Noname manuscript No.
(will be inserted by the editor)

Deep Unsupervised Endoscopic Image Enhancement Based on Multi-image Fusion

Dongjin Huang[1] · Jinhua Liu[1] · Shuhua Zhou[1] · Wen Tang[2]

Received: 4th July 2021 / Accepted: date

Abstract *Background and objective*: A deep unsupervised endoscopic image enhancement method is proposed based on multi-image fusion to achieve high quality endoscope images from poorly illuminated, low contrast and color deviated images through an unsupervised mapping and deep learning network without the need for ground truth.

Methods: Firstly, three image enhancement methods are used to process original endoscopic images to obtain three derived images, which are then transformed into HSI color space. Secondly, a deep unsupervised multi-image fusion network (DerivedFuse) is proposed to extract and fuse features of the derived images accurately by utilizing a new no-reference quality metric as loss function. I-channel images of the three derived images are inputted into the DerivedFuse network to enhance the intensity component of the original image. Finally, a saturation adjustment function is proposed to adaptive adjusting the saturation component of HSI color space to enrich the color information of the original input image.

Results: Three evaluation metrics: Entropy, Contrast Improvement Index (CII) and Average Gradient (AG) are used to evaluate the performance of the proposed method. The results are compared with that of fourteen state-of-the-art algorithms. Experiments on endoscopic image enhancement show that the Entropy value of our method is 3.27% higher than the optimal entropy value of comparison algorithms. The CII of our proposed method is 6.19% higher than that of comparison algorithms. The AG of our method is 7.83% higher than the optimal AG of comparison algorithms.

Conclusions: The proposed deep unsupervised multi-image fusion method can obtain image information details, enhance endoscopic images with high contrast, rich and natural color information, visual and image quality. Sixteen doctors and medical students have given their assessments on the proposed method for assisting clinical diagnoses.

Keywords Endoscopic image enhancement · Unsupervised deep learning · Image Fusion · Derived image · HSI color space

1 Introduction

Endoscopy imaging is a medical diagnostic and treatment procedure [1]. Endoscopy has been widely used in examining, diagnosing and treating esophageal digestive system, such as stomach and intestine. The image collected by endoscopy is usually collected by doctors operating endoscopy in narrow internal cavities. Endoscopic images are illuminated only by unidirectional point light source which causes problems of uneven illumination and low illumination. Tissues and blood vessels of human organs are mainly distributed in the mucosa and submucosa, which leads to which causes problems of the loss of image texture and color information. These problems will lead to uneven brightness, insufficient brightness, low contrast and clarity in the organ regions of interest to doctors in the image [2], thus affecting the accuracy of the doctors to analysis and diagnosis in clinical settings. Therefore, it is very important to improving endoscopic image quality in analysis and diagnosis in clinical settings.

In this paper, we study techniques aimed at improving endoscopic image quality. In the past, many image processing techniques have been proposed for brightness adjustment and contrast enhancement, such as gamma

✉ Dongjin Huang
djhuang@shu.edu.cn

1 Shanghai Film Academy, Shanghai University, Shanghai, China

2 The Faculty of Science, Design and Technology, University of Bournemouth, Poole, Dorset, UK

correction (GC) [3] based on gray level transformation and contrast-limited adaptive histogram equalization algorithm (CLAHE) [4] based on histogram and low-light image enhancement via illumination map estimation (LIME) [5] based on Retinex theory. Xia et al. proposed an endoscopic image enhancement algorithm based on Retinex theory to improve the insufficient illumination and prevent magnifying noise in the process of image enhancement [6]. Long et al. [7] combined histogram correction with wavelet transform to improve the contrast of endoscopic images and avoid image distortion. However, the traditional image enhancement algorithm only focuses on a specific problem, which is to correct the brightness in the image or to highlight the contrast of the blood vessels in mucosal layer of the human lumen. With the rapid development of deep learning networks, deep learning-based methods have made important progress in image classification, segmentation and detection [8,9]. LLNet [10], a CNN-based model, was successfully applied to low illumination image enhancement to natural images. Subsequently, LightenNet [11] and RetinexNet [12] networks were proposed and achieved significant brightness enhancement for low illumination images. Compared with traditional image processing technique, deep-learning methods have the advantages of getting more images features from large samples of datasets and more powerful feature expression ability. However, supervised neural networks require a large number of image pairs for training. The lack of medical ground truth datasets for network training is a major bottleneck in deep-learning based medical imaging processing. Unsupervised learning methods do not need ground truth medical images for training and is an effective way of processing endoscope image enhancement. More recently, Chen et al. [13] proposed a new unsupervised learning framework De-smokeGCN without the need for real image pairs as the ground truth, which has been used for smoke detection and smoke removal of endoscopic images. The enhancement effect is not ideal when using unsupervised deep learning-based image enhancement methods to enhance endoscopic images. Therefore, it is difficult to study endoscopic images before going straight to the previous deep learning-based image enhancement methods.

In this paper, we propose a deep unsupervised learning framework based on a multi-image fusion method for endoscopic image enhancement. It only uses low-quality endoscopic images to train the network without the need for ground truth, combining several mature image enhancement techniques to achieve contrast enhancement, uniform brightness, high clarity and natural color of endoscopic images. Firstly, we use three classical image enhancement techniques: GC, CLAHE and LIME to process the original endoscopic image to get three derived images. The RGB color space of the original images and the derived images have been transformed into HSI color space. Each channel in the HSI color space has relatively independent characteristics, and we deal with the luminance channel component I and saturation channel component S respectively to avoid the color distortion of the image without changing the hue channel component H. For I-channel components, we propose DerivedFuse to improve the details of the original image. Considering the characteristics of images obtained from endoscopy, we improve multiple exposure fusion structural similarity (MEF SSIM) [14] and propose a new loss function MDF SSIM to extract and fuse features of derived images more accurately. For S-channel component, we construct a saturation adjustment function to enrich the color information of the image.

The contributions of this work are as follows:

1) For the first time, we combine advantages of GC derived image, CLAHE derived image and LIME derived image with a deep learning network. The deep convolutional neural network architecture for unsupervised fusion of derived images is proposed to obtain enhanced endoscopic images with superior image quality.
2) DerivedFuse is proposed to fuse fine details of derived images. This model can accurately extract and fuse the features of the derived image without ground truth. A new loss function MDF SSIM was proposed to extract and fuse features of derived images more accurately.
3) A saturation adjustment function was constructed to enrich the color information of the image in HSI color space.
4) Our proposed method outperforms fourteen state-of-the-art algorithms and has shown high quality image enhancement. Sixteen medical doctors and students have evaluated our proposed algorithm for clinical endoscopic diagnosis and treatment. The proposed method is general and applicable to images obtained from gastroscopy, colonoscopy and laparoscopy.

2 Related work

Medical images obtained from endoscopy is illuminated by unidirectional point light source. This light source could cause uneven brightness or darkness in some areas of the image [15]. Therefore, research of endoscopic image enhancement techniques has been mainly focused on illumination adjustment, contrast and sharpness improvement. In recent years, researchers have proposed some classical image enhancement methods to improve endoscopic images quality [16–19]. With the rapid development of deep learning, deep convolutional neural network has gradually become the main driving force in the field of image enhancement. Deep learning algorithm builds the complex nonlinear mapping relationship between the degraded image and the real image by learning the deep convolution neural network, and then achieve the purpose of enhancing the degraded image.

In following sections, we divide these methods into two categories: classic conventional image enhancement methods and deep learning-based image enhancement methods.

2.1 Conventional image enhancement methods

Image enhancement is an active research topic in the field of computer vision. Many image enhancement algorithms have been proposed to address issues of natural images. Classical methods include algorithms based on histogram [4,20,21], gray level transformation [3,22,23], Retinex theory based algorithms [5,24–28]. Histogram based algorithm such as CLAHE [4] can significantly improve the contrast of the image to highlight the sense of layers and make details in bright areas of the image clear, but not effective for dark areas. Whilst gray level transformation method, such as GC [3], is commonly used gray non-linear transformation to improve the brightness of the image and enhance details in the dark areas, it is usually limited to a narrow gray scale and the contrast is low. Finally, LIME [5] based on Retinex theory can significantly improve the contrast and clarity of images in dark areas. However, it can cause excessive enhancement to areas with high-brightness in the image.

In endoscopic images enhancement, a method for processing endoscopy images using texture analysis is proposed by Hiroyasu et al. [16] to remove the noise when analyzing images of early gastric cancer taken with narrow-band imaging endoscopy using a gray level co-occurrence matrix and a gray level run-length matrix. A color image enhancement scheme termed as "Tri-scan" is proposed by Imtiaz et al. [17]. Firstly, the algorithm sharpens the surface and edges of tissue and vascular characteristics. Next, the R plane of the image is processed with an adaptive sigmoid function to enhance the vascular contrast of the mucosal layer. Finally, the subtle micro-vessels, mucosal and tissue characteristics were highlighted by color tone enhancement. Sdiri et al. [18] addressed stereo endoscopic image enhancement based on joint wavelet decomposition and binocular combination. These methods only focused on a specific problem such as light regulation, vascular contrast enhancement or smoke elimination.

