

Deep Polyp Image Enhancement Using Region of Interest with Paired Supervision

Dongjin Huang^a, Jinhua Liu^{a,*}, Yongsheng Shi^a, Canlin Li^b, Wen Tang^c

^a*Shanghai Film Academy, Shanghai University, Shanghai 200072, China*

^b*School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450001, China*

^c*Department of Creative Technology, University of Bournemouth, Poole BH12 5BB, United Kingdom*

Abstract

Endoscopic medical imaging in complex curved intestinal structures are prone to uneven illumination, low contrast and lack of texture information. These problems may lead to diagnostic challenges. This paper described the first supervised deep learning based image fusion framework to enable the polyp region highlight through a global image enhancement and a local region of interest (ROI) with paired supervision. Firstly, we conducted a dual attention based network in global image enhancement. The Detail Attention Maps was used to preserve more image details and the Luminance Attention Maps was used to adjust the global illumination of the image. Secondly, we adopted the advanced polyp segmentation network ACSNet to obtain the accurate mask image of lesion region in local ROI acquisition. Finally, a new image fusion strategy was proposed to realize the local enhancement effect of polyp image. Experimental results show that our method can highlight the local details of the lesion area better and reach the optimal comprehensive performance with comparing with 16 traditional and state-of-the-art enhancement algorithms. And 8 doctors and 12 medical students were asked to evaluate our method for assisting clinical diagnosis and treatment effectively. Furthermore, the first paired image dataset LHI was constructed, which will be made available as an open source to research communities.

Keywords: Polyp image enhancement, Image fusion, Region of interest, Supervised learning

1. Introduction

Polyps within the colon are one of the leading causes of prevalent cancer. Endoscopy is a standard visual diagnostic procedure for detecting intestinal polyps [1]. The intestinal polyp image directly obtained by endoscope is prone to some problems, such as highlight reflection [2], insufficient lighting covering the region of interest [3], and low contrast [4]. These problems will decrease the accuracy of manual examination. Image enhancement can provide detailed edge and texture information for clinical analysis. Therefore, the study of intestinal polyp image enhancement is very important to help colonoscopists reduce the rate of misdiagnosis and missed diagnosis.

Conventional image enhancement methods, including histogram equalization-based methods: CLAHE [5], BPDHE [6], and Retinex-based methods: MSR [7], NPE[8], SRIE [9], MF [10], LIME [11], applied to low-light and non-uniform illumination intestinal polyp images are prone to have problems such as over-enhancement of the brighter local areas, insufficient enhancement of lesion areas and unnatural enhancement results. In recent years, the method of deep learning has performed well in the field of natural image enhancement [12, 13, 14, 15]. Due to no public paired dataset

*Corresponding author

Email address: jinhua1427@hotmail.com (Jinhua Liu)

for intestinal polyp image enhancement, that seriously affects the development of supervised learning in intestinal polyp image enhancement. Scholars at home and abroad have also tried to apply unsupervised neural network framework to the task of endoscopic image enhancement [16, 17]. However, these learning methods are usually difficult to obtain high-quality image enhancement results. Therefore, it is of great significance to construct paired intestinal polyp image enhancement dataset.

In this paper, we constructed the first paired intestinal polyp image dataset LHI, including 2108 low-quality and high-quality synthetic image pairs, to promote the research of methods relying on paired images for training. And we proposed the supervised learning based image fusion framework to enhance local lesion area for attracting doctor’s more attention on accurate polyp region. The method was divided into three steps. In the first step, a new deep learning global image enhancement method, called Encoder Dual Attention U-Net Network (EDAU-Net), was proposed to recover the structure information that is difficult to observe in low-quality images. In order to preserve more image details and adjust the global image illumination, Detail Attention Maps and Luminance Attention Maps were added to the encoder part of the U-Net. In the second step, we used the advanced polyp segmentation network ACSNet [18] to obtain the ROI mask image of the polyp coverage lesion that the doctor was interested in. In the third step, the global enhanced image and the ROI mask image were fused by using the new fusion strategy to weaken the background area for achieving the local enhancement effect of the lesion area.

The contributions of this work are summarized as follows:

- 1) We constructed the first synthetic paired low-quality/high-quality intestinal polyp images dataset LHI, which can be downloaded at this link: <https://drive.google.com/drive/folders/1Lo7ctWolNgwNvkBXUcVMGUGYx6qF6Sbe?usp=sharing>.
- 2) We proposed the first supervised learning based image fusion framework to enhance the detailed information of the lesion area by fusing the global enhanced image and the ROI mask image.
- 3) We proposed a novel end-to-end global image enhancement network architecture EDAU-Net. The Detail Attention Map was proposed and added to the encoder of the network to highlight texture detail features. The Luminance Attention Map was introduced into the encoder of the network to improve the global image illumination.
- 4) Our method outperformed sixteen traditional and state-of-the-art algorithms on new dataset, and could obtain enhanced polyp images with superior quality.

The remainder of this paper is organized as follows. In Section 2, we introduce the related work on image enhancement. The proposed polyp image enhancement method is described in Section 3. We give the experiments and results in Section 4. And some issues need to be discussed in Section 5. Section 6 is the conclusion.

2. Related Work

2.1. Conventional Image Enhancement Methods

Currently, the most commonly used algorithms are methods based on Retinex theory and methods based on histogram equalization.

Retinex theory: these methods are often used in endoscopic image enhancement. Okuhata et al. [19] proposed a real-time image enhancement algorithm for gastroscopy. The algorithm combines Single Scale Retinex (SSR) [20] theory with gamma correction to enhance the brightness and contrast of endoscope image. Retinex focuses on local enhancement, which solves the problem of insufficient brightness enhancement of local area of gastroscopy image by global enhancement algorithm to a certain extent. However, for gastroscopy image with uneven illumination, it may lead to the problem of excessive image enhancement. In addition, Gamma correction improves image brightness by expanding the low gray value region and compressing the high gray value region. Therefore, the

contrast enhancement effect of endoscope image is not obvious. Multi-Scale Retinex (MSR) [7] algorithm can maintain the essential characteristics of the image in the enhancement process. However, when MSR is used to enhance the endoscope image, it will over-enhance the highlight area, and there will be color distortion, unclear texture and so on. Literature [21] used the improved MSR model to solve the problems of non-uniform illumination and color distortion of laparoscopic images. However, the problems of over-enhancement of high brightness areas and edge blur still exist. At present, many improved image enhancement algorithms based on Retinex theory have been published [8, 9, 10, 11, 22, 23]. The common point of these methods is that they first accurately estimate the illumination component of the image through manual design and parameter adjustment, and then obtain the enhanced image directly or indirectly according to Retinex theory. Because these methods depend on the illumination component, when they are used to process the brighter areas of the endoscope image with non-uniform illumination, there will be excessive enhancement, and the contrast enhancement effect is not significant. Therefore, the defects of algorithms based on Retinex theory also exist in this kind of methods.

Histogram equalization: these methods are widely used for contrast enhancement due to its simplicity and effectiveness [24]. However, when it is used to enhance the endoscope image, there will be some problems, such as detail losses, noise amplification, over-enhancement and so on [25]. Literature [5] proposed a contrast-limited adaptive histogram equalization (CLAHE) algorithm. CLAHE uses threshold clipping histogram to prevent over-enhancement, but color distortion occurs when used to enhance endoscopic images. Later, a series of improved algorithms for histogram equalization were produced [6, 26], which can effectively enhance contrast while maintaining image brightness. However, they are only suitable for specific scenes with moderate brightness, and the effect of brightness adjustment for low-illuminance images is not good. Literature [27] designed an endoscope image enhancement method using histogram information. Firstly, the endoscope image was processed by median filter to highlight the features of interest in the image, and then the histogram information was corrected to improve the image contrast. Although this technology enhances the overall contrast of the image, the endoscopic image requires excellent edge contrast and sufficient brightness.

Conventional image enhancement methods can enhance the natural image in a specific scene, but their nonlinear expression ability is not strong, and cannot well solve the problems of uneven illumination, low brightness and contrast, and lack of color in endoscopic images. Therefore, their practical application value is limited. In recent years, the deep learning method has good effect and high efficiency in image processing. Therefore, it is of great significance to apply the deep learning method to endoscopic image enhancement.

2.2. Deep Learning-based Image Enhancement Methods

Recently, deep learning methods have become the mainstream methods in image enhancement field, which are mainly consist of the supervised methods and unsupervised methods.

Supervised learning: at present, the research on the enhancement of uneven illumination, weak illumination and low-contrast image based on deep learning methods focuses on natural images in atmospheric environment [15, 28]. Most deep learning methods are fully supervised, and these methods require the construction of pairs of low illumination image and clear image datasets. Then, the mapping relationship between paired images is learned through various models to achieve the purpose of image enhancement. LLNet [29] proposed a data training method to simulate low illumination environment by using gamma check and artificially adding Gaussian noise, simultaneously enhance image brightness and reduce noise. Chen et al. [30] constructed a low exposure image dataset generated according to the camera exposure time, and designed a fully convoluted end-to-end network to achieve image enhancement under extreme low-light conditions. Inspired by Retinex theory, Retinex-Net [12] decomposes the low exposure image into illumination component and reflection component, and then uses the enhancement network to enhance the image brightness. This method can achieve the effect of improving the brightness of low exposure images, but it needs to be improved in denoising. KinD [13] has greatly improved compared with Retinex-Net in terms

of denoising and image enhancement effects. Although the above methods can achieve better results compared with the traditional image enhancement methods, these methods need to be trained with strictly aligned low/high quality images. For endoscopic images, due to the limitations of imaging instruments and imaging environment, it is difficult to obtain the low quality image and high quality image pair. Therefore, there are few reports on using supervised convolutional neural network to enhance endoscopic images.

Unsupervised learning: in order to solve the problem that it is difficult to obtain paired image training data, Jiang et al. [14] proposed for the first time to use unpaired low/normal brightness images to train the image enhancement network EnlightenGAN. This training strategy eliminates the dependence of deep learning method on paired training data. However, EnlightenGAN still needs to directly learn the mapping relationship from low exposure image to normal image end-to-end when there is a reference image. Therefore, reference images need to be selected carefully. Guo et al. Proposed unsupervised learning methods Zero-DCE [31] and Zero-DCE++ [32]. By setting a series of non-reference loss functions, the authors make the network end-to-end training without any reference images. Although the above unsupervised learning method can be adapted to the lack of datasets in the field of endoscopic image enhancement, this learning method is usually difficult to obtain high-quality image enhancement results.

Although deep learning-based image enhancement methods have achieved remarkable results in the field of natural images, a large number of paired datasets are difficult to obtain in the human gastrointestinal internal environment. Therefore, it is necessary to construct the endoscope image dataset. In this way, the deep learning network can be used to extract the endoscopic image features in

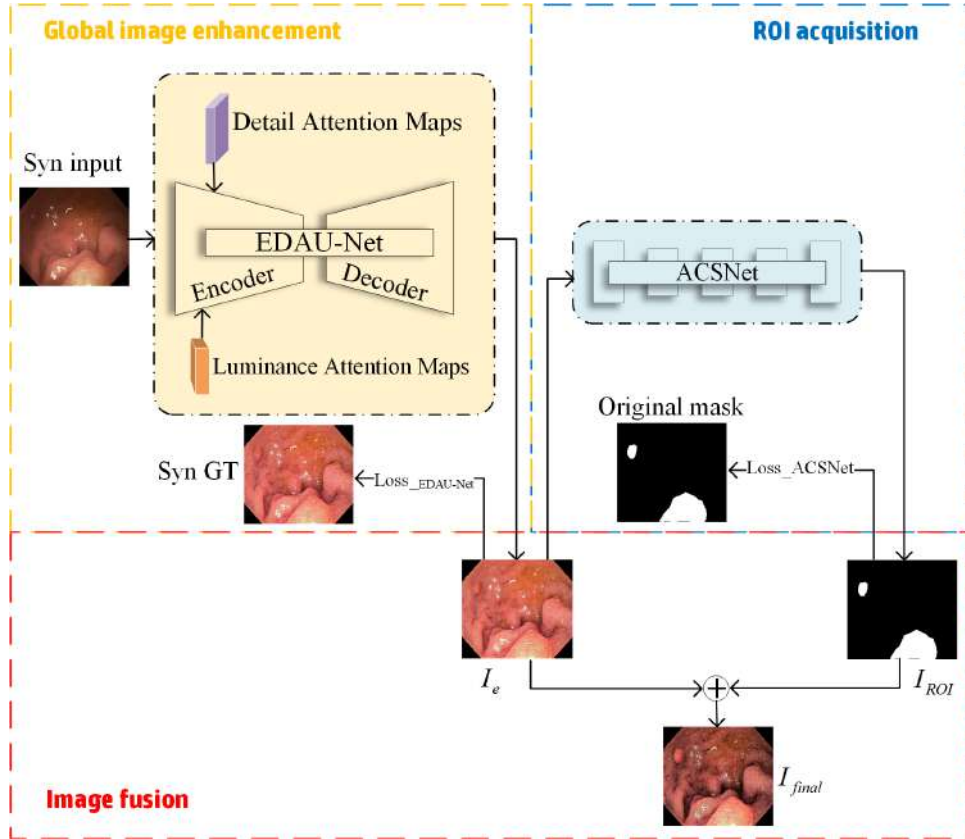


Figure 1: Flow chart of our proposed image fusion framework for the ROI of intestinal polyp images.

the data set, so that the network can be applied to the endoscopic image enhancement in real scenes. The existing deep learning-based image enhancement methods aim to achieve global enhancement. Due to the particularity of the internal environment, existing natural image enhancement algorithms are not completely suitable for intestinal polyp image.

