strict prediction monotonicity with SRCC=0.8477 and KRCC=0.7013.

4 Results

In this section, we firstly introduce the dataset and implementation settings. Then, a series of experiments are conducted for performance comparison, including quantitative comparison, qualitative comparison, and subjective evaluation, for performance comparison between the proposed CCGAN and state-of-the-art methods.

4.1 Dataset and implementation settings

Since the proposed CCGAN network is trained with unpaired low-light and normal-light images, we collected several unpaired images and divided them into a training set and a testing set without content duplication. This training set consisting of 1,000 low-light and 1,000 normal-light endoscopic images is collected from the department of Gastroenterology and Hepatology, Shenzhen University General Hospital. The testing set comprises 200 endoscopic images with real-world distortions. Our collected endoscopic images have undergone rigorous selection, primarily to ensure their quality and representativeness by screening out unclean endoscopic images. Additionally, to enhance the representativeness of our dataset, we specifically gathered some endoscopic images containing special cases such as colonic polyps and colonic inflammation. In this experiment, we emphasize on the translation from low-light endoscopic images to normal-light endoscopic images. Figure 8 provides some image examples from the training set.

Table 3 The results of PLCC, SRCC, and Kendall coefficient

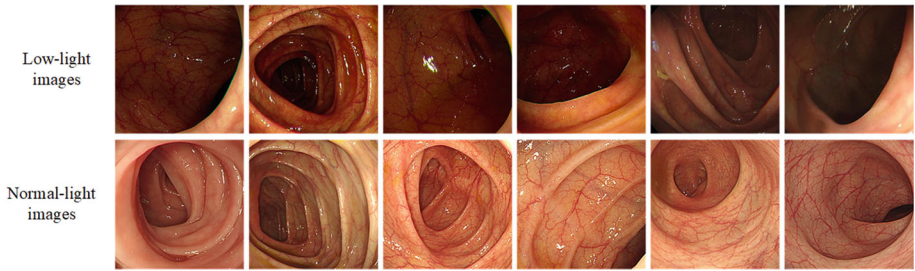| Criteria	Total
PLCC	0.8701
SRCC	0.8477
KRCC	0.7013

Fig. 8 Image samples from the training set

The proposed CCGAN is implemented with the PyTorch library and is trained on a workstation equipped with a single Nvidia GPU (GeForce RTX 3090, 24GB RAM). All images are converted to JPG format and resized into 256×256 pixels. A random flipping operation is applied for data augmentation. CCGAN is trained from scratch for 200 epochs with the learning rate of 1e-4. Adam optimizer is employed for network optimization and the batch size is set to 8.

4.2 Quantitative evaluation

For quantitative evaluation, we compare the proposed network with several image enhancement methods, including two classical handcrafted feature-based methods: contrast limited adaptive histogram equalization (CLAHE) [32] and Retinex [16], and two recently reported deep learning-based methods: RRDNet [58] and Zero DCE [21]. The parameters in conventional methods were set to the default values. For each deep learning method, we adopt the same training datasets as the proposed method, and follow their default settings. All experiments (training or test) are performed on the same workstation as the proposed method used.

In the experiment, the proposed BEIQE was used. In addition, two no-reference image quality assessment metrics were adopted: Natural Image Quality Evaluator (NIQE) [30] and Perception-based Image Quality Evaluator (PIQE) [40]. These two metrics are widely used for evaluating natural image distortions. The lower scores of these metrics indicate the better image quality achieved. Table 4 shows the quality scores of endoscopic images enhanced using different evaluation metrics. For convenient viewing, the best values of each evaluation metric are highlighted in the boldface. It can be seen that the proposed CCGAN exhibits the best performance in NIQE, PIQE and our evaluation method with average values of 3.2873, 11.1525 and 0.3725 across the 200 test images, respectively. CLAHE ranks second in PIQE

Table 4 Comparison among competing methods in terms of no-reference image quality assessment metrics