2.2 Deep learning-based image enhancement methods

Supervised learning and unsupervised learning are the two main methods of Deep learning-based methods. In supervised learning algorithms, they need a training data set including ground truth images, and build the complex nonlinear mapping relationship between the degraded image and the real image by learning the deep convolution neural network for the purpose of enhancing the degraded image. Unsupervised leaning methods could realize the transformation from low-quality raw images to high-quality enhanced images, not requiring image pairs.

Supervised learning: Lore et al. [10] applied a low-light net (LLNet) for low light image enhancement. LLNet was trained on a variant of the stacked-sparse denoising auto-encoder to enhance the contrast and denoise low illumination images. However, when LLNet is used to process color images of real scenes, there are more redundant parameters, so it is easy to overfitting. Li et al. [11] proposed a convolutional neural network (LightenNet) for weakly illuminated image enhancement. LightenNet uses four layer convolutional neural network (CNN) model to realize the illumination mapping between the low illumination images and the corresponding normal light images. However, the stability of the algorithm is poor, resulting in overexposure and redundant noise. Lv et al. [29] proposed a multi-branch weak light image enhancement network (MBLLEN) for the purpose that single branch or simple neural network can not enhance the brightness and contrast of image at the same time. MBLLEN can improve image quality in many ways.

The combination of CNN [30] and Retinex theory can further improve the visual quality of enhanced images, automatically learning the characteristics of the images and solving the problem of Retinex for being relying on manual setting of parameters. RetinexNet [12] included image decomposition and continuous image enhancement operations to improve image brightness. However, RetinexNet cannot estimate illumination correctly, and the edge information is lost when smoothing and denoising the reflectance image, which leads to the edge blur and color distortion of the enhanced image. In order to solve the problem of image color distortion, Ma et al. [31] combined the advantages of color model transformation and convolutional neural network, transformed the image from RGB color space to HSI space, and used deep convolutional neural network to enhance the image. Above methods are generally supervised learning which requires paired synthetic low illumination images and high-quality images as the ground truth to train the neural network. Obtaining high quality medical ground truth datasets is a big challenge. Using synthetic low illumination images adds additional problems of limited color and the image resolution range. Therefore, it is not ideal to train networks with synthetic datasets, especially in medical image enhancement.

Unsupervised learning: Generative adversarial networks (GAN) [32] provides a good foundation for unsupervised learning and have been widely applied in the field of image restoration. Inspired by GAN, CycleGAN [33] used unpaired data for network training without one-to-one matching pairs of images. At least two types of image datasets were required in CycleGAN to learn the mapping rules from one image to another. Guo et al. [34] proposed a none reference enhancement method Zero-DCE by setting a series of loss functions for the network end-to-end

training without reference images, which has good generalization for images under different illumination conditions. Then, the author redesigned the network structure of Zero-DCE and proposed an accelerated and light version Zero-DCE++ [35]. Jiang et al. [36] proposed an efficient EnlightGAN that can be trained without low-light/normal light image pairs, bringing great flexibility and adaptability for real life images in various fields. Although these unsupervised learning methods can solve the data set problem for low illumination image enhancement, the details of enhanced images still need to improve.

3 Background

This section presents a brief introduction to the necessary background of GC, CLAHE, LIME, and Color space involved in our algorithm.

3.1 Gamma correction

Gamma correction (GC), also known as exponential transformation or power transformation, is a commonly used gray non-linear transformation. It can not only change the brightness of the image, but also enhance the details, so that the overall effect of the image can be enhanced and improved. The basic form of gamma correction is as follows:

$$I_{gamma} = \lambda I^{\gamma} \tag{1}$$

where I is the brightness of the original image. I_{gamma} represents the image brightness after GC [3]. γ is the correction parameter, and its value directly determines the effect of GC. When $\gamma < 1$, GC can improve the brightness of the image and enhance the details in the dark areas of the image. Considering that the original intestinal polyp image has the characteristics of uneven illumination and low overall brightness. GC is used in the section 4.1 to generate the first derived image to adjust the global brightness of the original image. According to the characteristics of the original image, the parameter γ of gamma correction function is unified as 0.6, and the parameter λ is unified as 1.

3.2 Contrast-limited adaptive histogram equalization algorithm

Histogram based method can improve the contrast of image. Among them, the classical histogram equalization (HE) belongs to global equalization, which can cause over enhancement and color distortion for non-uniform illumination image processing. CLAHE is a local histogram equalization method, is an improvement of HE, it has better flexibility. The basic principle of CLAHE is: Firstly, the input image is divided into some sub-blocks, and the grayscale histogram of these sub-blocks is calculated. Then, a histogram clipping threshold is selected, and the gray level exceeding the threshold in the grayscale histogram of these sub-blocks is redistributed to other gray levels, which can avoid over-enhancement or over-amplification of the noise in the smooth area and reduce the boundary of artifacts. Next, histogram equalization is performed on the new histogram. Finally, the enhanced image is obtained by using bilinear interpolation between sub-blocks.

CLAHE [4] can significantly improve the contrast of the image, highlight the sense of layers of the image, and make the details of the image clear. The image collected by endoscopy are illuminated by a unidirectional point light source. Such light source may cause some areas to be brighten by direct illumination at a close range. In the section 4.1, CLAHE was selected to generate the second derived image to make the details of the local brightness area clear.

3.3 Low-light image enhancement via illumination map estimation

The traditional algorithm based on Retinex theory can significantly enhance the brightness of low illumination images, but the contrast of enhanced images is low and the color will deviate from the original image color. LIME is a new algorithm based on Retinex theory. The principle of this algorithm is to take the maximum value in the three RGB channels of the original low-illumination image, then modify the original illumination image continuously through the prior structure, and adjust the illumination map by gamma correction, so as to obtain the illumination map with globally smooth and clear edges.

LIME [5] is specifically proposed for low illumination images, which can significantly improve the contrast and clarity of images. The light source could also cause some areas to be darken in a long distance or back light. In the section 4.1, we incorporate LIME into the third derived image generation to make the details of the local dark area clear.

3.4 Color space and image details

Color spaces such as HSV, HSL, HSI and CIE L*a*b* are widely used in image enhancement. This section studies the brightness and detail information effect comparison of common color space in image enhancement technology and mainly introduces the related knowledge of HSI color space used in this paper.

The three color channels of RGB color space in the original images have strong correlation with brightness. When directly enhancing the image of RGB color space, it is generally necessary to enhance the R, G and B channels respectively, and then combine the three enhanced channels into the final enhanced image. If the enhancement of each channel is uneven, the inherent proportional relationship of R, G and B components will be changed, so the enhanced image is prone to produce color distortion. Transferring low illumination images of RGB color space into other color spaces and using relatively independent characteristics of each channel in the transferred color space to enhance the brightness component alone can improve the color distortion. This vision system based image processing can distinguish between the brightness and the information about hue and saturation [37]. HSL, HSV and HSI are color spaces composed of three attributes. H and S have the same meaning in the three color spaces, representing hue and saturation respectively. L, V, and I represent the lightness, value, and intensity of HSL, HSV, and HSI color spaces respectively. The CIE L*a*b* color space is a color model set by the International Commission on Illumination. L* component represents the brightness of pixels. a* and b* are two color channels. The L, V, I and L* components describe the brightness and detail information of the image, and the H, S, a* and b* components describe the color information of the image. The brightness information (L, V, I, and L* components) of the above color space closely match the brightness perception of human and are separated from the image color information (H, S, a* and b* components). Therefore, in the process of color image enhancement, scholars often convert the original image from RGB color space to HSL, HSV, HSI or CIE L* a* b* color spaces. Then, the brightness components of these color spaces are enhanced. Finally, it is converted back to RGB color space for storage and display.

Fig. 1 Comparison of brightness components of various color spaces of different exposure images.

A large number of experiments were conducted on 60 images with low illumination, high illumination and non-uniform illumination to compare the brightness components of various color spaces of different images. Considering of the layout limitation of the paper, we selected 3 representative images with different illumination conditions and their experimental results for display, which are low-illumination image 'Cloudy', high-illumination image 'Car' and non-uniform illumination image 'Sun'. As shown in Fig. 1, the original image Fig. 1 (a) is processed to obtain a single channel gray scale image Fig. 1 (b) to assess the brightness and detail information of the original image more intuitively without being influenced by the color components. We compare the characteristics between the brightness channels of CIE L*a*b*, HSV, HSI, and HSL color spaces and the gray scale image and select the color space that best match the brightness and detail information of the original image. The Just Noticeable Difference (JND) curve describes the minimum brightness deviation that the human eye can distinguish under different illumination conditions. According to the JND curve, the human eye is not sensitive to the brightness change in the high-light and low-light areas, but is sensitive to the brightness change in the middle-light (gray level

is about 128). Therefore, it is difficult for human eyes to observe the difference of the brightness components of the low-illumination image 'Cloudy' and the high-illumination image 'Car' in various color spaces. The non-uniform illumination image contains a large area of middle-light region. Next, we will mainly analyze the results for the image 'Sun'. It can be seen in Fig. 1 (c) that the brightness of the L component of HSL color space at the sun is darker than that of the original image, which makes some cloud textures with medium brightness level missing. On the contrary, the brightness of the V component of the HSV color space in the sun is brighter than that of the original image, which affects the texture details of the bright areas, as shown in Fig. 1 (d). The L* component of CIE L*a*b* color space also has the problem in that the brightness in the sun is brighter than that of the original image, which leads to the loss of the edge contour of the sun in the original image, as shown in Fig. 1 (e). The sun in the I component of HSI color space is the closest to the original color image in brightness, as shown in Fig. 1 (f).