3. Methods

In this section, we mainly introduce the image fusion framework for enhancing the local details in the polyp lesion area. The flow chart of that is shown in Fig. 1, which consists of three main modules: Global image enhancement, ROI acquisition and Image fusion.

3.1. Dataset Production

At present, there is no dataset containing both low-quality and their corresponding high-quality intestinal polyp images. It is very difficult or even impossible to capture low light images and normal light images in the same moment of the scene in the intestine. Therefore, we constructed a synthetic intestinal polyp dataset LHI, which includes low-quality images and their corresponding synthetic high-quality image pairs.

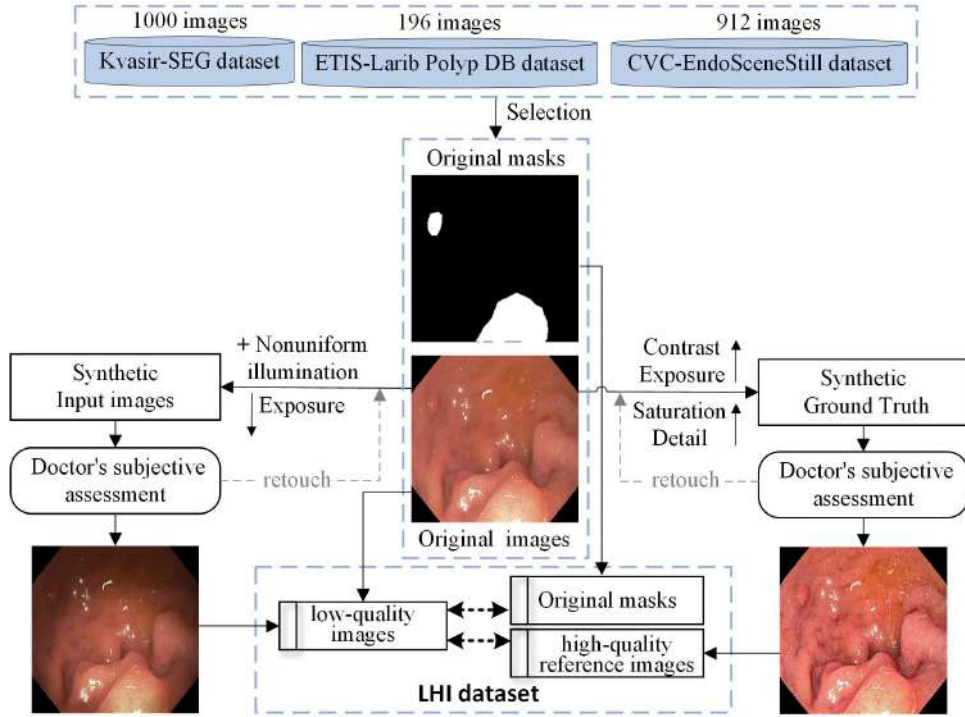


Figure 2: Pipeline of constructing the proposed dataset LHI. We selected the real intestinal polyp image and the corresponding segmentation mask from the public dataset: Kvasir-SEG [33], ETIS-Larib Polyp DB [34] and CVC-EndoSceneStill [35], and the exposure of real intestinal polyp images were reduced and non-uniform lighting effect was added to generate low illumination non-uniform lighting low-quality image (Synthetic Input image). At the same time, the real intestinal polyp image was enhanced by exposure correction, saturation adjustment and contrast/detail amplification, so as to generate a high-quality reference image (Synthetic Ground Truth).

In this paper, 2108 real intestinal polyp images and corresponding segmentation mask (Ground Truth) were collected from the open source medical imaging datasets Kvasir-SEG dataset [33], ETIS-Larib Polyp DB [34] and CVC-EndoSceneStill [35] related to gastrointestinal. This ground truth consists of a mask corresponding to the region covered by the polyp in the image. All these photos

were resized to 256×256 pixels. We recruited two experts majoring in digital media technology in the university to use Adobe Lightroom to adjust the real intestinal polyp images. They had received extensive training. The low-quality intestinal polyp images studied in this paper have the characteristics of uneven illumination, low illumination, low contrast and lack of texture information. In the collected dataset, all images were low contrast and lack texture information, while 775 images were well illuminated. In order to improve the availability of dataset images, it was necessary to change the illumination of well-illuminated images. one of the experts first fully analyzed the characteristics of a large number of real clinical low-quality polyp images, then carefully checked each collected image. The well-illuminance images were edited by Lightroom to reduce the exposure and add non-uniform lighting effect for obtain low-quality image (Synthetic Input image). Another expert manually retouched each collected image to obtain a high-quality reference image (Synthetic Ground Truth). In order to ensure the validity of the data set, we require the Synthetic Input image to be visually similar to the real clinical polyp image, and the Synthetic Ground Truth to achieve a visually pleasing effect for the doctor. Therefore, all Synthetic Input images and Synthetic Ground Truth in the dataset were subjectively evaluated by medical college volunteers on the same monitor. If Synthetic Input image is not similar to real clinical polyp image, the corresponding expert would be required to reprocess the image until the volunteers are satisfied. For Synthetic Ground Truth image sequences that did not produce satisfactory output, experts would be required to retouch the image to select the best output for each image. See Fig. 2 for the process of generating composite image pairs. Finally, we randomly divided the images in the dataset into two subsets: 1489 images for training and 619 images for testing.

3.2. Global Image Enhancement

In order to globally enhance the detail, contrast, and color of intestinal polyp images for more accurate ROI acquisition and Image fusion, we proposed EDAU-Net by improving the U-Net [36] with excellent feature extraction ability. Firstly, to address the problem of detail information loss when images are enhanced by U-Net, we proposed a new Detail Attention Map. And we added it to the encoding stage of the U-Net for highlighting detail features and reducing the loss of some details in the process of down-sampling and up-sampling of U-Net. Secondly, in order to better compensate for the influence of non-uniform illumination of the image by point light sources in the curved intestine, we introduced the Luminance Attention Map [14] in the encoding stage of the U-Net to improve the global image illumination. Thus, we achieved the effect of both effectively enhancing the local darker areas and avoiding the generation of overexposure of the local brighter areas. The network structure of EDAU-Net model is shown in Fig. 3.

As shown in Fig. 3, EDAU-Net consists of four parts: input, encoder dual attention, decoder, and output. The network input is the feature map with 4 channels and size of 256×256 obtained by concatenating a 3 channels RGB low-quality image and a single-channel Luminance Attention Map. The encoder dual attention and decoder modules consist of 8 down-sampling layers and 8 up-sampling layers respectively. In the encoding dual attention stage, the 8 down-sampling convolution layers use the filter of the size of 4×4 and the Leaky LeRU activation function to extract multi-scale features from the input image. The number of output layers is 64, 128, 256, 512, 512, 512, 512, and 512 respectively. The Luminance Attention Map is multiplied by the feature map output in the down-sampling process, and has the same size as the corresponding feature map. The Detail Attention Map is spliced with the feature map output in the down-sampling process, and has the same size as the corresponding feature map. In the decoding stage, the size of the filter used by the first 7 up-sampling convolution layers is 4×4 and the Leaky LeRU activation function is adopted. The last deconvolution layer uses the Tanh activation function to produce the globally enhanced image with more complete details. The number of output layers is 512, 512, 512, 512, 256, 128, 64, and 3 respectively. The network output is a globally enhanced image 3 channels and the size of 256×256 . The network parameters of EDAU-Net are shown in Table 1.

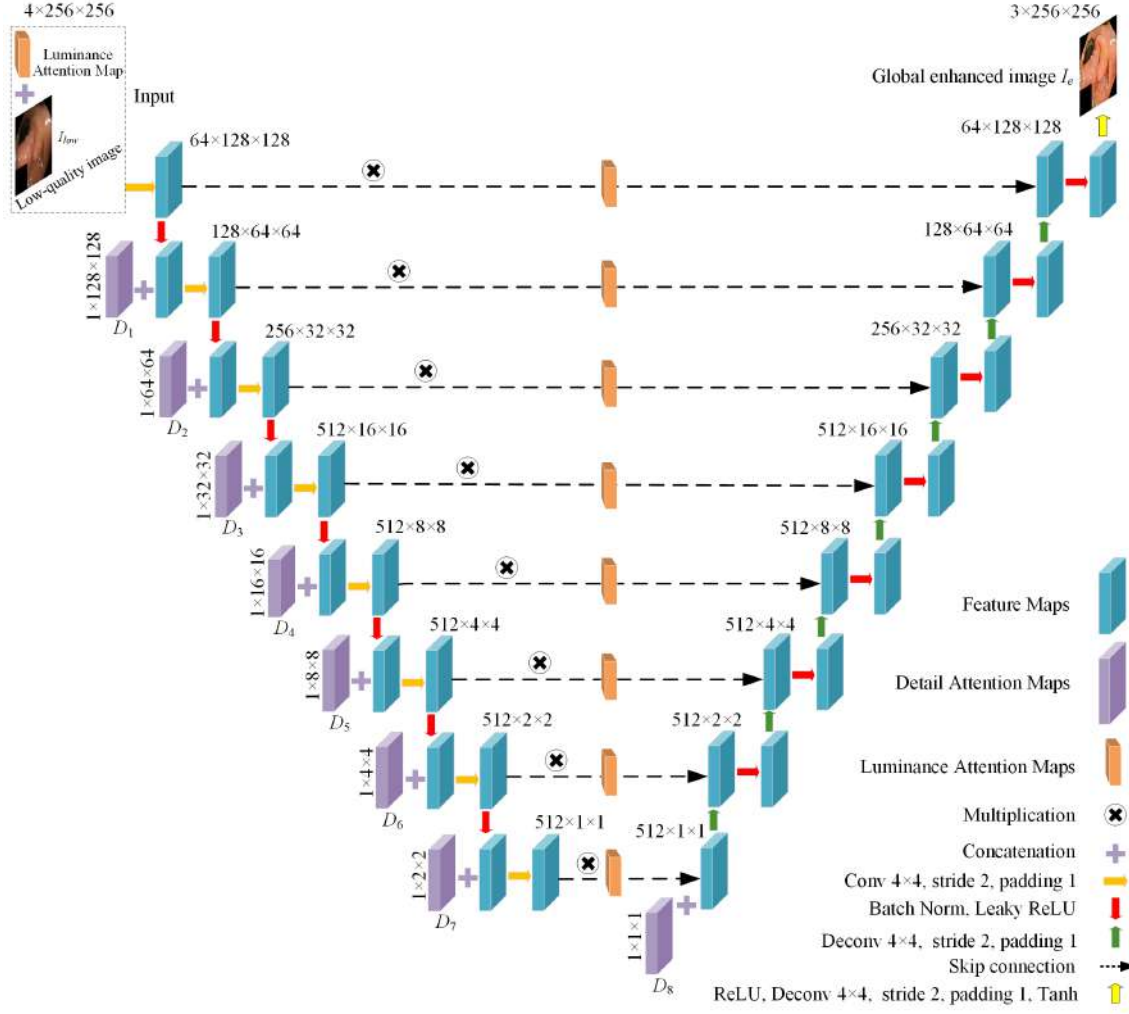


Figure 3: Network structure of EDAU-Net model.