Method	NIQE	PIQE	BEIQE
CLAHE [32]	3.6106	18.8555	0.4566
Retinex [16]	4.2957	22.9727	0.5610
RRDNet [58]	3.4178	27.2028	0.4249
Zero DCE [21]	3.6055	25.6279	1.0127
CCGAN(ours)	3.2873	11.1525	0.3725

Table 5 The metrics of FLOPs, Params, and Running times reveal the performance comparison of our approach with the deep learning-based methods RRDNet and ZERO DCE

Method	FLOPs(G)	Params(M)	Time(s)
RRDNet [58]	67.1964	0.1281	0.7162
Zero DCE [21]	41.6269	0.0794	0.5189
CCGAN(ours)	183.9775	2.0681	1.0040

with the average value of 18.8555, followed by Retinex, Zero DCE and RRDNet. RRDNet performs better than other competing methods in NIQE and takes the second position with a score of 3.4178, followed by Zero DCE, CLAHE, and Retinex.

Furthermore, Table 5 presents the performance of each competing method in terms of FLOPs (Floating point Operations Per second), Params (Parameters), and Running time. Specifically, our CCGAN model is inferior to the other image enhancement methods in these three aspects. This is because CCGAN model has a complex framework to ensure effective feature representation for better image enhancement. In the future, we will update our CCGAN model by replacing the current backbone with a light-weight one.

To check the quality scores distribution across the 200 tested images for all the methods, we present the result using the violin plot, as shown in Fig. 9. In these figures, each violin plot indicates the probability density distribution of all tested scenes for the different methods. The white dots in these plots are average values of compared methods. As seen, the proposed CCGAN ranks the first with the lowest values in terms of NIQE, PIQE and BEIQE. The conclusion from the distribution performance is consistent with that of the average values well.
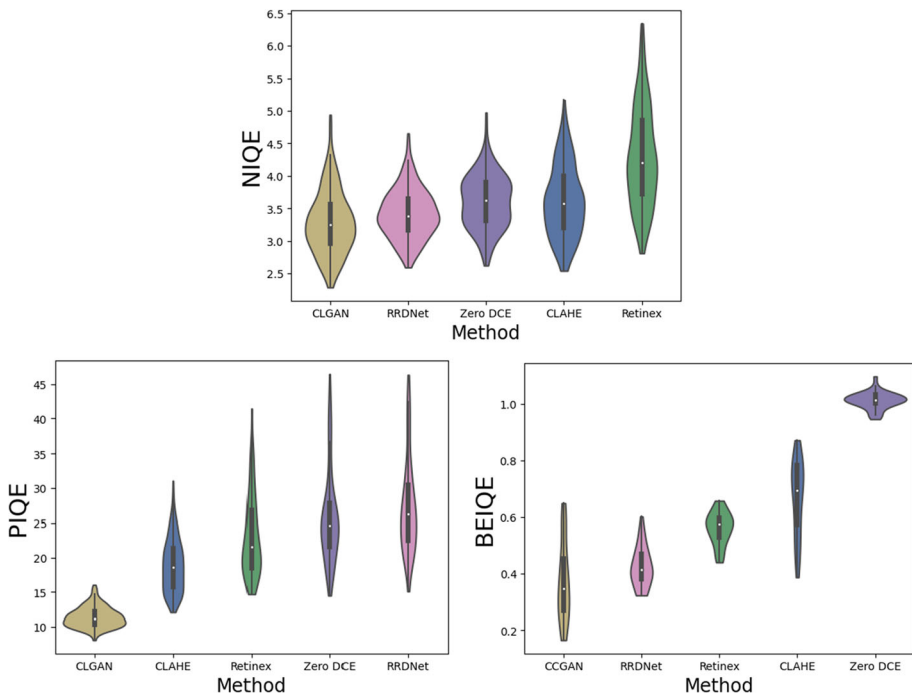