Quantitative metrics can intuitively show the difference of brightness components of various color spaces of different images. In order to more formally analyze the color space selection, we used two quantitative metrics, Structural Similarity Index Measurement (SSIM) [38] and Lightness Order Error (LOE) to compare the integrity of structural information and the distortion level of luminance in an image. SSIM is an index to measure the integrity of the image structure information, which is actually composed of luminance, contrast and structure. The higher its value, the more similar the structure features between the test image and the reference image. LOE can objectively measure the luminance distortion level of the image. The smaller the LOE value, the more similar the lightness order of the test image and the reference image, that is, the better the naturalness of the test image is maintained. In this paper, Fig. 1 (b) was taken as the reference image to calculate the values of SSIM and LOE. It can be seen from the quantitative comparison of brightness channel images in different color spaces. The SSIM values of the HSI color space images are all larger than those of the other three color spaces, indicating that the brightness component of the HSI color space is more similar to Fig. 1 (b). The LOE values of HSI color space images are less than those of the other three color spaces, indicating that HSI color space maintains a better brightness order and the best naturalness. Therefore, HSI color space is more consistent with human visual characteristics than other color spaces. Keeping H unchanged, the operation of I and S will not affect the proportion of the primary color hue component, which can better maintain the image color and avoid color distortion. We choose HSI color space to process the endoscope image.

4 Methods

Fig. 2 shows the flow chart of our proposed endoscope image enhancement framework, which consists of three main modules: derived image generation, deep unsupervised fusion of intensity channel image and saturation channel image adjustment.

1) The derived image generation module mainly uses the classical image enhancement technology to generate three derived images. These three kinds of derived images have their own advantages in improving the detailed information of dark areas, enhancing the contrast of bright areas and improving the global brightness of the image.
2) The deep unsupervised fusion of intensity channel image module mainly uses our proposed network DerivedFuse to extract and fuse the features of the three derived images accurately.
3) The saturation channel image adjustment module mainly uses the saturation adjustment function constructed by us to process the saturation component S of HSI color space to enrich the color information of the endoscope image.

4.1 The derived image generation

Taking into account of the complex internal environment of human body and the characteristics of endoscopy imaging processing, we propose to use a combined image enhancement framework by generating three derived images from GC, CLAHE and LIME. The three derived images are combined to complement each other to achieve the contrast enhancement and illumination adjustment for endoscopy images. Then, three derived images are input to the deep neural network to perform the merger for enhancing the intensity of them accurately.

GC can improve the global brightness of the original image. However, when GC is used to adjust the global brightness of the image, the gray value of the resulting image is usually limited to a narrow gray scale, and the contrast is low, so it is easy to lose some details. CLAHE can enhance the contrast of high bright areas, but it is not obvious for dark areas. LIME has an obvious effect on improving the details of the dark areas in the image obviously. However, it will cause excessive enhancement to the high-brightness areas in the image. Therefore, only using a single GC, CLAHE or LIME to process endoscope images cannot achieve a comprehensive effect. The three derived images each has its own advantages in improving the detailed information of the dark areas, enhancing the contrast of bright areas and improving the global brightness of the image. In order to achieve global enhancement of

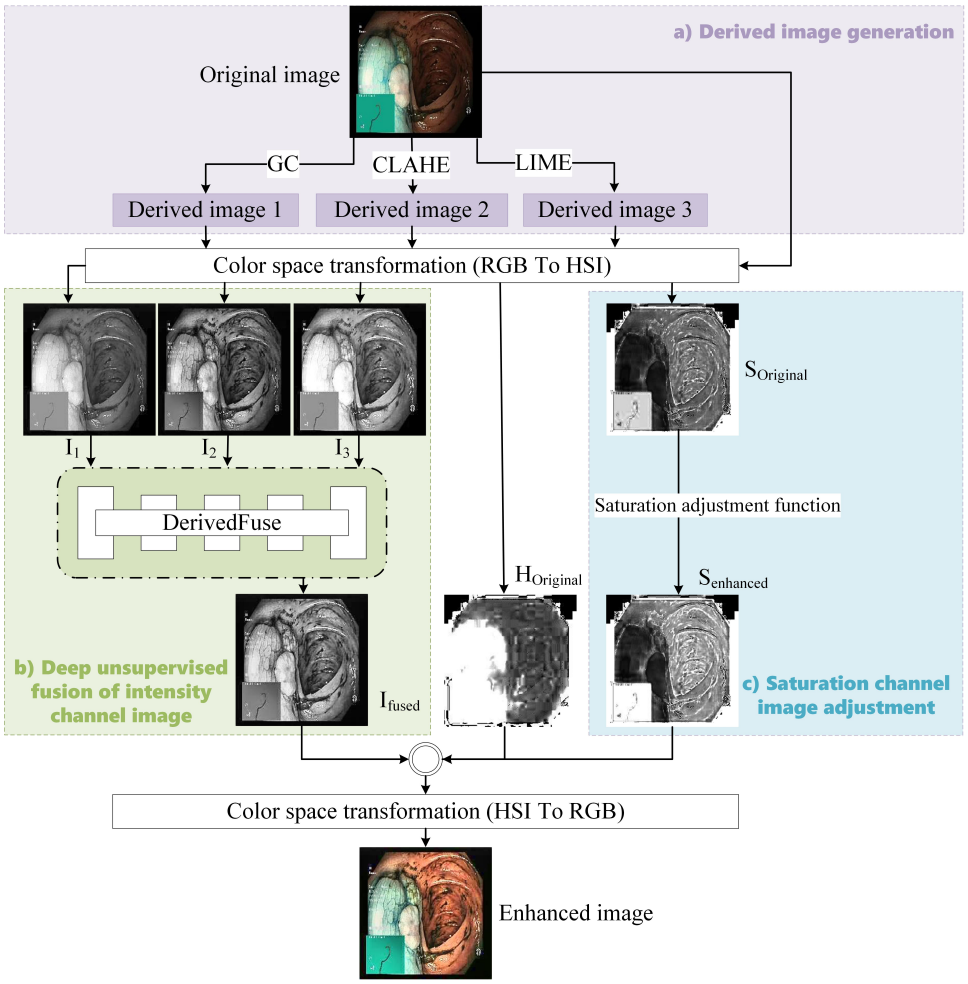


Fig. 2 Flow chart of our proposed endoscope image enhancement framework of three main modules: a) Derived image generation; b) Deep unsupervised fusion of intensity channel image; c) Saturation channel image adjustment.

contrast and brightness of endoscopic images, it is necessary to have these three derived images inputs to perform the merger.

4.2 DerivedFuse network structure

Considering the lack of datasets containing low-quality images and their corresponding high-quality image pairs of endoscopic images, we propose DerivedFuse fuses multiple derived images to effectively combine the advantages of the three derived images of GC, CLAHE and LIME. The I-channel images of three derived images are feed to DerivedFuse for enhancing the details of the original images. Different from the general deep learning method, DerivedFuse does not directly use the network to learn the end-to-end mapping relationship between the low-quality image and the high-quality image through training in RGB color space, but only enhances the intensity component I in HSI color space to obtain more details of original images.

GC mainly focuses on the brightness enhancement of the image and cannot solve the problem of image color restoration. CLAHE has a certain effect in enhancing the contrast of the image, but there will be the problem of color distortion. The overall color of the image enhanced by LIME is light, and the color information of the original image will be lost in the highlighted areas. Therefore, the saturation channel of the merged image will lose the color information of the original image. We directly use the saturation adjustment function to process the saturation component S of HSI color space of the original image instead of adjusting the saturation of the merged image, which will not produce color changes but enrich the color information of the image, so as to be more conducive to network training. DerivedFuse proposed in this paper is shown in Fig. 3. This model is mainly composed of image input, feature extraction, feature fusion, image reconstruction and image output.