3.2.1. Detail Attention Map

In the process of down-sampling and up-sampling of U-Net, it is inevitable that the image will lose details. Canny [37] is not easily disturbed by noise, and can obtain fine and accurate edge images. We added the Detail Attention Maps obtained by Canny, that is, the edge image, to the encoding structure for enhancing the network's ability to learn detail information. In the encoder, the Detail Attention Map is concatenated with the feature map after Leaky ReLU and participates in the training of the model. The Detail Attention Map is calculated as follows.

$$\begin{cases} D_i = f_{Mp}(D_{i-1}) \\ D_1 = f_{Mp}(I_{canny}) \\ I_{canny} = f_{canny}(I_{low}) \end{cases} \quad i = 2, 3, 4, 5, 6, 7, 8 \quad (1)$$

where I_{low} is the original low-quality image. f_{canny} indicates the image is processed using the Canny. f_{Mp} means Max-pooling is used for the edge image. $D_1, \dots, D_8$ constitute the Detail Attention Maps. As shown in Fig. 3, we spliced the Detail Attention Map to the feature map after Leaky ReLU of the corresponding size in the encoder for training, so the Detail Attention Maps needs to be consistent with the size of the corresponding convolution layer feature map. The acquisition of the detailed edge

Table 1: The specific parameter settings of EDAU-Net. Concat represents Concatenation, Conv represents convolution, and Deconv represents deconvolution.

Module	No.	Input size	Operation	Output size	Filter size	Filter number	Stride	Padding
Input	1	3*256*256 1*256*256	Concat	4*256*256	—	—	—	—
Encoder Dual Attention	1	4*256*256	Conv	64*128*128	4*4	64	2	1
	2	65*128*128	Conv	128*64*64	4*4	128	2	1
	3	129*64*64	Conv	256*32*32	4*4	256	2	1
	4	257*32*32	Conv	512*16*16	4*4	512	2	1
	5	513*16*16	Conv	512*8*8	4*4	512	2	1
	6	513*8*8	Conv	512*4*4	4*4	512	2	1
	7	513*4*4	Conv	512*2*2	4*4	512	2	1
	8	513*2*2	Conv	512*1*1	4*4	512	2	1
Decoder	8	513*1*1	Deconv	512*2*2	4*4	512	2	1
	7	1024*2*2	Deconv	512*4*4	4*4	512	2	1
	6	1024*4*4	Deconv	512*8*8	4*4	512	2	1
	5	1024*8*8	Deconv	512*16*16	4*4	512	2	1
	4	1024*16*16	Deconv	256*32*32	4*4	256	2	1
	3	512*32*32	Deconv	128*64*64	4*4	128	2	1
	2	256*64*64	Deconv	64*128*128	4*4	64	2	1
	1	128*128*128	Deconv	3*256*256	4*4	3	2	1
Output	1	—	—	3*256*256	—	—	—	—

image of the corresponding size (i.e. $D_1, \dots, D_8$) is shown in Fig. 4. We used Max-pooling to process the edge image to achieve the effect that the size of the edge image decreases with the number of down-sampling increases while retaining the main information of the image.

3.2.2. Luminance Attention Map

Intestinal endoscopic images have problems of uneven illumination and insufficient illumination. The self-regularized attention map proposed in literature [14] can effectively adjust the image illumination. We introduce it into the encoder of our proposed network called Luminance Attention Map to make the network pays more attention to low illumination areas and avoid over-enhancement of high brightness areas. The generation process of Luminance Attention Map is shown in Fig. 5. We convert the low-quality image I_{low} in RGB color space to a single-channel gray image I_g , normalize it to $[0, 1]$, and then use $1 - I_g$ (element-wise difference) as our Luminance Attention Map. As shown in Fig. 3, we used low-quality image and Luminance Attention Map as input, rather than just the low-quality image directly into the network. The motivation is that the Luminance Attention Map

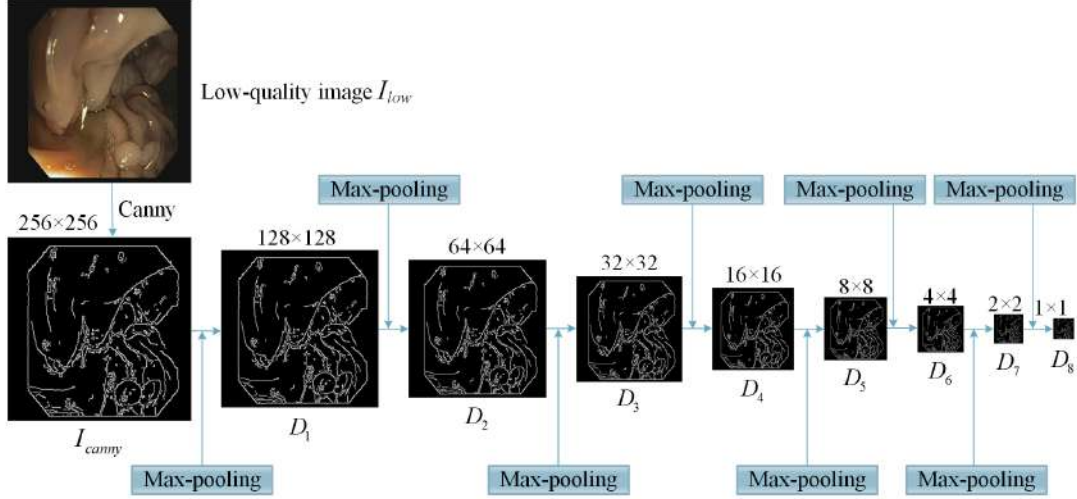


Figure 4: The Detail Attention Map down-sampling process.

provides more priori information for the network. Then, the Luminance Attention Maps is multiplied by the feature map output in the down-sampling process, and has the same size as the corresponding feature maps. We also used Max-pooling to change size of the Luminance Attention Maps for keeping the same size of the corresponding convolution feature maps.

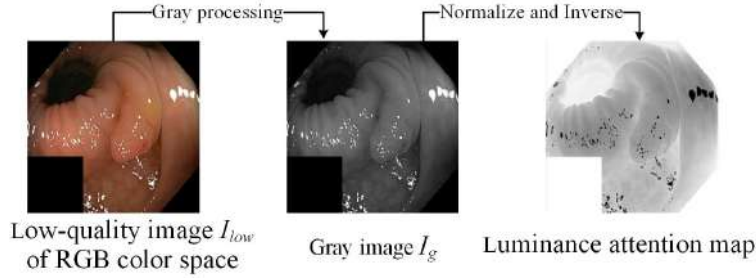


Figure 5: Detailed generation process of Luminance Attention Map.

3.2.3. Loss function

In order to train the network parameters, we used four loss functions: mean abs error loss(L_1), mean square error loss(L_2), structural similarity loss ($L_{Structure}$) and perceptual loss ($L_{Perceptual}$) to characterize the difference between the predicted image and Ground Truth. The overall loss function for training EDAU-Net was written as:

$$L_{EDAU-Net} = \lambda_1 L_1 + \lambda_2 L_2 + \lambda_3 L_{Structure} + \lambda_4 L_{Perceptual} \quad (2)$$

where λ_1 , λ_2 , λ_3 , and λ_4 denote the weight value for each loss term, respectively. We set their values to $\lambda_1=1$, $\lambda_2=1000$, $\lambda_3=1$, and $\lambda_4=1$. The details of the four loss functions are given below.

Mean abs error loss and mean square error loss. L_1 loss is widely used in the field of image enhancement. It can reconstruct the clear edge of the image. However, it cannot effectively capture the high-frequency information of the image, resulting in the overall prediction image is too smooth and the visual experience is general. Although L_2 loss can better reconstruct the high-frequency information of the image, it has the disadvantage of producing artifacts. Therefore, in order to

better obtain the high-frequency and low-frequency information of the image without producing artifacts, we introduce L_1 loss and L_2 loss at the same time. The loss function of L_1 and L_2 can be expressed as:

$$L_1 = E(I_e, I_{gt}) [\|I_{gt} - I_e\|_1] \quad (3)$$

$$L_2 = E(I_e, I_{gt}) [\|I_{gt} - I_e\|_2]^2 \quad (4)$$

where I_e and I_{gt} are the enhanced image output by EDAU-Net and the corresponding Ground Truth respectively.

Structural similarity loss. The structural similarity of images is also an important index to measure the quality of image enhancement. While obtaining the global information, the network also learns the structure information through the convolution of feature maps of multiple sizes. Therefore, we use the MultiScale Structural Similarity (MS-SSIM) [38] quality evaluation method as the loss function to maintain the image structure information and avoid ambiguity. Finally, the structural similarity loss function of $L_{Structure}$ is defined as follows:

$$L_{Structure} = 1 - l(I_{gt}, I_e)_m^{\alpha m} \prod_i^m c(I_{gt}, I_e)_i^{\beta i} s(I_{gt}, I_e)_i^{\theta i} \quad (5)$$

where l , c , and s roughly represent the luminance, contrast and structure, respectively. i is the pixel coordinate. m is the total number of pixels. α , β , and θ are parameters to adjusting the importance of the three components l , c , and s . To simplify parameter selection, we usually set $\alpha = \beta = \theta = 1$.

Perceptual loss. In order to make the generated image by EDAU-Net have more detail information, this paper introduces perceptual loss $L_{Perceptual}$. We use the feature map generated by the VGG-16 network pre-trained on the ImageNet [39] dataset to calculate the perceptual loss, so as to measure the global difference between the enhanced image I_e and the corresponding Ground Truth. $L_{Perceptual}$ is defined as follows:

$$L_{Perceptual} = \frac{1}{CWH} \sum_{a=1}^C \sum_{b=1}^W \sum_{c=1}^H \|\Phi(I_e)_{a,b,c} - \Phi(I_{gt})_{a,b,c}\| \quad (6)$$

where C , W , H are the channel, height and width of the feature map respectively.

3.3. ROI Acquisition

In order to locate the ROI in the image to the doctor, we needed to segment and highlight the polyps region for attracting his attention. We used the advanced intestinal polyp segmentation network ACSNet [18] to obtain the segmentation results of polyp position. The segmentation mask of polyp image locates the region of interest to the doctor, which is defined as ROI. Very cleverly, our dataset LHI not only has synthetic image pairs that can be used for image enhancement model training and testing, but also has the segmentation mask of polyp that can be used for image segmentation model training and testing. ACSNet has high accuracy in polyp boundary location, specially is very robust to some complex situations (such as polyp region sizes and image brightness changes). ACSNet was trained and tested on the dataset LHI. The accuracy of ACSNet on the testing set was very high (98.15%), and the false positive rate was very low (1.1%). This shows that the use of ACSNet for the aided diagnosis of colorectal polyps can effectively detect polyps and reduce the rate of misdiagnosis. In the ROI acquisition module, the ACSNet network input is the globally enhanced image predicted by EDAU-Net, and the network output is a single channel mask image. In addition, we can judge whether the input image contains polyps according to whether there are white areas in the mask image. We used ACSNet to predict mask images of 300 intestinal endoscopy images with and without polyps collected from a Grade-A tertiary hospital in Zhengzhou city and invited doctors to evaluate whether polyps can be accurately determined. The feedback results show that

ACSNet can effectively distinguish between polyp-free images and polyp images, and can perform accurate polyp segmentation for polyp images. Therefore, ACSNet is used to determine whether the input image is to be globally enhanced or further locally enhanced. That is, if there is no polyp in the image, image fusion will not be performed; if there is polyp in the image, the image fusion operation in Section 3.4 will be performed.