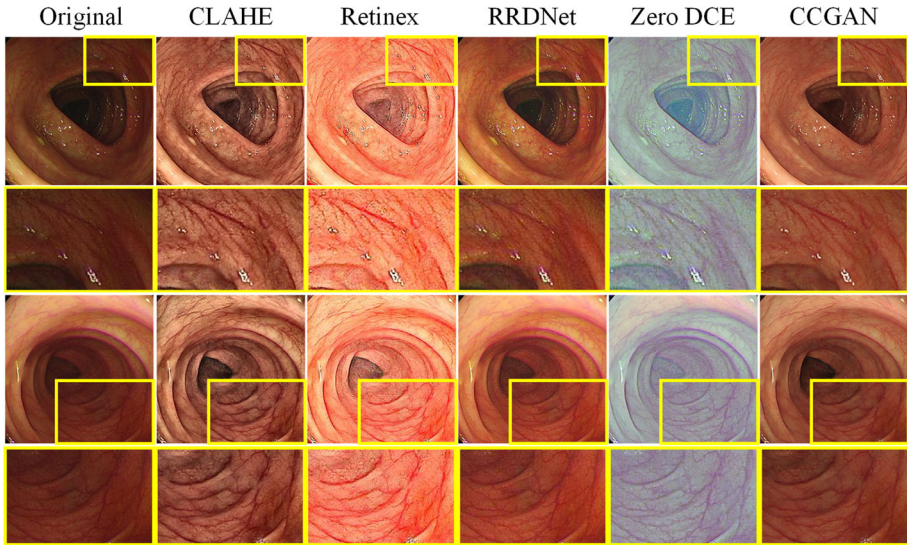**Fig. 9** Comparison of the performance distributions among competing methods

Fig. 10 Examples of endoscopy using different approaches (zoom in for a better view)

4.3 Qualitative comparison

Figure 10 shows the enhanced image results achieved by different enhancement methods. The first column shows the original low-light endoscopic images, and the second to the last columns are the images enhanced by: CLAHE, Retinex, RRDNet, Zero DCE, and the proposed CCGAN.

To analyze the details of the enhanced images, we enlarge some details in the yellow bounding boxes. As seen, CLAHE, Retinex, and Zero DCE all cause color distortion to some extent. Specifically, CLAHE expands the blood vessels and amplifies noise in the overall images. Retinex results in severe color distortions and misses the detail information. Zero DCE brings the overall illumination improvement, but it also leads to severely baised color. RRDNet easily leads to high saturated colors, leading to some information missing. In conclusion, the deep learning methods, RRDNet and Zero DCE, generate unsatisfactory visual result in terms of detailed information and color reproduction. In contrast, CCGAN not only enhances the low-light areas but also shows the details and colors well.

4.3.1 Subjective evaluation

We conducted a subjective evaluation to compare the performance of the proposed method with competing ones. In this experiment, a graphical user interface (GUI) is used for displaying the 200 endoscopic scenes [31]. For each endoscopic scene, it is first enhanced by five methods (CLAHE, Retinex, RRDNet, Zero DCE, and the proposed CCGAN). These enhanced images are presented in the GUI randomly. Figure 11 briefly shows the subjective experiment setup. The display presents five thumbnails (labeled A-E) obtained from the five low-light image enhancement methods on the left side of the screen. Each thumbnail is displayed in full-screen when participants double-click on it. One professional gastroenterologist with over ten years of experience is invited to rank these images from the quality evaluation perspective depending on his clinical experience.
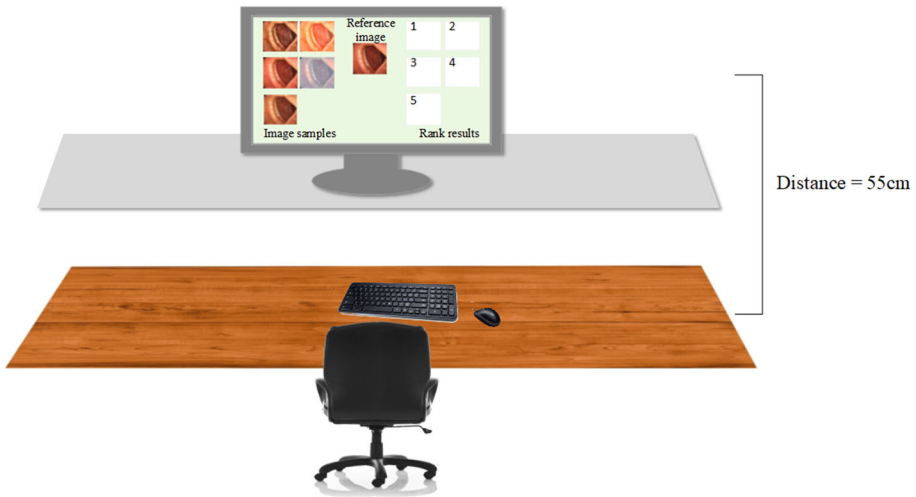
Fig. 11 Schematic diagram of the subjective experimental environment