1) Image input and image output. As shown in Fig. 3, I-channel images (I_1, I_2 and I_3) in HSI color space of three derived images are input into DerivedFuse model respectively. DerivedFuse only processes the brightness component, so the input image is the intensity component in HSI color space of three derived images with the

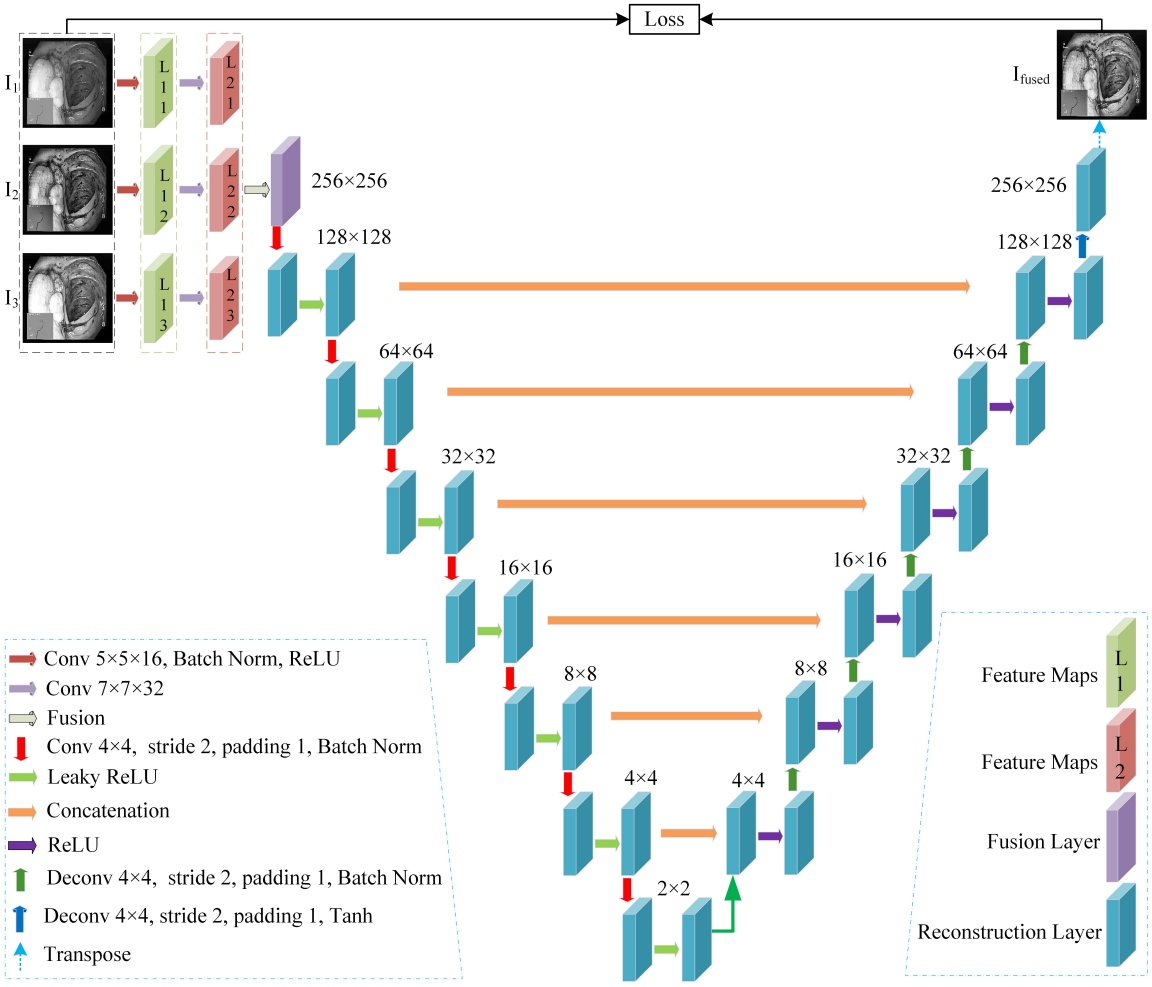


Fig. 3 Network structure of DerivedFuse.

number of channels of 1, and the output image (I_{fused}) is the fused intensity component in HSI color space with the number of channels of 1.

2) Feature extraction and feature fusion of the unsupervised deep net. The feature extraction module is composed of two layers of partial convolution layer. The size of the three convolution kernels of the first layer is 5×5, and the number of channels is 16. The size of the three convolution kernels of the second layer is 7×7, and the number of channels is 32. The 5×5 convolution kernel can be used to extract low-level features of image. By extracting feature maps with convolution of different scales, different features of the data can be extracted from the intensity component in HSI color space of three derived images, which improves the speed of network training. We simply add the pixel values at the same position of feature maps through the fusion layer to combine each feature map. The fusion layer is used to fuse the features of L21, L22 and L23, as shown in Fig. 3. The size of the fused feature map is 256×256 and contains 32 channels.

3) Image reconstruction. For the image reconstruction module, we use the U-Net model to extract the deep features of the fused feature map. The U-Net model consists of 7 upsampling layers and 7 subsampled layers respectively. The image reconstruction module input is the fused feature map with the channel number of 32 and the size of 256×256 output by the fusion layer. The output of the reconstruction module is the final fused intensity channel image with the channel number of 1 and the size of 256×256. The 7 subsampled convolution layers uses a convolution kernel of size of 4×4, the stride is 2, the padding is 1, and the Leaky LeRU activation function is adopted. The size of the convolution kernel used by the first six upsampling convolution layers is 4×4, the stride is 2, the padding is 1, and the LeRU activation function is adopted. The last deconvolution layer uses the Tanh activation function to produce the final intensity channel image with more complete details.

The specific parameter settings of DerivedFuse are shown in Table 1.

Table 1 The network structure details of DerivedFuse. Block represents the convolution block type. Layer represents the layer type. Conv represents convolution and Deconv represents deconvolution.

Block	#	Layer	Output size	Convolution kernel size	Convolution kernel number	Stride	Padding
Feature Extraction	Input	—	1*256*256	—	—	—	—
	1, 2, 3	Conv	16*256*256	5*5	16	1	2
	1, 2, 3	Conv	32*256*256	7*7	32	1	3
Feature Fusion	1	—	32*256*256	—	—	—	—
Reconstruction	1	Conv	64*128*128	4*4	64	2	1
	2	Conv	128*64*64	4*4	128	2	1
	3	Conv	256*32*32	4*4	256	2	1
	4	Conv	512*16*16	4*4	512	2	1
	5	Conv	512*8*8	4*4	512	2	1
	6	Conv	512*4*4	4*4	512	2	1
	7	Conv	512*2*2	4*4	512	2	1
	8	Deconv	512*4*4	4*4	512	2	1
	9	Deconv	512*8*8	4*4	512	2	1
	10	Deconv	512*16*16	4*4	512	2	1
	11	Deconv	256*32*32	4*4	256	2	1
	12	Deconv	128*64*64	4*4	128	2	1
	13	Deconv	64*128*128	4*4	64	2	1
	14	Deconv	1*256*256	4*4	1	2	1

4.3 New loss function-MDF SSIM

SSIM [38] is not only a quantitative index, but also is a widely used loss function that guides the network learning process by minimizing the difference between the enhanced image and the reference image, so that the structural information of the enhanced image is consistent with the reference image and speeds up the convergence of the network. By introducing structural loss SSIM can maintain image details and avoid image blur. However, it is impossible to use SSIM directly in this paper because it needs a single perfect quality reference image. DerivedFuse is an unsupervised network without ground truth reference images. MEF SSIM [14] calculates the loss without using reference images. However, our experimental results show that MEF SSIM is not directly suitable as the loss function in our network training, since it causes low contrast and brightness in the fused image, as well as the loss of texture details. In order to extract and fuse the features(luminance, contrast and structure) of each derived image more accurately, and make the details of the fused image clear, we improved MEF SSIM and proposed a new loss function MDF SSIM by taking into account of the characteristics of three derived images.

Let $\{i_n\} = \{i_n \mid n = 1, 2, 3\}$ denote the image patches extracted from the same spatial location in the input HSI color space luminance channel images of multi-derived images, i_f denote the patch extracted from DerivedFuse output fused image at same spatial location. The main goal of MDF SSIM is to calculate a value to evaluate the fusion performance of given i_n and i_f.