3.4. Image Fusion

The proposed EDAU-Net can obtain images with good global enhancement effect. Through ACSNet, the target area and the background area in the enhanced images can be segmented with high accuracy. For highlighting the details of the target area, we proposed a new image fusion strategy to weaken the background area. We used gamma correction to process the non-lesion area to achieve the local enhancement effect of polyp image. The ROI image I_{ROI} obtained by ACSNet and the global enhanced image I_e obtained by the proposed EDAU-Net are fused to obtain the final enhanced image I_{final} . The specific expression is as follows:

$$I_{final} = I_e \times \frac{I_{ROI}}{255} + f_{gamma}^\gamma(I_e) \times (255 - I_{ROI}) \quad (7)$$

where $f_{gamma}(I_e)$ is the image generated by gamma correction of I_e . Considering the influence of background area on doctors, gamma correction is used to eliminate the influence of disturbing doctors' background area details. Gamma correction is used to correct image brightness by non-linear operation on the pixels of image I_e [40]. γ is the correction parameter, which can control the degree of image stretching. When $\gamma=1$, $f_{gamma}(I_e)$ is the true representation of the input image I_e . If $\gamma < 1$, $f_{gamma}(I_e)$ will be brighter than I_e . If $\gamma > 1$, $f_{gamma}(I_e)$ will be darker than I_e . This paper sets $\gamma=2.5$. Fig. 6 shows the comparison of effects before and after polyp image fusion. As can be seen from Fig. 6(b) and Fig. 6(d), after polyp image fusion, the interval between the pixel value distribution area in the background area and the pixel value distribution area in the polyp area becomes larger, and the image contrast increases.

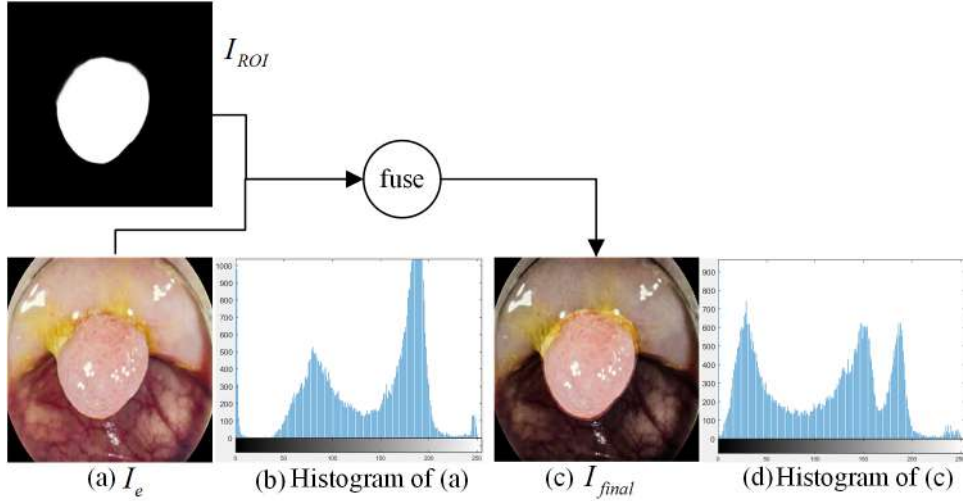


Figure 6: Effect Comparison before and after image fusion strategy. (a) Global enhanced image; (b) Histogram corresponding to global enhanced image; (c) Final local enhanced image after image fusion; (d) Histogram corresponding to final local enhanced image.

4. Experiments and Results

4.1. Experiments and Results of the Proposed EDAU-Net

We use Python programming language and pytorch1.5.0 as the deep learning framework to build, train and test the network on an Inter(R) Xeon(R) E5-2620 CPU, 2.10GHz processor, 64GB RAM, and a Nvidia Titan Xp GPU. We train our model using Adam optimizer, and set the initial value of network learning rate to 0.0002. The batch size is set to 32, and the iteration number is 200.

In order to verify the effectiveness of the proposed EDAU-Net, we compare EDAU-Net with seven most advanced image enhancement methods based on deep learning, namely RetinexNet [12], MBLLEN [28], KinD [13], DeepUPE [15], Zero-DCE [31], Zero-DCE++ [32], EnlightenGAN [14] and nine traditional image enhancement methods, namely BPDHE [6], Dong [22], AGCWD [41], NPE [8], SRIE [9], MF [10], LIME [11], BIMEF [42], RRM [23]. We conduct qualitative and quantitative analysis on the dataset LHI constructed in this paper.

4.1.1. Qualitative Evaluation

Due to the limitation of page space, we randomly selected two images of different scenes in the testing set of our dataset LHI for display. The experimental results are shown in Fig. 7 and Fig. 8.

Fig. 7(a) is the original input image. According to Fig. 7(a), the original image is a low exposure image, and the polyp is located in the dark area. Fig. 7(b)-Fig. 7(j) are the results of processing Fig. 7(a) by traditional image enhancement methods. Fig. 7(k)-Fig. 7(r) are the results of processing Fig. 7(a) by image enhancement methods based on deep learning. From Fig. 7, we can see that BPDHE causes some problems such as underexposure and poor color processing effect. The method proposed by Dong et al. can improve the brightness of image, but it produces some artifacts when the processing in reflective areas. So this method is not suitable for the intestinal polyp image. The sharpness of intestinal polyp image is restored well by AGCWD and LIME, but AGCWD does not significantly enhance the darker area and LIME over-enhances the lighter area, which leads to the loss of image details. The enhancement effects of NPE and MF were not obvious. BIMEF has a poor processing effect on image contrast and color, resulting in a slight color difference in the result image. RRM and DeepUPE can better solve the problem of low brightness of the endoscopic image, but the enhanced image has the problem of low contrast. Retinex-Net will have very obvious patches in different colors and light transitions, and the visual effect is poor. MBLLEN causes excessive color enhancement in the image. the sharpness of the image after KinD enhancement is good, the overall brightness of the image is dark. The enhanced images of Zero-DCE and Zero-DCE++ appear abnormal green or yellow. Although EnlightenGAN solves the color distortion problem of Zero-DCE and Zero-DCE++, EnlightenGAN will produce obvious sawtooth noise distortion at the edge of the light. In contrast, our proposed EDAU-Net has better visual effect, because it can well restore the texture, detail and color of intestinal polyp image without producing artifacts.

Fig. 8 shows the enhancement effects of different enhancement methods on non-uniform illumination images. Due to the fluctuation of the intestine, the original image Fig. 8(a) contains both dark and bright areas. Fig. 8(b)-Fig. 8(j) are the results of processing Fig. 8(a) by traditional image enhancement methods. Fig. 8(k)-Fig. 8(r) are the results of processing Fig. 8(a) by image enhancement methods based on deep learning. BPDHE still has no significant effect on non-uniform illumination image enhancement. The method proposed by Dong et al. and LIME can enhance the details of dark areas to a certain extent. However, LIME has the problem of over-enhancement, and the method proposed by Dong et al. has black artifacts on the reflective edges. AGCWD can cause over-enhancement of the brighter area and blurring of blood vessel edges, and the enhancement effect of polyp in the dark area is not obvious. NPE, SRIE, MF and BIMEF have the phenomenon of loss of vascular edge details and insufficient brightness enhancement in local areas. RRM not only leads to the loss of image edge details, but also produces over-enhancement in the area where the polyp with high brightness is located. Retinex-Net has obvious loss of details, and there is an undesirable enhancement in some edge structures of the image, and it will produce obvious sawtooth noise distortion at the edge of the light. Although MBLLEN, KinD and EnlightenGAN produce good effects

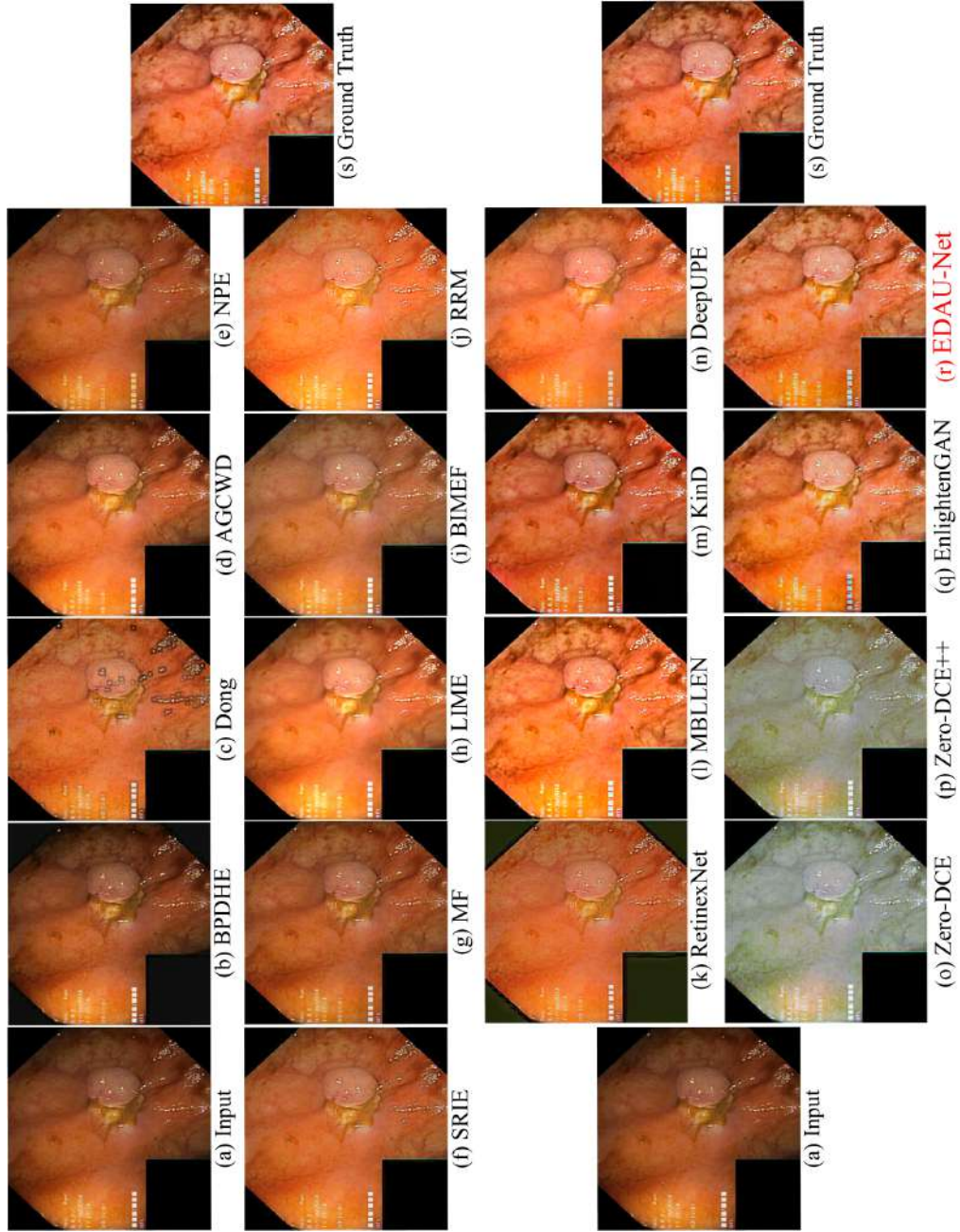


Figure 7: Visual comparison with state-of-the-art low-light image enhancement methods for polyp image A in the testing set of LHI.