The gastroenterologist can view each image displayed in full-screen mode and rank the images on the right side by dragging his preferred choice to its corresponding position (labelled 1-5), where 1 denotes the best one and 5 represents the worst one.

Figure 12 provides five histograms, each of which indicates the rank distribution of 200 endoscopic images generated by a method. For example, the proposed CCGAN ranks the first for 110 images out of 200 images, the second for 89 images, and the third for 1 image out of 200 images. By comparing the five histograms, it is clear that CCGAN receives the best results from the gastroenterologist, with an average rank score of 1.455 across over 200 samples. CLAHE and Zero DCE are not well scored because of the severe color distortion and noise.

4.4 Ablation studies

In this work, the proposed CCGAN benefits from two novel terms, i.e., color consistency loss and the ARAM. We conducted the following ablation studies to investigate their contributions. Here, our baseline is the regular CycleGAN method.
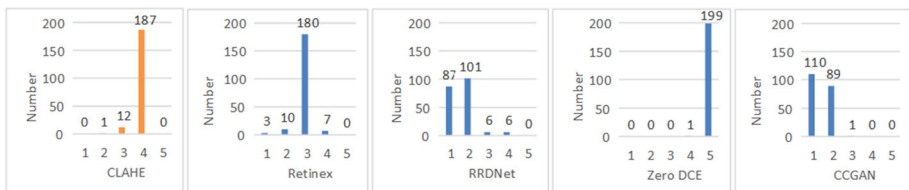

Fig. 12 The results of five methods in the subjective evaluation. In each histogram, the x-axis denotes the ranking index (1-5, 1 represents the highest value), and the y-axis denotes the number of images in each ranking index. As seen, CCGAN ranks the most top-ranking images and obtains the best performance with the smallest average ranking value

Table 6 Ablation studies. Here, our baseline is the standard CycleGan framework

Networks	NIQE	PIQE	BEIQE
Baseline	3.9121	14.5358	0.6374
Baseline+CA	3.7967	12.1274	0.5587
Baseline+ARAM	3.5825	11.4071	0.4913
Baseline+L_{color}	3.5440	13.4600	0.3779
Baseline+L_{color}+ARAM	3.2873	11.1525	0.3725

- **Color consistency loss:** Firstly, we verify the impact of the proposed color consistency loss. As shown in the second row of Table 6, the application of the color consistency loss brings improvements in NIQE (3.9121 vs. 3.5440), PIQE (14.5358 vs. 13.4600) and BEIQE(0.6374 vs. 0.3779) compared with the baseline method. This demonstrates that this color consistency loss is effective in assisting the proposed CCGAN to improve the image quality.
- **Adaptive reverse attention module:** To explore the effectiveness of the ARAM, we compared the performance of the CycleGAN baseline and that with ARAM over the collected endoscopic images dataset. As illustrated in the last line of Table 6, the ARAM brings significant improvements (11.1525) in PIQE compared with the CycleGAN baseline (14.5358). In NIQE, the ARAM also exhibits a significant improvement from 3.9121 to 3.2873. The results show that the application of ARAM contributes to the overall performance.