SSIM decomposes any given image patch into luminance, contrast and structure, then input patches i_n can be represented by Equation (2).

$$i_n = \|i_n - \mu_{i_n}\| \cdot \frac{i_n - \mu_{i_n}}{\|i_n - \mu_{i_n}\|} + \mu_{i_n} = \left\|\widetilde{i}_n\right\| \cdot \frac{\widetilde{i}}{\left\|\widetilde{i}\right\|} + \mu_{i_n} = c_n \cdot s_n + l_n \tag{2}$$

where $\|\cdot\|$ denotes the l^2 norm of a vector, μ_{i_n} denotes the mean value of i_n. l_n, c_n, and s_n roughly represent the luminance, contrast and structure components of i_n, respectively. In order to get a high-contrast fused image, the desired contrast $\widehat{c}$ of i_f is calculated from the maximum contrast of three different image patches in HSI color space luminance channel. The calculation formula of $\widehat{c}$ is as follows:

$$\widehat{c} = \max_{\{1 \le n \le 3\}} c_n = \max_{\{1 \le n \le 3\}} \left\|\widetilde{i}_n\right\| \tag{3}$$

In MEF SSIM, the desired structure $\widehat{s}$ of i_f is obtained by weighted sum of structures of all input different exposure image patches. MEF SSIM is suitable for extracting structural features of all multi-exposure images. GC derived image mainly improves the brightness of the original image, but the image contrast is low, and it is easy to lose part of the structure information. We want to combine only texture structures of CLAHE and LIME derived images to improve the contrast and clarity of the original image. In MDF SSIM, the desired structure $\widehat{s}$ of i_f is obtained by weighted sum of structures of input patches of luminance channel in HSI color space for CLAHE

derived image and LIME derived image as follows:

$$\begin{cases} \widehat{s} = \frac{\overline{s}}{\|\overline{s}\|} \\ \overline{s} = \frac{\sum_{n=1}^{2} \omega(\widetilde{i}_n) s_n}{\sum_{n=1}^{2} \omega(\widetilde{i}_n)} \\ \omega(\widetilde{i}_n) = \left\|\widetilde{i}_n\right\|^{tan(\frac{\pi}{2} \cdot \frac{\left\|\sum_{n=1}^{2} \widetilde{i}_n\right\|}{\sum_{n=1}^{2} \|\widetilde{i}_n\|})} \end{cases} \tag{4}$$

It is worth noting that the weighting function assigns equal ω to patches, when all input patches s_n have dissimilar structures. However, when all input patches s_n have similar structure components, the patch with high contrast is given more weight.

In MEF SSIM, the brightness of local patch is not obvious, so the desired fusion result $\widehat{i}$ does not consider the brightness. If the endoscopic image has the characteristics of low overall brightness, all the derived images can improve the overall brightness of the original image with our multi-image fusion approach based on the brightness information of all the derived images. The formula of $\widehat{l}$ is as follows:

$$\widehat{l} = \underset{\{1 \le n \le 3\}}{mean}\, l_n = \underset{\{1 \le n \le 3\}}{mean} \left\|\widetilde{i}_n\right\| \tag{5}$$

Finally, we can get the desired fusion result patch $\widehat{i} = \widehat{c} \cdot \widehat{s} + \widehat{l}$. Using SSIM framework to calculate the value to evaluate the fusion local image quality:

$$Quality = \frac{2\sigma_{\widehat{i}i_f} + C}{\sigma_{\widehat{i}}^2 + \sigma_{i_f}^2 + C} \tag{6}$$

where, $\sigma_{\widehat{i}}^2$ is variance, $\sigma_{\widehat{i}i_f}$ is covariance between $\widehat{i}$ and i_f. C is a constant. The total loss is calculated as,

$$Loss = 1 - Quality \tag{7}$$

4.4 Saturation adjustment

In order to avoid color changes in the original endoscopic image and enrich the color information of the original image, this paper constructed a function to adaptively adjust the saturation channel S of HSI color space and keep the hue component unchanged. If $S_{Original}$ represents the saturation of the original image and $S_{enhanced}$ is the saturation of the adjusted image. The mathematical expression of $S_{enhanced}$ is as follows:

$$S_{enhanced} = \alpha \times T \times S_{Original} \tag{8}$$

$$T = \frac{mean(R,G,B) + M(R,G,B) + m(R,G,B)}{mean(R,G,B)} \tag{9}$$

where the values of $M(R,G,B)$, $m(R,G,B)$ and $mean(R,G,B)$ are obtained adaptively in RGB color space. $M(R,G,B)$ represents the maximum value of pixels corresponding to R, G and B color channels in RGB color space. $m(R,G,B)$ is the minimum value of pixels corresponding to R, G and B color channels in RGB color space. $mean(R,G,B)$ is the average value of pixels corresponding to R, G and B color channels in RGB color space. α and T are parameter variables, which can control the enhancement degree of S channel image. The range of the α and T values is [0, 1] and [2, 6]. As shown in the Fig. 4, we get a large number of saturation adjustment images with different α values. Due to the space limitation, we choose four representative saturation adjustment images to enlarge and display. It can be seen that if the α value of image is too high or too low, the visual effect (color over-saturation or color loss) of the image will be affected in the Fig. 4. We according to the doctor's judgment to decide if the level of saturation is suitable for the required enhancement. A large number of experiments on the low-light gastrointestinal tract image dataset show that the α values corresponding to the enhancement results selected by doctors is approximately 0.6. When the α value is approximately 0.6, the color of the enhanced image is more bright than that of the original image, and the phenomenon of color distortion and over-saturation is avoided. Therefore, doctors can choose different α values to achieve different saturation enhancement effects according to their needs. Here, we set $\alpha = 0.6$, the saturation adjustment function will adaptively adjust the saturation according to the images with different T values.

Under the condition that the hue component H of HSI color space remains unchanged, the proposed saturation adjustment function is used to adaptively and non-linearly stretch the S-channel information of HSI color space according to the color information of the original image. Finally, the intensity component I_{fused} obtained in Section 4.2 and the saturation component $S_{enhanced}$ obtained in this Section are integrated into the original hue component $H_{Original}$ and reversed into the RGB color space to obtain the final enhanced color image with high contrast, uniform brightness, clear details and natural color.

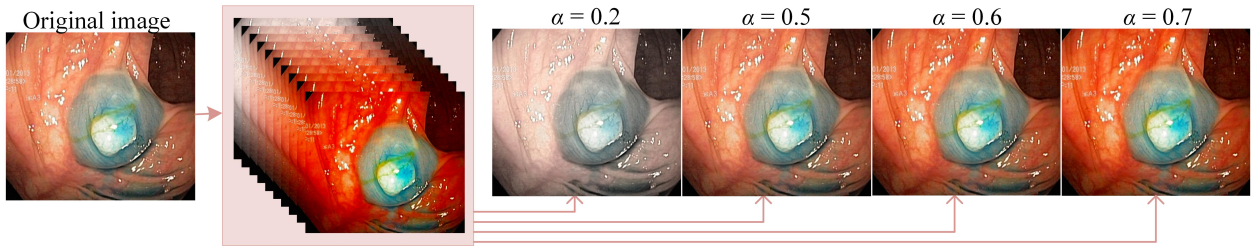


Fig. 4 The visual effect of different α values of image.

5 Results and Discussions

We use MATLAB2019 to generate the derived images and pytorch1.5.0 as the deep learning framework on an Inter(R) Xeon(R) E5-2620 CPU, 2.10GHz processor, 64GB RAM, and a Nvidia Titan Xp GPU. Under this environment, one training procedure takes 20 hours. We train our model using Adam optimizer and set the learning rate to 0.0002. The batch size is set to 16, and the iteration number is 100.

5.1 Training Data

The training data directly determines what mapping rules the deep CNN can learn. We evaluated our framework on the open endoscopic image data set of the gastrointestinal. There are 868 authentic low-light gastrointestinal tract images from the open source data set, including 325 endoscopic images from Kvasir Dataset [39], 113 images from Kvasir-SEG [40], 93 images from CVC-ClinicDB [41], and 26 images from ETIS-Larib Polyp DB [42], 299 images in CVC-EndoSceneStill [43] and 12 images in CVC-ClinicSpec [44]. All these images are resized to 256× 256 pixels. These images are trained, verified and tested respectively.

5.2 Experimental Results and Analysis

Fig. 5 shows the loss graph in Equation (7) during the training procedure. As shown in Fig. 5, it can be seen that the loss function value of our network is large at first, and then rapidly converges to a stable range as the iterative number increasing, and the fluctuation range is small during this period, which indicates our method is stable. After about 100 iterations, the loss function value of our network tends to be stable, so we set the iteration number to 100 during network training.

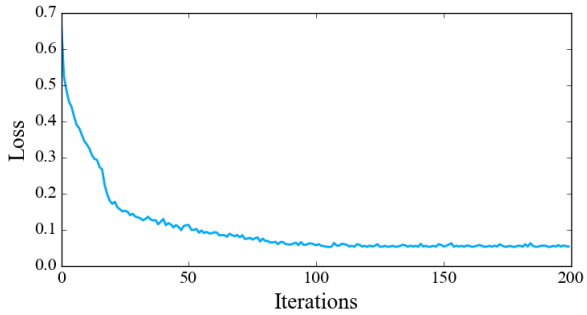


Fig. 5 Convergence map of training. The abscissa represents the number of iterations, and the ordinate represents the loss function value of our network.