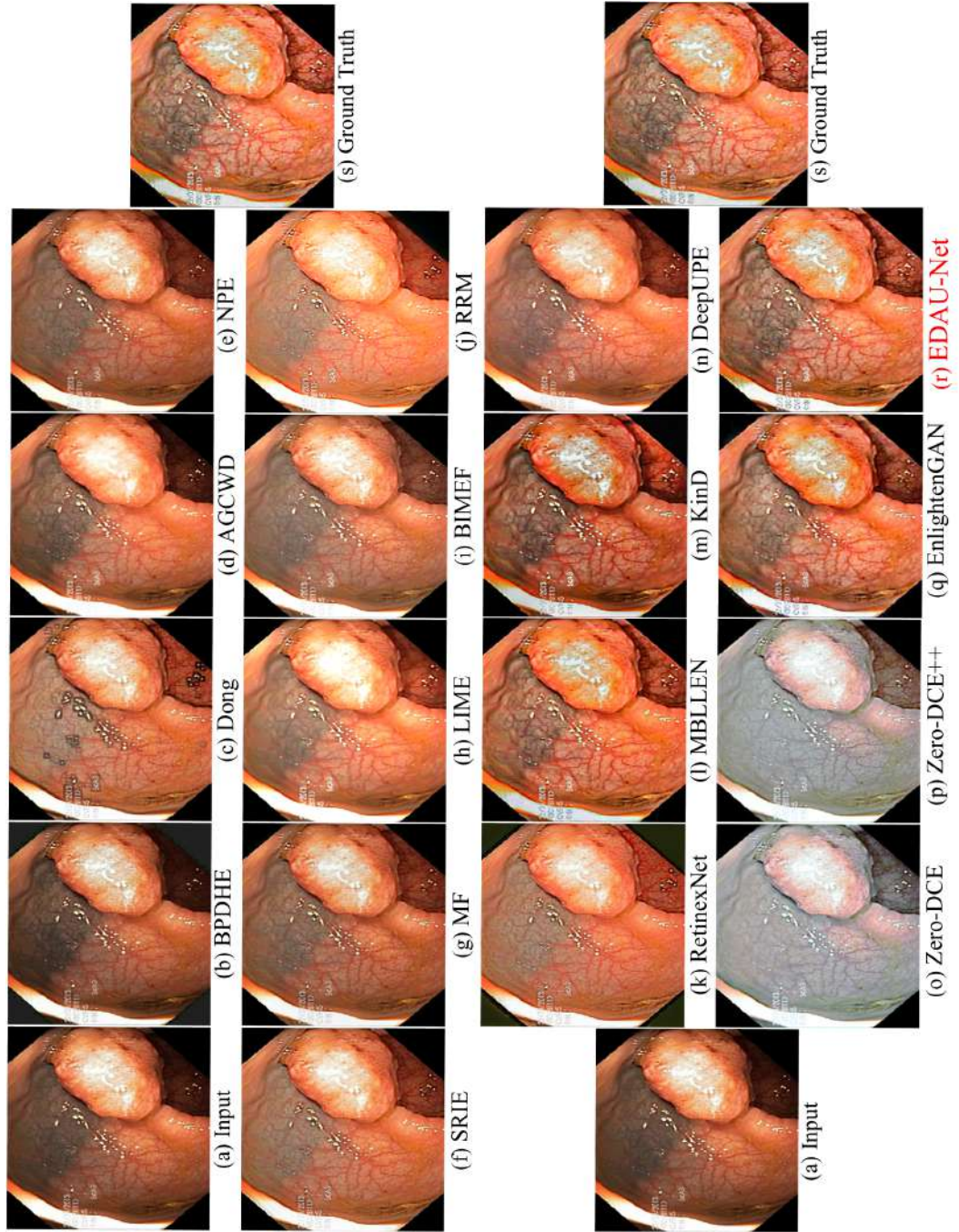


Figure 8: Visual comparison with state-of-the-art low-light image enhancement methods for polyp image B in the testing set of LHI.

in contrast and detail enhancement, they cause color distortion. DeepUPE has no significant change to the bright region and can significantly improve the brightness of the dark region, but the overall image is blurry. Zero-DCE and Zero-DCE++ not only have serious color distortion, but also reduce the visibility of the image. The proposed EDAU-Net can significantly enhance the darker areas of the image, and still maintain richer texture details for the brighter areas of the original image. The image contrast has also been significantly improved, and it is more consistent with the original color of the object in the image.

4.1.2. Quantitative Evaluation

In order to effectively and objectively evaluate the proposed EDAU-Net, three full reference image quality assessment indexes including Peak Signal-to-Noise Ratio(PSNR), Structural Similarity Index Measurement(SSIM) [43] and Lightness Order Error(LOE), as well as three no-referenced image quality assessment indexes including Naturalness Image Quality Evaluator(NIQE), Contrast Improvement Index(CII) [44] and Average Gradient(AG) are selected to evaluate the results of the experiment. It should be noted that LOE is also often used to measure the brightness order error between the enhanced image and the original image. However, using the original image to compute LOE is problematic. One should choose a reliable Ground Truth as a reference [13]. Therefore, LOE_{ref} used in this paper belongs to the full reference image quality index.

Table 2: SSIM, PSNR, LOE_{ref} , NIQE, CII, AG metrics of each algorithm on polyp images of the testing set LHI. $\downarrow$ means that the smaller the value of the corresponding objective index, the better the enhancement result. $\uparrow$ means that the larger the value of the corresponding objective index, the better the enhancement result.

Algorithm	SSIM $\uparrow$	PSNR $\uparrow$	$LOE_{ref}\downarrow$	NIQE $\downarrow$	CII $\uparrow$	AG $\uparrow$
BPDHE [6]	0.7163	15.3104	1372.4284	4.8981	0.7755	4.2645
Dong [22]	0.8870	19.5481	1545.8305	5.3263	0.9475	5.6325
AGCWD [41]	0.8722	18.8017	1371.2160	4.6730	0.7470	4.1633
NPE [8]	0.8620	17.9272	1420.0461	4.6757	0.6802	4.0914
SRIE[9]	0.9051	20.3136	1332.7079	4.3695	0.7057	4.3086
MF [10]	0.8888	19.2371	1215.5981	4.6964	0.6482	4.3377
LIME [11]	0.8998	19.2184	1308.0230	4.7895	0.6719	4.3346
BIMEF [42]	0.8831	20.1190	1259.5562	4.7423	0.6249	4.1973
RRM [23]	0.8913	18.6979	1553.4079	5.4251	0.7507	4.6677
RetinexNet [12]	0.8012	17.4408	1556.5680	5.0568	0.7838	5.3250
MBLLEN [28]	0.8805	22.3009	969.4290	4.4759	1.5285	6.1699
KinD[13]	0.9141	21.4660	990.8599	4.9926	1.0236	5.9252
DeepUPE [15]	0.9162	22.8697	1139.5569	4.6340	0.6871	4.2341
Zero-DCE[31]	0.3344	13.1093	1549.5628	4.6223	0.5989	4.7106
Zero-DCE++ [32]	0.3193	13.6593	1577.3043	4.6836	0.6066	4.6728
EnlightenGAN [14]	0.9310	23.5917	1070.3859	4.3454	1.2368	5.7481
EDAU-Net	0.9628	26.9513	701.1326	4.3147	1.3927	6.2345

In order to make objective quantitative comparison, the average value of image quality assessment indexes obtained after processing 619 test images of dataset LHI was calculated during the experiment. The comparison results are shown in Table 2.

As can be seen from Table 2, except that the CII of the proposed EDAU-Net is lower than that of MBLLEN, the average values of SSIM, PSNR, NIQE, AG and LOE_{ref} are better than other algorithms. The PSNR and SSIM values of EDAU-Net are significantly higher than those of the comparison algorithm, which shows that EDAU-Net has better effect in reducing image distortion and restoring structure. In order to accurately and stably depict the discrete distribution of data, we draw the LOE_{ref} values obtained from 619 test images processed by 17 different enhancement methods into a box diagram, as shown in Fig. 9. It can be seen from Fig. 9 that the average

value of LOE_{ref} of EDAU-Net is the lowest, and the distance between the upper quartile and the lower quartile is small, which shows that EDAU-Net can effectively adjust the image brightness and maintain the image naturalness while obtaining a relatively stable LOE_{ref} value. EDAU-Net obtains the best NIQE value, indicating that the image enhancement result of EDAU-Net the overall performance is more natural and more consistent with the subjective visual perception of human eyes. The CII and AG values of EDAU-Net and MBLEN are significantly higher than those of other algorithms, which shows that EDAU-Net effectively improves the contrast and clarity of polyp images. MBLEN has the highest CII, indicating that MBLEN effectively improves the contrast of intestinal polyp images. However, high contrast itself may also affect the shadow removal of dark areas in the image, so the SSIM value of MBLEN is low, which directly affects the image quality and adversely affects the clinical use. In addition, the large LOE_{ref} value of MBLEN indicates that the brightness order of the image is destroyed and the naturalness is maintained poorly. The proposed EDAU-Net is superior to other algorithms in other five image quality indexes except that the average value of CII is lower than MBLEN, indicating that the comprehensive performance of the enhanced image by EDAU-Net is the best.

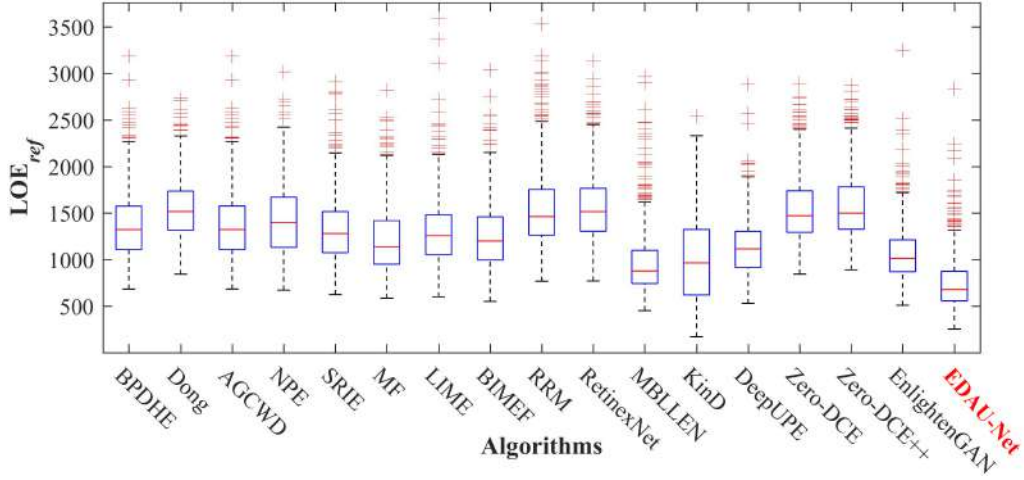


Figure 9: Box diagram of image quality assessment index LOE_{ref} .

4.2. Experiments and Results of Image Fusion

In order to highlight the details of the lesion region of interest, our proposed image fusion strategy weakens the background region of the global enhanced image. Below, we use the objective evaluation method and the subjective evaluation method of user study to analyze the results.

4.2.1. Objective Evaluation

Fig. 10 displays the salient change before and after the enhancement of polyp images by our proposed image fusion strategy. Fig. 10(e) is the ground truth for the polyps corresponding to the original input images Fig. 10(a). Fig. 10(d) shows the attention heat map obtained from the original input image without enhancement. The redder the attention heat map, the more attention the corresponding region has received from the observer. The bluer the attention heat map, the less attention the corresponding region has received from the observer. It can be seen from the observation that after the intestinal polyp image is locally enhanced by our proposed image fusion strategy, people pay more attention to the lesion region. Therefore, the image fusion framework for the ROI of internal polyp images proposed in this paper is expected to assist doctors in accurate diagnosis.

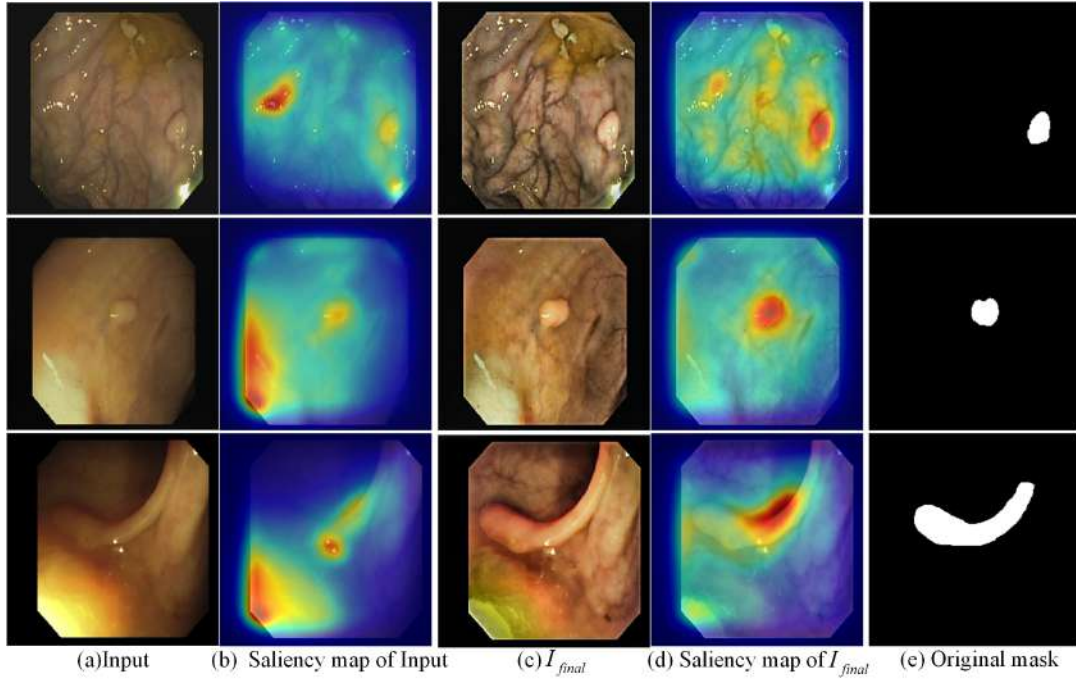


Figure 10: Saliency map for example images.