After incorporating ARCAM into the baseline, our experimental results show an obvious improvement in the three evaluation metrics NIQE, PIQE, and BEIQE. This indicates that ARCAM can help the model better focus on important channel information while reducing unnecessary computation, thus improving the accuracy and efficiency of the model. Specifically, the introduction of ARCAM allows the network to focus more on beneficial feature channels and filter out some useless channels, making the model's decisions more accurate. In addition, our spatial attention module can also adaptively focus on more important spatial position information on the image, helping the model learn useful features and improve the enhancement effect. Finally, combining AR-CAM and RASM can further improve the enhancement effect.

The above ablation studies show that the ARAM and color consistency loss play a positive role in performance improvement. The former focuses on enhancing the low-light areas, while the latter tends to preserve color consistency when the luminance increases in the scene. The combination of the ARAM and color consistency loss can obtain an impressive performance in endoscopic image enhancement.

5 Discussion

Low-light endoscopic images affect the observation of important tissues and even lead to missed diagnoses. However, it is difficult to obtain high-quality images due to the diverse illumination conditions and the low-quality imaging sensors. Low-light endoscopic image enhancement is an effective way to improve image quality. High-quality images can assist physicians in improving the accuracy of diagnosis. However, very few existing image enhancement methods focus on low-light endoscopic images. Additionally, due to the local

information loss and color distortion, most enhancement algorithms are not suitable for the endoscopic image enhancement task.

In this paper, we present a novel method that can handle the low-light endoscopic image enhancement. In the proposed method, we introduce the ARAM and color consistency loss to deal with low-light area enhancement and color distortion problems. In addition, we propose a blind quality evaluation method. Finally, we investigate the impact of the proposed method in terms of quantitative evaluation, visual inspective and subjective evaluation.

For quantitative evaluation, two conventional and two deep-learning image enhancement methods were selected. The proposed CCGAN exhibits overall best performance in NIQE, PIQE and BEIQE compared with competing image enhancement methods. Specifically, the other competing methods show a relatively inferior performance in the low-light area enhancement and color information preservation. This may be attributed to that, these methods are developed for natural images, thereby being incapable of tackling the endoscopic image enhancement task. In the proposed CCGAN, the ARAM and the color consistency loss are applied, which are used for enforcing the network to focus on specific low-light areas and extracting local features of original images, and alleviating color distortion. To explore their contributions, we further conducted two ablation studies. By comparing the results in Table 6, we can find that the performance of the baseline leaves considerable room for improvement. The color consistency loss, as shown in the second line of Table 6, brings approximately 0.4 increments of NIQE, 1.1 increments of PIQE, and 0.26 increments of BEIQE compared with the baseline. The combination of the ARAM and the color consistency, as shown in the last row, brings approximately 0.6 increments of NIQE and 3.4 increments of PIQE. By comparing all data in the table, we conclude that the proposed color consistency loss and the ARAM are effective for improving image quality.

For visual inspection, we present the results of two representative low-light images generated by five image enhancement methods. As illustrated, our CCGAN exhibits obvious superiority against competing methods in two aspects. First, CCGAN is more suitable for preserving color information. For instance, CLAHE leads to excessive enhancement of blood vessels, while Retinex and Zero DCE bring severe color distortion. Second, CCGAN can focus on low-light areas thanks to the proposed attention module ARAM. RRDNet exhibits good performance in color preservation and luminance improvement, but it can not enhance the low-light areas well and cause low contrast in these areas. Retinex and Zero DCE brighten the image as a whole but ignore local information CLAHE causes distortion in local areas. By contrast, CCGAN not only provides high contrast and sufficient color information, but also preserves details of local areas. Overall, the proposed CCGAN is more conducive to dealing with the endoscopic image enhancement task.

In the last experiment, we invited one professional expert with more than ten years of clinical experience to observe the enhanced endoscopic images obtained using different methods. The results show that CCGAN produces the overall most favored results by the expert subject, with an average ranking of 1.455 over 200 images. This also verifies the effectiveness of the proposed method.