5.2.1 Compare with the conventional methods and unsupervised learning methods

We compare the performance of different methods on the same images, including HE [20], CLAHE [4], SSR [26], MSRCR [25], MSRCP [27], RRM [28], LIME [5], AGCWD [22], Al_Ameen [23], Zero-DCE [34], Zero-DCE++ [35], EnlightenGAN [36]. We use the recommended parameter settings in the original papers, and the parameter Lambda proposed by the reference [23] are unified set to 5. We randomly select 167 gastrointestinal tract images from the data set as the testing set A, and get the results of 13 algorithms respectively. Testing set A was composed of various types of gastrointestinal tract images obtained by endoscopy, including normal cecum images, pathological

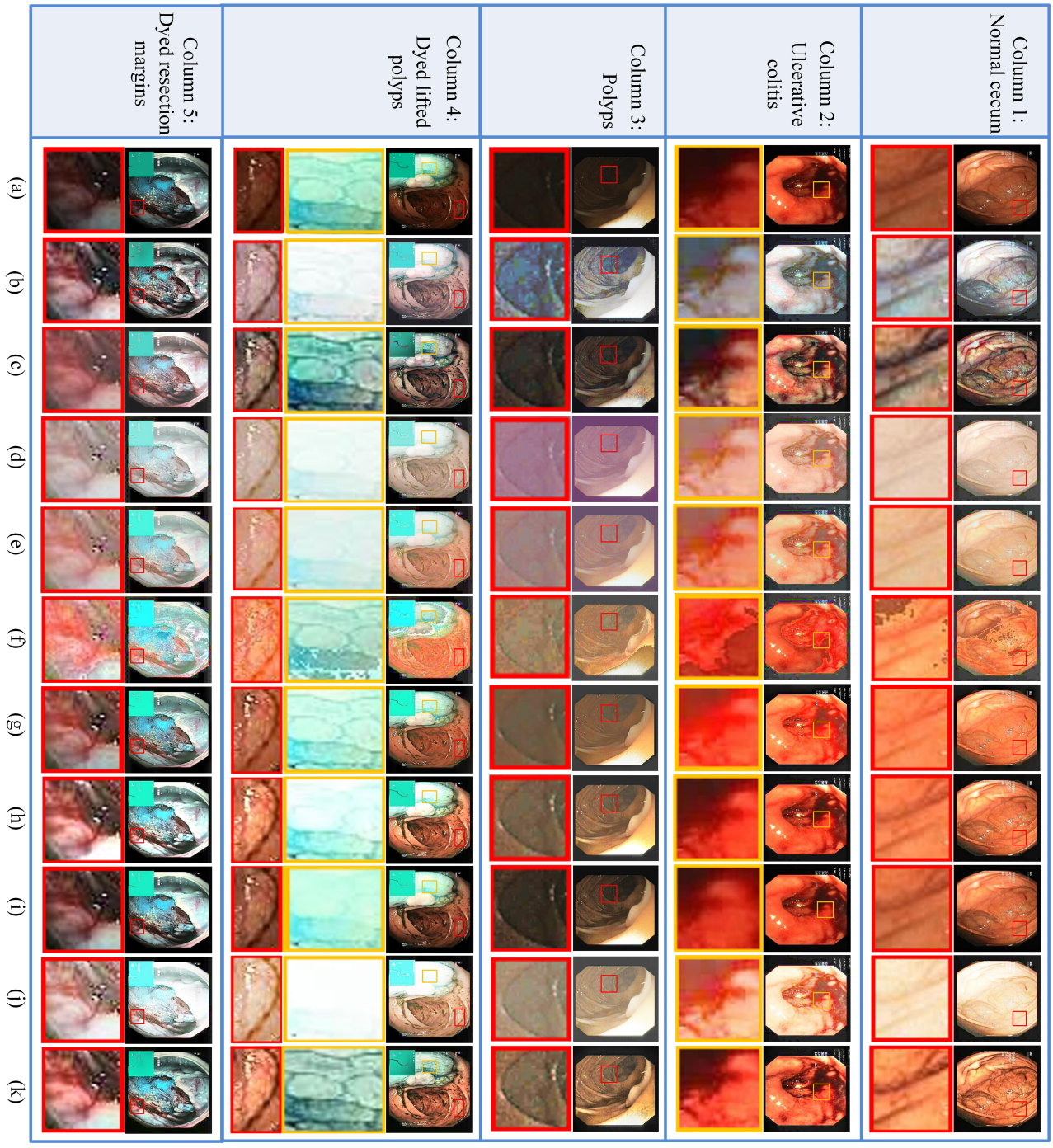


Fig. 6 Visual comparison of different conventional image enhancement methods for gastrointestinal tract images of the testing set A. (a) Original image; (b) HE; (c) CLAHE; (d) SSR; (e) MSRCR; (f) MSRCP; (g) RRM; (h) LIME; (i) AGCWD; (j) Al_Ameen; (k) Proposed method.

finding images of polyps and ulcerative colitis. In addition, it includes images related to removal of lesions, e.g., "dyed and lifted polyp", the "dyed resection margins".

Fig. 6 shows the visual comparison of the proposed method and conventional image enhancement methods on the testing set A for processing gastrointestinal images. Column 1 shows an endoscopic image (containing an appendiceal orifice) of a normal cecum observed from the inside of the intestine. Column 2 shows an example of ulcerative colitis with bleeding and swelling of the mucosa. An endoscopic image of polyp is shown in Column 3. Column 4 illustrates an endoscopic image of a polyp lifted by injection of saline and indigocarmine. Column 5 shows the resection site after removal of a polyp by injection of saline and indigocarmine. Analysis from the subjective observation, the method in HE (shown in Fig. 6(b)) clarify the images well but cannot effectively correct the color of gastrointestinal images, since the overall color is slightly white as well as the bright region is over-enhanced. That's because this algorithm belongs to the histogram equalization algorithm, which can improve the contrast

of the image but has little effect on color correction. The method of CLAHE (shown in Fig. 6(c)) and AGCWD (shown in Fig. 6(i)) can improve the brightness of the image to a certain extent, but the overall brightness is still low. The CLAHE has little effect on enhancing details in dark region, while the AGCWD may cause the loss of some details owing to over-enhancement in bright region. The images enhanced by SSR (shown in Fig. 6(d)) and MSRCR (shown in Fig. 6(e)) methods are not natural, which is slightly purple or white compared from the original image. The enhancement results of MSRCP (shown in Fig. 6(f)) are generally too bright, and the produced artifacts in the bright region cause the loss of texture details. The RRM (shown in Fig. 6(g)) significantly improves the brightness of the image and makes the color bright, but over-enhances the bright region which resulting in a loss of details. The LIME (shown in Fig. 6(h)) preserves the color and edge well, but fails to enhance the texture details at the bright region. The Al_Ameen (shown in Fig. 6(j)) method performs well on improving the brightness, but has the problem of overexposure, which degrade the image details. By contrast, our method remarkably increases the overall brightness and preserves the details well in both dark region and bright region. More specifically, after the enhancement of our method, we can more clearly see the position of the appendiceal orifice on the normal cecum image, the mucosal bleeding and swelling on the ulcerative colitis image, the polyp site on the polyp image, the light blue polyp margins against the darker normal mucosa on the dyed lifted polyp image and the polyp resection margin on the dyed resection margin image. In addition, our method avoids the color distortions while enriching the color information of the original images, which has a better performance.