Our proposed framework can locally enhance images with polyp. In fact, the EDAU-Net in the proposed framework can be applied to images without polyp for global enhancement. We trained our model with the intestinal polyp images in dataset LHI. To better evaluate the theoretical advantages and practical application value of the proposed method, we selected 2000 images from the web public dataset Kvasir [45] and 17 images from the actual Clinical dataset collected from a Grade-A tertiary hospital in Zhengzhou city, Henan Province as the test images (images with and without polyps). The actual Clinical dataset was collected by the high-definition colonoscope CF-H260AI from OLYMPUS in Japan. CF-H260AI is the world’s most advanced digestive endoscopic diagnosis and treatment equipment recommended by the World Health Organization and the International Gastroenterological Association. Its electronic endoscope system contains image enhancement technology that can enhance the quality of clinical images. We validated the image enhancement algorithm by directly comparing the high-quality image obtained by CF-H260AI with the image obtained by the proposed method. The test images included normal cecum images, normal pylorus images, pathological finding images of polyps, and ulcerative colitis. We randomly selected 10 images without and with polyps from 2017 test images and their experimental results to display in our paper. The comparison between the original endoscopic images and the enhanced images of the proposed method in Kvasir dataset is shown in Fig. 11. The comparison of the enhancement effect of other traditional techniques BPDHE [6], Dong [22], AGCWD [41], NPE [8], SRIE [9], MF [10], LIME [11], BIMEF [42], RRM [23], Wang [46] and CF-H260AI with our method in the Clinical dataset is shown in Fig. 12. It can be seen that the proposed method can locally enhance the images with polyps to make the polyps appear clearer and more prominent, and can also globally enhance the images without polyps to make details of the intestinal lining clearer. More specifically, our method can more clearly see the position of the appendiceal orifice on the normal cecum image, the area around the opening from the stomach into the first part of the small bowel on the pylorus image, the ulceration and mucosal bleeding on the ulcerative colitis image. This shows that our proposed framework can not only enhance polyp images, but it is also suitable for image enhancement tasks of some images w/o polyp

(normal cecum images, normal pylorus images, and ulcerative colitis), with good generalizability. To further verify the effectiveness of the proposed algorithm, we used two no-referenced image quality assessment indexes, NIQE and AG to evaluate the image quality. These evaluation metrics are not applicable for the locally enhanced images with polyps. Therefore, we only show the quantitative comparison results of images w/o polyps in Fig. 11 and Fig. 12. It can be seen from the quantitative comparison of the images, all NIQE and AG values of images obtained by our proposed method are significantly better than those of the images obtained by Kvasir, CF-H260AI, and the traditional technology, which shows that our method the overall performance is more natural and effectively improves the clarity of images.

	Kvasir(with polyp)	Kvasir(w/o polyp)			
Original					
		NIQE=6.24 AG=4.34	NIQE=6.32 AG=5.80	NIQE=4.24 AG=4.61	NIQE=11.41 AG=5.78
Our					
		NIQE=5.17 AG=7.64	NIQE=5.96 AG=8.94	NIQE=4.11 AG=6.17	NIQE=8.39 AG=7.54

Figure 11: Image enhancement results for images with and without polyp in Kvasir dataset.

	CF-H260AI	BPDHE	Dong	AGCWD	NPE	SRIE	MF	LIME	BIMEF	RRM	Wang	Our
Clinical (with polyp)												
Clinical (two polyp)												

Figure 12: Image enhancement results for images with and without polyp in Clinical dataset from hospital.

For an endoscopic image with size 256×256 , we record the average running time of the proposed method to process the actual Clinical dataset from a Grade-A tertiary hospital on CPU is 0.1018 seconds, and the average running time on GPU is 0.0289 seconds. That is to say, the proposed method enhances about 8 images per second on the CPU and about 35 images per second on the GPU. The speed of imaging with GPU mode meets the requirements of clinical 2D real-time imaging. We also

invited three senior endoscopists from the hospital in Zhengzhou City, Henan Province to make a subjective evaluation on the real-time performance of the proposed algorithm and the enhancement results of Fig. 12. The results show that the proposed method not only meets the clinical needs of physicians in terms of processing speed but also outperforms the enhancement effect of CF-H260AI and other traditional techniques. In addition, we trained ACSNet on a dataset containing both polyp and non-polyp images to predict masks for 300 clinical images (143 without polyps and 157 with polyps). However, these clinical images do not have corresponding segmentation masks (Ground Truth). To address this problem, we invited three experienced endoscopists to use Labelme, an open-source image annotation tool, to label and generate masks for these clinical images. All clinical image masks were annotated on the same monitor indoors. The consensus results from two of the doctors were used as Ground Truth, and we obtained a false positive rate of 0.19% for the predicted images. Therefore, ACSNet can effectively distinguish between regions with and without polyps in the images. That is to say, the enhancement method can accurately enhance the area covered by polyps, and also has a good suppressive effect on the non-lesioned areas. This further validates the usefulness of the proposed method for clinical diagnosis and treatment.

4.2.2. Human Subjective Evaluation

In order to subjectively evaluate the image fusion framework for the ROI of intestinal polyp images, we conduct a user study to compare the performance of our method and 17 image enhancement methods including BPDHE, the algorithm proposed by Dong et al., AGCWD, NPE, SRIE, MF, LIME, BIMEF, RRM, RetinexNet, MBLLEN, KinD, DeepUPE, Zero-DCE, Zero-DCE++, EnlightenGAN, and EDAU-Net.

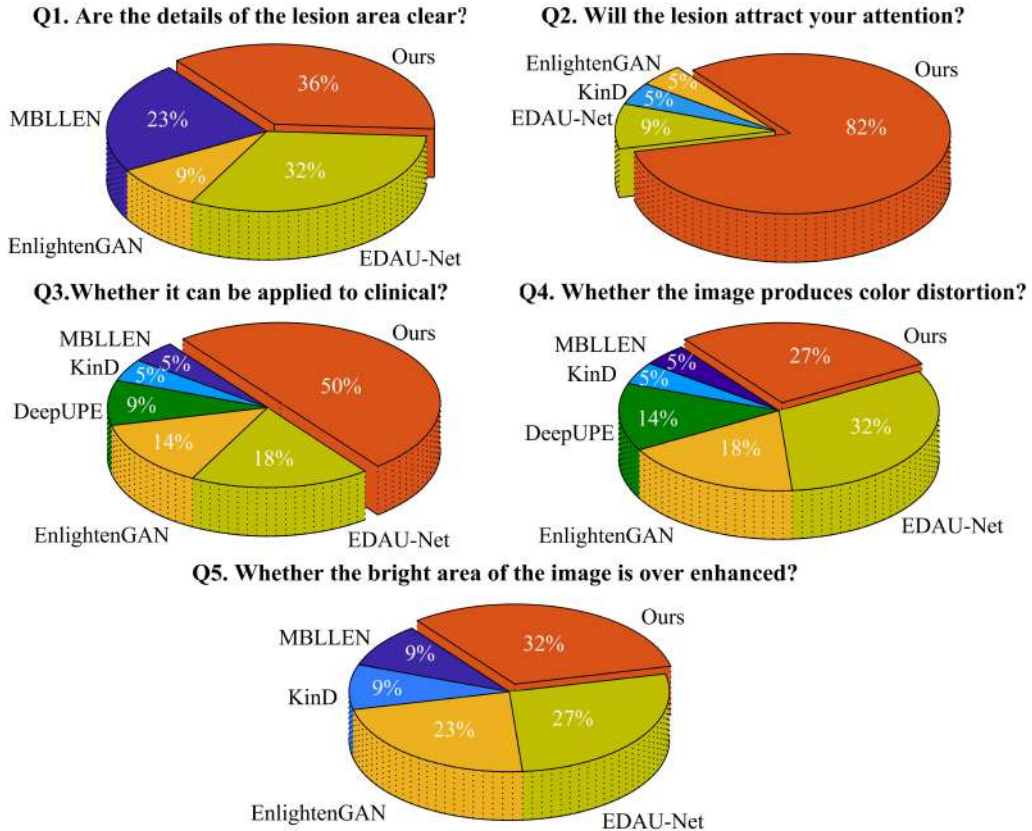


Figure 13: Results of our user study.

We consulted 8 colorectal surgeons with more than 3 years of clinical experience and 12 medical students, and obtained their evaluation feedback on the enhancement results of 22 randomly selected images from the synthetic dataset LHI. Five questions were designed in the investigation sheets (1. Are the details of the lesion area clear? 2. Will the lesion attract your attention? 3. Whether it can be applied to clinical? 4. Whether the image produces color distortion? 5. Whether the bright area of the image is over enhanced?). We repeat the above for all 22 images.

The results of human user study are presented in Fig. 13, which displays the five pie charts. The distribution of satisfactory methods for different problems shows that our method obtains more distribution proportion than other methods. Methods based on deep learning such as RetinexNet, Zero-DCE and Zero-DCE++ have zero satisfaction distribution in all five questions. This is because they cause color distortion and over-enhancement, that seriously affects people’s observation of the details of intestinal polyp images. As can be seen from Q1, Q4 and Q5, the distribution proportion of EDAU-Net enhancement results and ROI local enhancement results proposed in this paper are higher than other comparison methods. Obviously, the effect of EDAU-Net and image fusion framework for the ROI of intestinal polyp images proposed by us is superior to other methods in terms of enhancing details of lesion area, avoiding color distortion and over-enhancing brighter areas. By observing Q2, it can be seen that the distribution proportion of the proposed local enhancement framework for the ROI of intestinal polyp images is far exceeds that of EDAU-Net, which shows that the proposed image fusion framework for the ROI of intestinal polyp images can effectively attract people’s attention to polyp ROI. This conclusion is also consistent with the analysis of our Objective evaluation. Furthermore, according to Q3’s feedback, compared with the global enhancement method EDAU-Net, the image fusion framework for the ROI proposed by us is more favored by doctors and has medical clinic value. Therefore, it is necessary to perform image fusion after EDAU-Net to obtain the local enhancement result of intestinal polyp images.

5. Discussion

In the present work, we have proposed a new polyp image quality enhancement framework based on the region of interest with paired supervision to simultaneously realize two different tasks: global detail enhancement and local detail enhancement. To preserve more image details and adjust the global image illumination, we proposed a supervised global image enhancement network EDAU-Net. We also proposed a new fusion strategy to achieve the local enhancement effect of the region covered by the polyp in the image.

From the results of the comparison with traditional image enhancement techniques and recent image enhancement methods based on deep learning, as shown in Table 2, we can conclude that deep learning-based methods are more important than traditional image enhancement method for endoscopy image enhancement. Although MBLEN, KinD and EnlightenGAN have achieved remarkable image enhancement results. All full reference image quality indexes values of EDAU-Net are significantly better than those of the comparison algorithm. In addition, compared to previous studies, e.g. Ref. [6, 14, 46], the proposed method can locally enhance and weaken specific information in the images with polyps. As shown in Fig. 10, selectively enhancing the polyps on the surface of the intestinal lumen. From both a qualitative and a quantitative point of view, the results obtained by our proposed framework are satisfactory in the two different tasks. Human subjective evaluation from colorectal surgeons and medical students further solidifies the advantage of the proposed framework. Although extensive experiments have been performed to validate the effectiveness of our proposed framework, some issues still exist that need to be further discussed.