6 Conclusion

In this work, we proposed an unsupervised deep learning framework named CCGAN for endoscopic image enhancement. To cope with the color distortion, we introduced a color consistency loss to constrain the color change between the original images and the generated

images. By carefully analyzing the characteristics of the low-light areas, we proposed an adaptive reverse attention module named ARAM. Owing to the collaboration of the consistency loss and ARAM, CCGAN can preserve local area information and relieve color distortion. To validate the effectiveness of the proposed CCGAN, we propose a blind evaluation metric by extracting K-L divergence, entropy, and spatial features. Finally, extensive experiments were conducted to compare the proposed CCGAN with four recently reported methods. The results show that our CCGAN is competent for addressing the challenging low-light endoscopic image enhancement task, and performs better than others.

Data availability statements The datasets generated during and/or analysed during the current study are not publicly available due to the need for hospital consent but are available from the corresponding author on reasonable request

Declaration

Conflicts of interest The authors declare that there are no conflicts of interest related to this article

References

1. Abdullah-Al-Wadud M, Kabir MH, Dewan MAA, Chae O (2007) A dynamic histogram equalization for image contrast enhancement. IEEE Trans. Consum. 53(2):593–600
2. Asif M, Chen L, Song H, Yang J, Frangi AF (2021) An automatic framework for endoscopic image restoration and enhancement. Appl. Intell. 51(4):1959–1971
3. Billah M, Waheed S (2020) Minimum redundancy maximum relevance (mrmr) based feature selection from endoscopic images for automatic gastrointestinal polyp detection. Multimed Tools Appl. 79(33):23633–23643
4. Chang C-C, Lin C-J (2011) Libsvm: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2(3)
5. Chen Wei WYJL, Wenjing Wang (2018) Deep retinex decomposition for lowlight enhancement. In: British Machine Vision Conference
6. Chou Y-C, Chen C-C (2022) Improving deep learning-based polyp detection using feature extraction and data augmentation. Multimed Tools Appl. 1–21
7. Das A, SS S (2021) Efficient quality enhancement of gastrointestinal endoscopic video by a novel method of color salient bilateral filtering. Multimed Tools Appl. 80(4):6235–6245
8. Fu X, Liao Y, Zeng D, Huang Y, Zhang X-P, Ding X (2015) A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. IEEE Trans. Image Process. 24(12):4965–4977
9. Gu S, Guo S, Zuo W, Chen Y, Timofte R, Van Gool L, Zhang L (2019) Learned dynamic guidance for depth image reconstruction. IEEE Trans. Pattern Anal. Mach. Intell. 42(10):2437–2452
10. Guo X, Li Y, Ling H (2016) Lime: Low-light image enhancement via illumination map estimation. IEEE Trans. Image Process. 26(2):982–993
11. Guo X, Yang C, Liu Y, Yuan Y (2020) Learn to threshold: Thresholdnet with confidence-guided manifold mixup for polyp segmentation. IEEE Trans. Med. Imaging. 40(4):1134–1146
12. Häfner M, Liedlgruber M, Uhl A, Vécsei A, Wrba F (2012) Color treatment in endoscopic image classification using multi-scale local color vector patterns. Medical Image Analysis. 16(1):75–86
13. Ibrahim H, Kong NSP (2007) Brightness preserving dynamic histogram equalization for image contrast enhancement. IEEE Trans. Consum. 53(4):1752–1758

14. Lai W-S, Huang J-B, Ahuja N, Yang M-H (2018) Fast and accurate image super-resolution with deep laplacian pyramid networks. IEEE Trans. Pattern Anal. Mach. Intell. 41(11):2599–2613

15. Lei H, Liu W, Xie H, Zhao B, Yue G, Lei B (2022) Unsupervised domain adaptation based image synthesis and feature alignment for joint optic disc and cup segmentation. IEEE Journal of Biomedical and Health Informatics. 26(1):90–102

16. Li M, Liu J, Yang W, Sun X, Guo Z (2018) Structure-revealing low-light image enhancement via robust retinex model. IEEE Trans. Image Process. 27(6):2828–2841

17. Liao L, Xiao J, Wang Z, Lin C-W, Satoh S (2021) Uncertainty-aware semantic guidance and estimation for image inpainting. IEEE Journal of Selected Topics in Signal Processing. 15(2):310–323