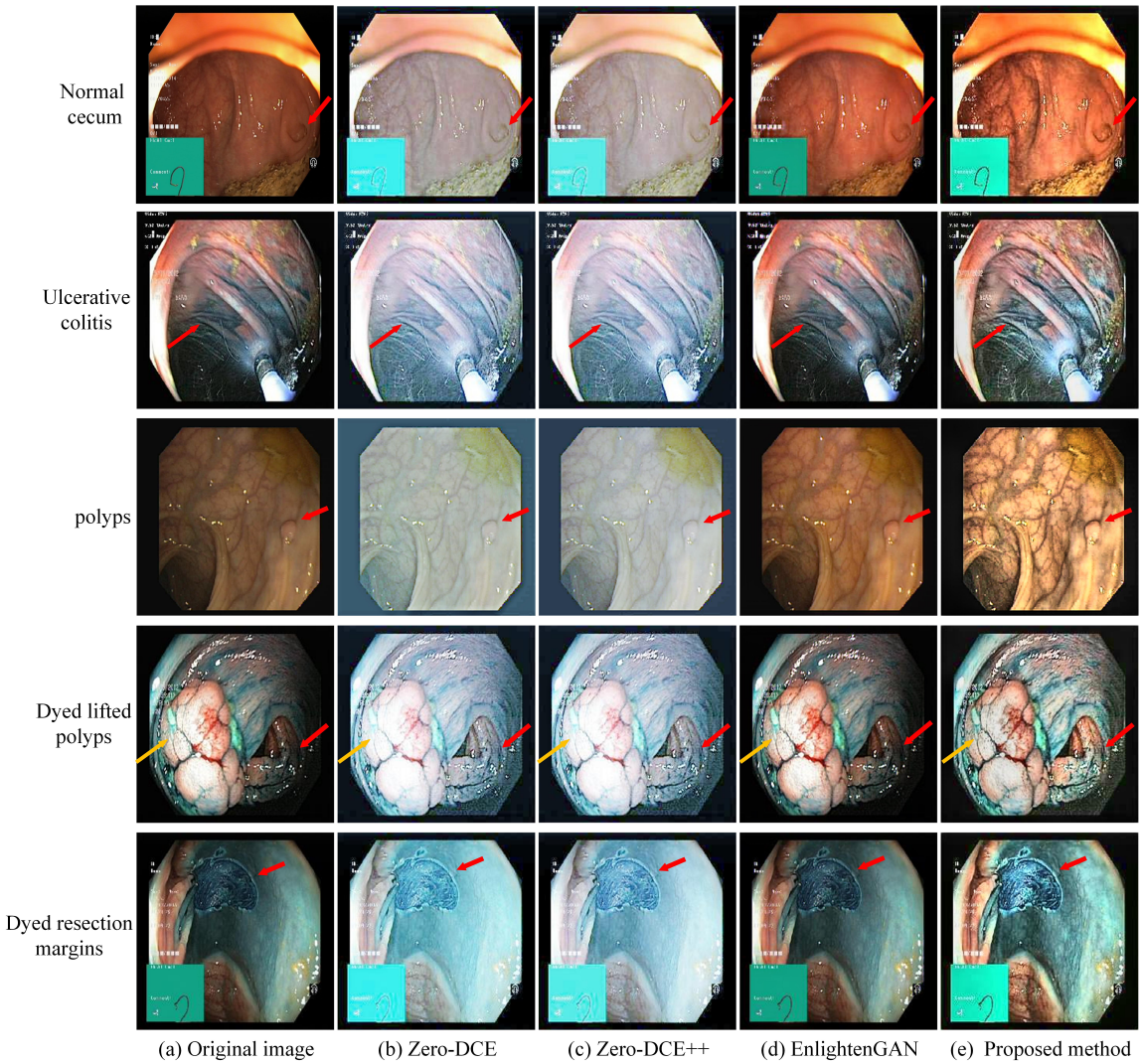


Fig. 7 Visual comparison of different unsupervised learning methods for gastrointestinal tract images of the testing set A. (a) Original image; (b) Zero-DCE; (c) Zero-DCE++; (d) EnlightenGAN; (e) Proposed method.

Fig. 7 shows the visual comparison of the proposed method and unsupervised leaning methods on the testing set A for processing gastrointestinal images such as normal cecum images, ulcerative colitis images, polyp images, dyed lifted polyp images and dyed resection margin images. It can be seen from Fig. 7 that Zero-DCE (shown

in Fig. 7(b)) and Zero-DCE++ (shown in Fig. 7(c)) greatly improves the overall image brightness, but some image edges get blurry, the overall tone is white and the bright region is over-enhanced after enhancement. The enhancement effect of EnlightenGAN (shown in Fig. 7(d)) is better than Zero-DCE and Zero-DCE++ on the color and contrast. However, on dark regions, that is not remarkable. Compared with these algorithms, the results of our method have moderate brightness, rich texture details and natural visual effect.

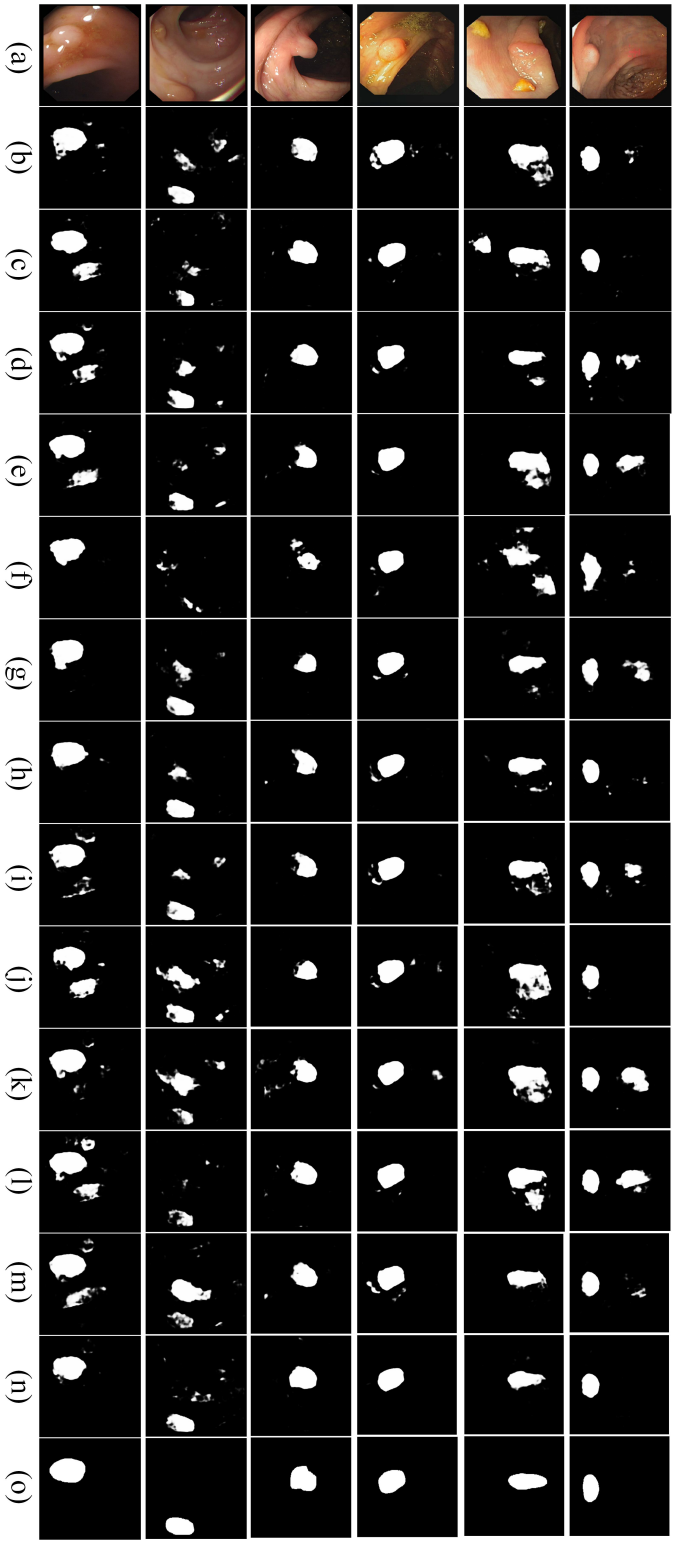


Fig. 8 The salient object detection results of different image enhancement methods for example images on CVC-EndoSceneStill dataset. (a) Example images; (b) HE; (c) CLAHE; (d) SSR; (e) MSRCR; (f) MSRCP; (g) RRM; (h) LIME; (i) AGCWD; (j) Al_Ameen; (k) Zero-DCE; (l) Zero-DCE++; (m) EnlightenGAN; (n) Proposed method; (o) Polyp mask (Ground truth).

In order to verify the significant of the results of the proposed method, we made a significance analysis. This study performed 13 experiments. The proposed method and other comparison algorithms are used to enhance the images in the CVC-EndoSceneStill dataset. The trained with enhanced CVC-EndoSceneStill dataset using the same configuration so that we can evaluate the impact of the 13 datasets on the significance detection performance of U-Net [45]. Fig. 8 displays the comparison of polyp detection results for the 13 methods with the ground truths. Fig. 8 (o) is the ground truth for the polyps corresponding to the original example images (shown in Fig. 8 (a)). This ground truth consists of a mask corresponding to the region covered by the polyp in the image. Fig. 8 (b)-(n) shows the salient detection results obtained from the image enhanced by 13 methods. It can be shown that our attention focusing more on the lesion area of the polyp images after using the proposed method to enhance the image. The proposed method significantly improves the saliency detection performance. Therefore, the enhancement method proposed by this paper could have applications in the performance optimization for salience detection.

5.2.2 Compare with the supervised learning methods

We use natural image datasets to evaluate our proposed method, and compare the test results with the supervised learning methods RetinexNet [12] and MBLLEN [29]. One purpose is to verify the better enhancement effect of our proposed method than the supervised learning methods. The other is to verify that the proposed method is not only effective for low illumination endoscope images, but also can enhance the real low illumination natural images, and has better robustness.

In this paper, Low-Light datasets (LOL) [12] is selected as the training set, which includes 500 real scene low/normal light image pairs and 1000 synthetic ones. And, we choose the public low illumination image data set DICM [46] as the testing set B. DICM contains 64 real low illumination images such as night scene image, backlight image and non-uniform illumination image.