5.1. The reason for using the enhanced image as input to ACSNet

The ROI acquisition module was implemented by feeding globally enhanced images into ACSNet. To further investigate the mechanisms and rationales behind our proposed framework, we present a brief analysis to prove that using the enhanced image as input to ACSNet could obtain the polyp

region more accurately. Although ACSNet has shown excellent performance in polyp segmentation tasks, the visual complexity associated with images such as low contrast and uneven illumination can hamper the optimal performance of deep convolutional neural networks based methods [47]. In order to obtain ROI images more accurately, we conducted two experiments, one was to train ACSNet with the Synthetic input image in dataset LHI, and the other was to train ACSNet with the enhanced Synthetic input image I_e by EDAU-Net. Fig. 14 shows the comparison results between the different ROI and Ground Truth by using two different types of image input ACSNet, the unenhanced image and the enhanced image. It reveals that the ACSNet model trained with enhanced image dataset by EDAU-Net could detect the polyp region more accurately than trained with unenhanced image dataset. The explanation of this accuracy enhancement is that the proposed EDAU-Net reduces artifacts near the polyp, which causes the polyp segmentation borders obtained by ACSNet to come closer to that of the Ground Truth. Therefore, it is necessary to perform global image enhancement for intestinal polyp images before ACSNet training to achieve ROI acquisition.

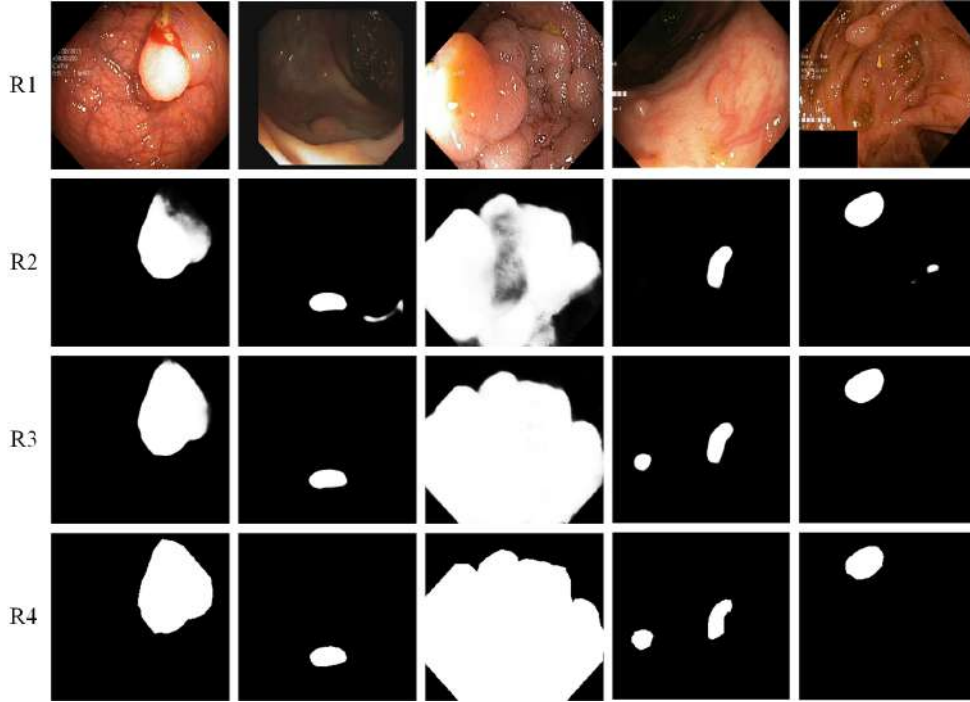


Figure 14: Visual comparison of ROI output from ACSNet. Row 1~4 display the without enhanced Syn input images, ROI images obtained by inputting without enhanced Syn input images into ACSNet, ROI images obtained by inputting the enhanced image output from EDAU-Net into ACSNet, and Original masks (Ground Truth), respectively.

5.2. Discussion of the contribution of each component in the EDAU-Net

In order to prove the effectiveness of each component in the EDAU-Net model, we performed a series of ablation studies to analyze EDAU-Net and compared it with the following network structures: 1) Without Detail Attention Maps, only the U-Net based model composed of Luminance Attention Maps is introduced; 2) Without Luminance Attention Maps, only the U-Net based model composed of Detail Attention Maps is added; 3) The EDAU-Net model of Detail Attention Maps and Luminance Attention Maps is added on the basis of U-Net. For the above models, the same parameters as EDAU-Net are used for training on the synthetic dataset LHI for fair comparison. The qualitative and quantitative results are shown in Fig. 15 and Table 3.

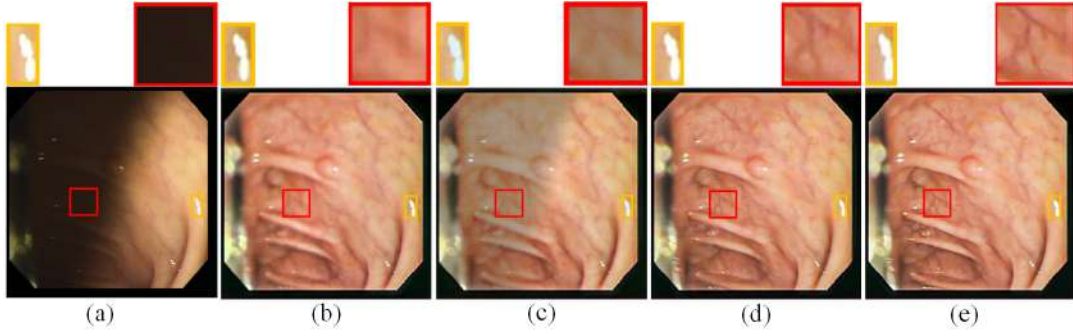


Figure 15: Qualitative comparison of models with different structures on synthetic polyp images. (a) Input; (b) Without Detail Attention Maps; (c) Without Luminance Attention Maps; (d) EDAU-Net; (e) Ground Truth.

Table 3: Quantitative comparison of models with different structures on polyp images of dataset LHI.

Model	SSIM $\uparrow$	PSNR $\uparrow$	$LOE_{ref}\downarrow$	NIQE $\downarrow$	CII $\uparrow$	AG $\uparrow$
(b) Without Detail Attention Maps	0.9570	26.2328	760.4334	4.3297	1.2181	6.1689
(c) Without Luminance Attention Maps	0.9609	26.7480	740.7765	4.3380	1.3671	6.2128
(d) EDAU-Net	0.9628	26.9513	701.1326	4.3147	1.3927	6.2345

By observing Fig. 15, it can be seen that Fig. 15(d) has clearer texture and detail information than Fig. 15(b), which shows that Detail Attention Maps can improve the clarity of the image. Comparing Fig. 15(c) with Fig. 15(d), it can be seen that the overall brightness of Fig. 15(c) is more uneven, which shows that Luminance Attention Maps has a better adjustment effect on the non-uniform illumination of the original image Fig. 15(a). In addition, the edge of the highlighted area in Fig. 15(c) produces a lot of noise, which shows that Luminance Attention Maps can also suppress noise generation. It can be seen from Table 3 that the values of CII and AG have been optimized to varying degrees after adding Detail Attention Maps. All evaluation criteria have been optimized after adding Luminance Attention Maps, especially PSNR and LOE_{ref} . Therefore, the quantitative results in Table 3 are also consistent with the qualitative results in Fig. 15. The attention map is helpful to extract the image features of intestinal polyps.

5.3. Discussion of the contribution of each loss function

In addition, we analyze the contributions of the L_1 loss, L_2 loss, $L_{Structure}$ loss, and $L_{Perceptual}$ loss. More specifically, the w/o L_1 means that only the L_2 , $L_{Structure}$ and $L_{Perceptual}$ are taken to train the EDAU-Net, the w/o L_2 means that only the L_1 , $L_{Structure}$ and $L_{Perceptual}$ are adopted to constrain the EDAU-Net, the w/o $L_{Structure}$ presents that only the L_1 , L_2 and $L_{Perceptual}$ are taken to train the EDAU-Net, and the w/o $L_{Perceptual}$ means that the term of $L_{Perceptual}$ is discarded in the total loss. The results of different combinations of loss functions are shown in Table 4.

Table 4: Quantitative comparison of models with different loss functions on polyp images of dataset LHI. “w/o” means without. Bold means the best results.

Model	SSIM $\uparrow$	PSNR $\uparrow$	$LOE_{ref}\downarrow$	NIQE $\downarrow$	CII $\uparrow$	AG $\uparrow$
w/o L_1	0.9621	26.9192	707.6696	4.3190	1.3758	6.2218
w/o L_2	0.9570	26.1741	761.5694	4.3636	1.1865	6.1470
w/o $L_{Structure}$	0.9627	26.9418	705.0991	4.3657	1.3730	6.2323
w/o $L_{Perceptual}$	0.9619	26.8982	708.2144	4.3531	1.3881	6.2274
EDAU-Net	0.9628	26.9513	701.1326	4.3147	1.3927	6.2345

As illustrated in Table 4, L_1 urges EDAU-Net to learn the global similarity between the predicted image and the target image and suppresses the network to generate noise. L_2 can significantly improve the overall performance of images. $L_{Structure}$ encourages the network to learn the global similarity between the predicted image and the target image and makes the results look natural. $L_{Perceptual}$ improved all quantitative evaluation indexes. It is worth noting that the value of PSNR increased by 0.0531 and the values of LOE_{ref} and NIQE decreased by 7.0818 and 0.0366 respectively, with $L_{Perceptual}$ compared to without $L_{Perceptual}$. This proves that $L_{Perceptual}$ has certain advantages in noise removal, brightness enhancement, and natural visual effects maintenance of polyp images, and $L_{Perceptual}$ encourages the generated images to have similar content as the target images. The results show that the quality of enhancement is improving by combining the above loss functions reasonably.

5.4. Discussion of the loss weights setting

In this part, we discuss the effect of λ_1 , λ_2 , λ_3 and λ_4 on the performance. Here we performed some ablation experiments to explore the performance of each parameter by setting different combinations of loss weights and running over 150 iterations on EDAU-Net. It is understandable that such an approach is time-consuming and not efficient. However, the nature of the problem does not permit the derivation of a direct relation of those parameters with the enhancement technique. We find that even though these hyperparameters need fine-tuning carefully, their settings do follow certain principles. The weighted coefficient λ_2 of L_2 loss in fact controls the high-frequency information between an enhanced image and its Ground Truth image. When λ_2 is large, the comprehensive performance of the image can be significantly improved. Thus λ_2 should be large enough to ensure the correspondence of the enhanced high-frequency information. For L_1 loss, too large a value of λ_1 often leads to low contrast and sharpness in the whole image. An excessively large λ_3 value also leads to low contrast. In addition, too large λ_4 usually causes the enhanced image to deviate from the brightness order of the original image. After extensive experiments, we concluded that when the values of the parameters λ_1 , λ_2 , λ_3 , and λ_4 are equal to 1, 1000, 1, and 1 respectively, the achieved result is more similar to the reference image and the objectively evaluation metrics can get better value, which means that the proposed method can achieve wonderful performance. Therefore, the parameters λ_1 , λ_2 , λ_3 , and λ_4 are finally set as to 1, 1000, 1, and 1 through testing results.

5.5. Limitations

Although the proposed framework shows the generalizability in challenging polyp image datasets, some limitations of our method still exist. In a few exceptional cases, our method cannot accurately distinguish polyps from background of the polyp images and perform targeted enhancement of lesion regions. For example, inappropriate bowel preparation (with a large amount of fluid and feces) may lead to more error prone polyp segmentation. In fact, even experienced colonoscopists may disagree on the polyp boundary segmentation of such challenging images. Fortunately, ACSNet in our proposed framework is replaceable. With the development of deep learning, the segmentation accuracy of the polyp segmentation algorithm is constantly optimized. It is an important research direction in the future for us to accurately segment polyp images obtained under improper bowel preparation.

When ACSNet was trained on a dataset containing both polyp and non-polyp images, the false positive rate tested on clinical images with and without polyps was 0.19%. This is because ACSNet may produce false positive results for non-polyp images that are close to the polyp characteristic region. According to the doctors' feedback, the proposed method can significantly reduce the diagnostic workload and effectively meet clinical needs. The false positives in non-polyp images will not lead to unnecessary and potentially dangerous surgery, as computer images are only a tool for assisting medical decisions. In practical applications, doctors can switch between local and global image enhancement effects of the proposed algorithm at their discretion, and perform further confirmations as needed.