18. Liao L, Chen W, Xiao J, Wang Z, Lin C-W, Satoh S (2022) Unsupervised foggy scene understanding via self spatial-temporal label diffusion. IEEE Transactions on Image Processing. 31:3525–3540

19. Liao L, Xiao J, Wang Z, Lin C-W, Satoh S (2020) Guidance and evaluation: Semantic-aware image inpainting for mixed scenes. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16. Springer, pp. 683–700

20. Liao L, Xiao J, Wang Z, Lin C-W, Satoh S (2021) Image inpainting guided by coherence priors of semantics and textures. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6539–6548

21. Li C, Guo C, Loy CC (2021) Learning to enhance low-light image via zeroreference deep curve estimation. arXiv preprint. arXiv:2103.00860

22. Li J, Li J, Fang F, Li F, Zhang G (2020) Luminance-aware pyramid network for low-light image enhancement. IEEE Trans. Multimedia PP. (99):1–1

23. Lim S, Kim W (2020) Dslr: Deep stacked laplacian restorer for low-light image enhancement. IEEE Trans. Multimedia PP. (99):1–1

24. Li C, Wand M (2016) Precomputed real-time texture synthesis with markovian generative adversarial networks. In: European Conference on Computer Vision. Springer, pp. 702–716

25. Lore KG, Akintayo A, Sarkar S (2017) Llnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognit. 61:650–662

26. Luo X, Zeng H-Q, Wan Y, Zhang X-B, Du Y-P, Peters TM (2019) Endoscopic vision augmentation using multiscale bilateral-weighted retinex for robotic surgery. IEEE Trans. Med. Imaging 38(12):2863–2874

27. Ma Y, Liu J, Liu Y, Fu H, Hu Y, Cheng J, Qi H, Wu Y, Zhang J, Zhao Y (2021) Structure and illumination constrained gan for medical image enhancement. IEEE Trans. Med. Imaging. 40(12):3955–3967

28. McLeod AI (2005) Kendall rank correlation and mann-kendall trend test. RPackage Kendall. 602:1–10

29. Mittal A, Moorthy AK, Bovik AC (2012) No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. 21(12):4695–4708

30. Mittal A, Soundararajan R, Bovik AC (2012) Making a "completely blind" image quality analyzer. IEEE Signal Process Lett. 20(3):209–212

31. Mukherjee R, Debattista K, Bashford-Rogers T, Vangorp P, Mantiuk R, Bessa M, Waterfield B, Chalmers A (2016) Objective and subjective evaluation of high dynamic range video compression. Signal Process. Image Commun. 47(C):426–437

32. Okuhata H, Nakamura H, Hara S, Tsutsui H, Onoye T (2013) Application of the real-time retinex image enhancement for endoscopic images. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 3407–3410

33. Öztürk Ş, Özkaya U (2020) Gastrointestinal tract classification using improved lstm based cnn. Multimed Tools Appl. 79(39):28825–28840

34. Pannu HS, Ahuja S, Dang N, Soni S, Malhi AK (2020) Deep learning based image classification for intestinal hemorrhage. Multimed Tools Appl. 79(29):21941–21966

35. Ren W, Liu S, Ma L, Xu Q, Xu X, Cao X, Du J, Yang M-H (2019) Low-light image enhancement via a deep hybrid network. IEEE Trans. Med. Imaging. 28(9):4364–4375

36. Ren X, Yang W, Cheng W-H, Liu J (2020) Lr3m: Robust low-light enhancement via low-rank regularized retinex model. IEEE Trans. Image Process. 29:5862–5876

37. Ruderman DL (1994) The statistics of natural images. Network. 5(4):517–548

38. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, Cercek A, Smith RA, Jemal A (2020) Colorectal cancer statistics, 2020. CA Cancer J. Clin. 70(3):145–164

39. Venkatanath N, Praneeth D, Bh MC, Channappayya SS, Medasani SS (2015) Blind image quality evaluation using perception based features. In: 2015 Twenty First National Conference on Communications (NCC). IEEE, pp. 1–6

40. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13(4):600–612

41. Wang LW, Liu ZS, Siu WC, Lun D (2020) Lightening network for lowlight image enhancement. IEEE Trans. Image Process. 29:7984–7996

42. Wu H, Chen C, Liao L, Hou J, Sun W, Yan Q, Lin W (2023) Discovqa: temporal distortion-content transformers for video quality assessment. IEEE Trans, Circuits Syst, Video Technol

43. Xie H, Lei H, Zeng X, He Y, Chen G, Elazab A, Yue G, Wang J, Zhang G, Lei B (2020) Amd-gan: attention encoder and multi-branch structure based generative adversarial networks for fundus disease detection from scanning laser ophthalmoscopy images. Neural Networks. 132:477–490

44. Xu J, Zhang Q, Yu Y, Zhao R, Bian X, Liu X, Wang J, Ge Z, Qian D (2022) Deep reconstruction-recoding network for unsupervised domain adaptation and multi-center generalization in colonoscopy poly pdetection. Comput. Meth. Prog. Bio. 214:106576

45. Yue G, Hou C, Yan W, Choi LK, Zhou T, Hou Y (2019) Blind quality assessment for screen content images via convolutional neural network. Digital Signal Processing 91:21–30

46. Yue G, Li S, Zhou T, Wang M, Du J, Jiang Q, Gao W, Wang T, Lv J (2020) Adaptive context exploration network for polyp segmentation in colonoscopy images. IEEE Trans, Emerg, Topics Comput

47. Yue G, Han W, Jiang B, Zhou T, Cong R, Wang T (2022) Boundary constraint network with cross layer feature integration for polyp segmentation. IEEE J Biomed Health Inform. 26(8):4090–4099

48. Yue G, Cheng D, Li L, Zhou T, Liu H, Wang T (2022) Semi-supervised authentically distorted image quality assessment with consistencypreserving dual-branch convolutional neural network. IEEE Transactions on Multimedia. https://doi.org/10.1109/TMM.2022.3209889

49. Yue G, Wei P, Zhou T, Jiang Q, Yan W, Wang T (2022) Towards multi-center skin lesion classification using deep neural network with adaptively weighted balance loss. IEEE Trans Med Imaging. https://doi.org/10.1109/TMI.2022.3204646

50. Yue G, Han W, Li S, Zhou T, Lv J, Wang T (2022) Automated polyp segmentation in colonoscopy images via deep network with lesionaware feature selection and refinement. Biomedical Signal Processing and Control 78:103846

51. Yue G, Wei P, Liu Y, Luo Y, Du J, Wang T (2023) Automated endoscopic image classification via deep neural network with class imbalance loss. IEEE Transactions on Instrumentation and Measurement. 72:1–11

52. Yue G, Li S, Cong R, Zhou T, Lei B, Wang T (2023) Attention-guided pyramid context network for polyp segmentation in colonoscopy images. IEEE Transactions on Instrumentation and Measurement 72:1–13

53. Yue G, Li Y, Zhou T, Zhou X, Liu Y, Wang T (2023) Attention-driven cascaded network for diabetic retinopathy grading from fundus images. Biomedical Signal Processing and Control. 80:104370

54. Zar JH (1972) Significance testing of the spearman rank correlation coefficient. Journal of the American Statistical Association. 67(339):578–580

55. Zhong J, Wang W, Wu H, Wen Z, Qin J (2020) Polypseg: An efficient context-aware network for polyp segmentation from colonoscopy videos. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 28–294

56. Zhu J-Y, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2223–2232

57. Zhu A, Zhang L, Shen Y, Ma Y, Zhao S, Zhou Y (2020) Zero-shot restoration of underexposed images via robust retinex decomposition. In: 2020 IEEE International Conference on Multimedia and Expo (ICME).IEEE, pp. 1–6

58. Mukherjee R, Debattista K, Bashford-Rogers T, Vangorp P, Mantiuk R, Bessa M, Waterfield B, Chalmers A (2016) Objective and subjective evaluation of high dynamic range video compression. Signal Process. Image Commun. 47(C):426–437