6. Conclusion

In this paper, an intestinal polyp image dataset LHI composed of paired low-quality images and high-quality images was constructed for the first time. LHI solves the problem of the lack of paired intestinal polyp image enhancement datasets, contributes to the development of deep learning in the field of intestinal polyp image enhancement, and is of great significance for improving the quality of polyp images. We proposed the EDAU-Net model by introducing Detail Attention Maps and the Luminance Attention Maps. EDAU-Net can effectively eliminate non-uniform luminance and improve image texture and detail information. EDAU-Net is compared with the classic algorithm and the latest algorithm on the dataset LHI. The results show that the proposed EDAU-Net has the best comprehensive performance. By weakening the background region to highlight the details of the region of interest, we proposed a new image fusion strategy that can help doctors quickly and clearly observe the texture and details of the lesion region. Compared with the traditional image enhancement methods and the state-of-the-art deep learning methods, the proposed has better local enhancement effect and has certain clinical practicability. In the future work, we will improve the algorithm for the further research on 3D reconstruction of the lesion region.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Shanghai Natural Science Foundation of China (Grant No. 19ZR1419100) and the Shanghai talent development funding of China (Grant No. 2021016).

References

- [1] M. Taş, B. Yılmaz, Super resolution convolutional neural network based pre-processing for automatic polyp detection in colonoscopy images, *Comput. Electr. Eng.* 90 (2021) 106959, <http://doi.org/10.1016/j.compeleceng.2020.106959>.
- [2] Z. Yang, J. Pan, R. Li, H. Qin, Scene-graph-driven semantic feature matching for monocular digestive endoscopy, *Comput. Biol. Med.* 146 (2022) 105616, <http://doi.org/10.1016/j.compbiomed.2022.105616>.
- [3] S. J. Mahato, D. Banik, D. Bhattacharjee, Exploring hand-crafted features and transfer learning for polyp segmentation, in: *Computational Intelligence in Communications and Business Analytics*, 2021, pp. 68–76, https://doi.org/10.1007/978-3-030-75529-4_6.
- [4] K. Patel, A. M. Bur, G. Wang, Enhanced u-net: A feature enhancement network for polyp segmentation, in: *2021 18th Conference on Robots and Vision*, 2021, pp. 181–188, <http://doi.org/10.1109/CRV52889.2021.00032>.
- [5] S. Pizer, R. Johnston, J. Erickson, B. Yankaskas, K. Muller, Contrast-limited adaptive histogram equalization: speed and effectiveness, in: *Proceedings of the First Conference on Visualization in Biomedical Computing*, 1990, pp. 337–345, <https://doi.org/10.1109/VBC.1990.109340>.
- [6] H. Ibrahim, N. S. P. Kong, Brightness preserving dynamic histogram equalization for image contrast enhancement, *IEEE Trans. Consumer Electron.* 53 (2007) 1752–1758, <http://doi.org/10.1109/TCE.2007.4429280>.
- [7] Z.-u. Rahman, D. J. Jobson, G. A. Woodell, Multi-scale retinex for color image enhancement, in: *Proceedings of 3rd IEEE International Conference on Image Processing*, 1996, pp. 1003–1006, <https://doi.org/10.1109/ICIP.1996.560995>.

- [8] S. Wang, J. Zheng, H.-M. Hu, B. Li, Naturalness preserved enhancement algorithm for non-uniform illumination images, *IEEE Trans. Image Process.* 22 (2013) 3538–3548, <https://doi.org/10.1109/TIP.2013.2261309>.
- [9] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, X. Ding, A weighted variational model for simultaneous reflectance and illumination estimation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2782–2790, <https://doi.org/10.1109/CVPR.2016.304>.
- [10] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, J. Paisley, A fusion-based enhancing method for weakly illuminated images, *Signal Process.* 129 (2016) 82–96, <https://doi.org/10.1016/j.sigpro.2016.05.031>.
- [11] X. Guo, Y. Li, H. Ling, Lime: Low-light image enhancement via illumination map estimation, *IEEE Trans. Image Process.* 26 (2016) 982–993, <https://doi.org/10.1109/TIP.2016.2639450>.
- [12] C. Wei, W. Wang, W. Yang, J. Liu, Deep retinex decomposition for low-light enhancement, 2018 arXiv preprint arXiv:1808.04560, <https://doi.org/10.48550/arXiv.1808.04560>.
- [13] Y. Zhang, J. Zhang, X. Guo, Kindling the darkness: A practical low-light image enhancer, in: *Proceedings of the 27th ACM international conference on multimedia*, 2019, pp. 1632–1640, <https://doi.org/10.1145/3343031.3350926>.
- [14] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, Z. Wang, Enlightengan: Deep light enhancement without paired supervision, *IEEE Trans. Image Process.* 30 (2021) 2340–2349, <https://doi.org/10.1109/TIP.2021.3051462>.
- [15] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, J. Jia, Underexposed photo enhancement using deep illumination estimation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6849–6857, <https://doi.org/10.1109/CVPR.2019.00701>.
- [16] Y. Li, J. Fan, D. Ai, H. Song, Y. Wang, J. Yang, A general endoscopic image enhancement method based on pre-trained generative adversarial networks, in: *2020 IEEE International Conference on Bioinformatics and Biomedicine*, 2020, pp. 2403–2408, <https://doi.org/10.1109/BIBM49941.2020.9313443>.
- [17] Y. Ma, Y. Liu, J. Cheng, Y. Zheng, M. Ghahremani, H. Chen, J. Liu, Y. Zhao, Cycle structure and illumination constrained gan for medical image enhancement, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020, pp. 667–677, https://doi.org/10.1007/978-3-030-59713-9_64.
- [18] R. Zhang, G. Li, Z. Li, S. Cui, D. Qian, Y. Yu, Adaptive context selection for polyp segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020, pp. 253–262, https://doi.org/10.1007/978-3-030-59725-2_25.
- [19] H. Okuhata, H. Nakamura, S. Hara, H. Tsutsui, T. Onoye, 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*, 2013, pp. 3407–3410, <https://doi.org/10.1109/EMBC.2013.6610273>.
- [20] D. J. Jobson, Z.-u. Rahman, G. A. Woodell, Properties and performance of a center/surround retinex, *IEEE Trans. Image Process.* 6 (1997) 451–462, <https://doi.org/10.1109/83.557356>.
- [21] X. Luo, H.-Q. Zeng, Y. Wan, X.-B. Zhang, Y.-P. Du, T. M. Peters, Endoscopic vision augmentation using multiscale bilateral-weighted retinex for robotic surgery, *IEEE Trans. Med Imaging.* 38 (2019) 2863–2874, <https://doi.org/10.1109/TMI.2019.2916101>.

- [22] X. Dong, G. Wang, Y. Pang, W. Li, J. Wen, W. Meng, Y. Lu, Fast efficient algorithm for enhancement of low lighting video, in: 2011 IEEE International Conference on Multimedia and Expo, 2011, pp. 1–6, <https://doi.org/10.1109/ICME.2011.6012107>.
- [23] M. Li, J. Liu, W. Yang, X. Sun, Z. Guo, Structure-revealing low-light image enhancement via robust retinex model, *IEEE Trans. Image Process.* 27 (2018) 2828–2841, <https://doi.org/10.1109/TIP.2018.2810539>.
- [24] R. Kumar, A. K. Bhandari, Spatial mutual information based detail preserving magnetic resonance image enhancement, *Comput. Biol. Med.* 146 (2022) 105644, <https://doi.org/10.1016/j.compbiomed.2022.105644>.
- [25] H. Zhang, W. Qian, M. Wan, K. Zhang, Infrared image enhancement algorithm using local entropy mapping histogram adaptive segmentation, *Infrared Phys. Technol.* 120 (2022) 104000, <https://doi.org/10.1016/j.infrared.2021.104000>.
- [26] Y.-T. Kim, Contrast enhancement using brightness preserving bi-histogram equalization, *IEEE Trans. Consumer Electron.* 43 (1997) 1–8, <https://doi.org/10.1109/30.580378>.
- [27] A. Attar, X. Xie, C. Zhang, Z. Wang, S. Yue, Wireless micro-ball endoscopic image enhancement using histogram information, in: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014, pp. 3337–3340, <https://doi.org/10.1109/EMBC.2014.6944337>.
- [28] F. Lv, F. Lu, J. Wu, C. Lim, Mblen: Low-light image/video enhancement using cnns, in: Proceedings of the 2018 British Machine Vision Conference, 2018, pp. 1–13, <https://www.researchgate.net/publication/360481784>.
- [29] K. G. Lore, A. Akintayo, S. Sarkar, Llnet: A deep autoencoder approach to natural low-light image enhancement, *Pattern Recogn.* 61 (2017) 650–662, <https://doi.org/10.1016/j.patcog.2016.06.008>.
- [30] C. Chen, Q. Chen, J. Xu, V. Koltun, Learning to see in the dark, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3291–3300, <https://doi.org/10.1109/CVPR.2018.00347>.
- [31] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, R. Cong, Zero-reference deep curve estimation for low-light image enhancement, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1780–1789, <https://doi.org/10.1109/CVPR42600.2020.00185>.
- [32] C. Li, C. Guo, C. C. Loy, Learning to enhance low-light image via zero-reference deep curve estimation, 2021 arXiv preprint arXiv:2103.00860, <https://doi.org/10.48550/arXiv.2103.00860>.
- [33] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, H. D. Johansen, Kvasir-seg: A segmented polyp dataset, in: International Conference on Multimedia Modeling, 2020, pp. 451–462, https://doi.org/10.1007/978-3-030-37734-2_37.
- [34] J. Silva, A. Histace, O. Romain, X. Dray, B. Granado, Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer, *Int. J. Comput. Assist. Radiol. Surg.* 9 (2014) 283–293, <https://doi.org/10.1007/s11548-013-0926-3>.
- [35] D. Vázquez, J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, A. M. López, A. Romero, M. Drozdal, A. Courville, A benchmark for endoluminal scene segmentation of colonoscopy images, *J. Healthc Eng.* 2017 (2017) 1–9, <https://doi.org/10.1155/2017/4037190>.

- [36] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.
- [37] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach Intell. 8 (1986) 679–698, <https://doi.org/10.1109/TPAMI.1986.4767851>.
- [38] Z. Wang, E. P. Simoncelli, A. C. Bovik, Multiscale structural similarity for image quality assessment, in: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, 2003, pp. 1398–1402, <https://doi.org/10.1109/ACSSC.2003.1292216>.
- [39] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, 2009, pp. 248–255, <https://doi.org/10.1109/CVPR.2009.5206848>.
- [40] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughaier, M. S. Khan, et al., Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images, Comput. Biol. Med. 132 (2021) 104319, <https://doi.org/10.1016/j.compbiomed.2021.104319>.
- [41] S.-C. Huang, F.-C. Cheng, Y.-S. Chiu, Efficient contrast enhancement using adaptive gamma correction with weighting distribution, IEEE Trans. Image Process. 22 (2012) 1032–1041, <https://doi.org/10.1109/TIP.2012.2226047>.
- [42] Z. Ying, G. Li, W. Gao, A bio-inspired multi-exposure fusion framework for low-light image enhancement, arXiv preprint arXiv:1711.00591, <https://doi.org/10.48550/arXiv.1711.00591>.
- [43] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612, <https://doi.org/10.1109/TIP.2003.819861>.
- [44] S. Saravanan, P. S. Kumar, Image contrast enhancement using histogram equalization techniques, Int. J. Adv. Comput. Sci. Technol. 3 (2014) 163–172.
- [45] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, P. T. Schmidt, et al., Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection, in: Proceedings of the 8th ACM on Multimedia Systems Conference, 2017, pp. 164–169, <https://doi.org/10.1145/3083187.3083212>.
- [46] L. Wang, B. Wu, X. Wang, Q. Zhu, K. Xu, Endoscopic image luminance enhancement based on the inverse square law for illuminance and retinex, Int. J. Med. Robotics. Comput. Assist. Surg. 18 (2022) e2396, <https://doi.org/10.1002/rcs.2396>.
- [47] O. Sule, S. Viriri, M. Gwetu, Contrast enhancement in deep convolutional neural networks for segmentation of retinal blood vessels, in: Asian Conference on Intelligent Information and Database Systems, 2021, pp. 278–290, https://doi.org/10.1007/978-981-16-1685-3_23.