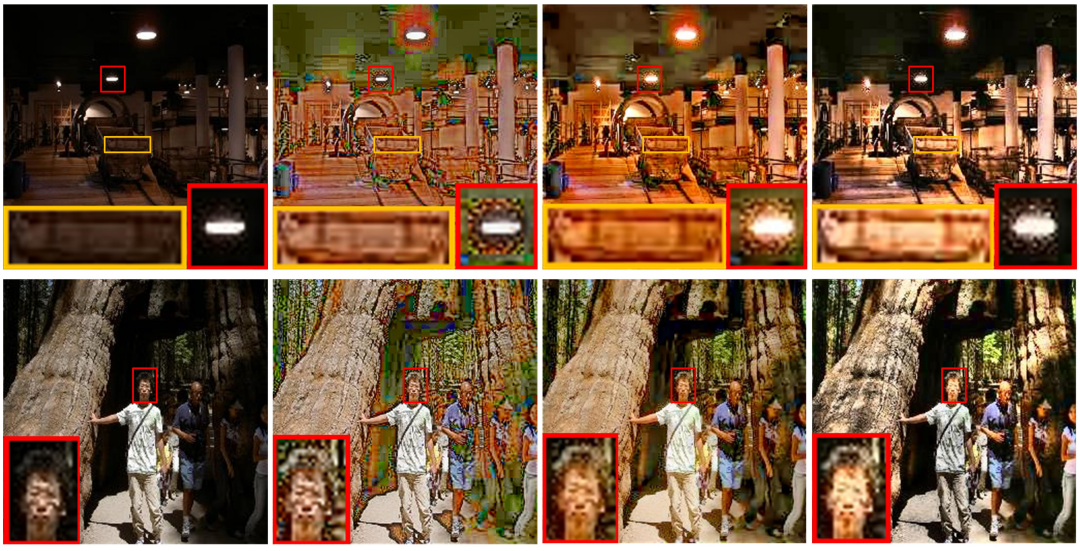


Fig. 9 Visual comparisons of different supervised learning methods and our method on a low-light image sampled from the testing set B.

Fig. 9 is the visual comparison of the proposed method and supervised leaning methods on the testing set B. The RetinexNet would produce problems of artifacts, noise and the image texture details are not clear on dealing with the dark areas in low-light natural images, as shown in Fig. 9(b). The MBLLEN would produce a lot of noise when enhancing the extremely dark area near the light, and the over-enhancement problem when dealing with the bright area in the low-light image. In addition, the overall hue of the image processed by MBLLEN is reddish, which makes the enhanced image look unnatural, as shown in Fig. 9(c). The proposed method can restore the clear image from the dark image, and keep the details and texture of the image well. Moreover, the contrast and brightness of the image have also been well enhanced. This shows that our proposed method is suitable for low light image enhancement task, and performs the better result than the supervised learning methods.

Table 2 Entropy, Contrast, Sharpness metrics of each algorithm on polyp images of the testing set A. ↑ means that the larger the value of the corresponding objective index, the better the enhancement result.

Algorithm	Entropy↑	CII↑	AG↑
HE[20] (2002)	6.6176	0.9443	6.1487
CLAHE[4] (1994)	7.3892	1.2331	7.8895
SSR[26] (1997)	6.8867	0.6180	4.2280
MSRCR[25] (1997)	7.1956	0.5870	4.3810
MSRCP[27] (2014)	6.8261	0.6000	7.2964
RRM[28] (2016)	7.2351	0.8950	5.1143
LIME[5] (2017)	7.2086	0.9814	6.1469
AGCWD[22] (2013)	7.1229	1.0275	4.8848
Al_Ameen[23] (2019)	6.4152	0.7865	5.0396
Zero-DCE[34] (2020)	7.1611	0.8002	5.4717
Zero-DCE++[35] (2021)	7.1734	0.8180	5.3344
EnlightenGAN[36] (2021)	7.1087	1.1710	4.5634
Our method	**7.6314**	**1.3095**	**8.5074**

5.3 No-Referenced Image Quality Assessment

To measure the results quantitatively, we need to test the objective evaluation indexes of the experimental results. Different from synthetic images, the real polyp images and natural images don't have corresponding high-quality images of the same scene, so full reference image quality indexes such as MSE and PSNR cannot be tested. We use three No-Referenced Image Quality Assessment indexes to evaluate the image quality including: Entropy, Contrast Improvement Index(CII), Average Gradient(AG).

Entropy is a statistical form of feature, which reflects the quantity of average information in an image. The value of entropy represents the amount of image information, that is, the detail of the image. A higher entropy value represents more information and details in image. Entropy is expressed as:

$$Entropy = -\sum_{i=0}^{255} p_i log p_i \tag{10}$$

where p_i indicates the probability of i-th gray level.

CII can be used to reflect the degree of contrast enhancement before and after image enhancement. The larger the value is, the more obvious the contrast enhancement is. CII is defined as:

$$CII = C_{enhancement}/C_{original} \tag{11}$$

We divide the image into 3×3 small patches, where C is the average of local contrast measured by a window of 3×3, $C_{enhancement}$ and $C_{original}$ represent the average of local contrast in output and original image, respectively. The contrast can be defined as: $C = (max - min)/(max + min)$, where max and min are the maximum and minimum gray value of the image patch.

AG is used to indicate the image clarity, which reflects the sharpness and texture change of image. The higher the value, the clearer the image. The formula of AG is as follows:

$$AG = \frac{1}{M * N} \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{\frac{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}{2}} \tag{12}$$

where $M * N$ is the image size, $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ represents the gradient in the horizontal and vertical directions respectively.

5.3.1 Evaluation of endoscopic images

Table 2 shows the average evaluation value of the three objective indexes of the endoscopic image enhancement results processed by the 13 algorithms on testing set A, including Entropy, CII and AG. The results demonstrate that our method has the best performance on Entropy, which indicates that the proposed method enlarges the image information further after the enhancement of the polyp image. Thus more image information can be extracted, which is beneficial for detail display. However, the Entropy values of HE, SSR, MSRCP and Al_Ameen were low, indicating the details and textures aren't enhanced obviously. Moreover, the CII average value of the proposed algorithm is also the highest among the 13 algorithms, indicating that the proposed method has a better performance on contrast enhancement. The average CII values of SSR, MSRCR, MSRCP, Al_Ameen, Zero-DCE and Zero-DCE++ were slightly low among the 13 algorithms, which indicates that the contrast improvements of SSR, MSRCR, MSRCP, Al_Ameen, Zero-DCE and Zero-DCE++ on polyp images are not significant. More

Table 3 Entropy, Contrast, Sharpness metrics of each algorithm on natural images of the testing set B. ↑ means that the larger the value of the corresponding objective index, the better the enhancement result.

Algorithm	Entropy↑	CII↑	AG↑
RetinexNet[12] (2018)	7.3983	0.8702	12.2892
MBLLEN[29] (2018)	7.3875	2.1162	9.7879
Our method	**7.5398**	**4.6192**	**14.0950**

Table 4 Background of the participants.

Backgroud	Category	Number of cases
Gender	Male	9
	Female	7
Work seniority/y	≥3	6
	<3	10
Job	Surgeon	6
	Medical student	10
Evaluation time/min	>30	5
	20~30	6
	<20	5

specifically, as shown in Table 2, after processing by the 13 methods respectively, the AG value of the proposed method is obviously higher than that of other methods, indicating that the image enhanced by our method is relatively clearer. While the AG results of AGCWD, SSR, MSRCR and EnlightenGAN were all low, indicating that the enhancement results were not clear enough. The Entropy value of our method is 3.27% higher than the optimal entropy value of comparison algorithms. The CII of our proposed method is 6.19% higher than that of comparison algorithms. The AG of our method is 7.83% higher than the optimal AG of comparison algorithms. In conclusion, the image enhanced by the proposed method contains richer information, and the contrast and sharpness are also greatly improved. It can be seen that the proposed method has achieved significant effect on enhancing the brightness and contrast of polyp images with different contents.

5.3.2 Evaluation of natural images

Table 3 shows the average evaluation value of the three objective indexes of the natural image enhancement results processed by the two algorithms on testing set B, including Entropy, CII and AG. From Table 3, the proposed method achieves the optimal values in Entropy, CII and AG. The Entropy value of our method is 1.91% higher than the optimal entropy value of comparison supervised learning algorithms. The CII value of our proposed method is 118.27% higher than the optimal CII value of comparison supervised learning algorithms. The AG value of our method is 14.69% higher than the optimal AG value of comparison supervised learning algorithms. This shows that the detail information, contrast and clarity of the low light natural image are improved significantly after the enhancement by our proposed method.

5.4 User Study

To evaluate the results by 13 enhancement methods on endoscopic images, we consulted 6 colorectal surgeons with more than 3 years of clinical experience and 10 medical students, and obtained their evaluation feedback on the enhancement results of 8 images, randomly selected from the testing set A, as shown in Table 4.

We asked these 16 participants to evaluate the enhancement results of the 13 methods on endoscopic images independently. The concrete steps were taken: participants were shown 16 images for every method, including 8 original polyp images and their corresponding enhanced images. 6 questions were designed in the investigation sheets (1. Are the details of the lesion area clear? 2. Is the bright area of the image over-enhanced? 3. Does the image contain underexposed artifacts? 4. Is there color distortion? 5. Does the result look realistic? 6. Whether it can be applied to clinical diagnosis?). The participants were asked to take these 6 questions into consideration when evaluating. The criteria were used to evaluate our method subjectively, and the score of each question in the range of 1-5 (5 represents the most satisfying and 1 represents the least satisfying) was given. At last, we can get feedback summary based on the evaluation of the participants.

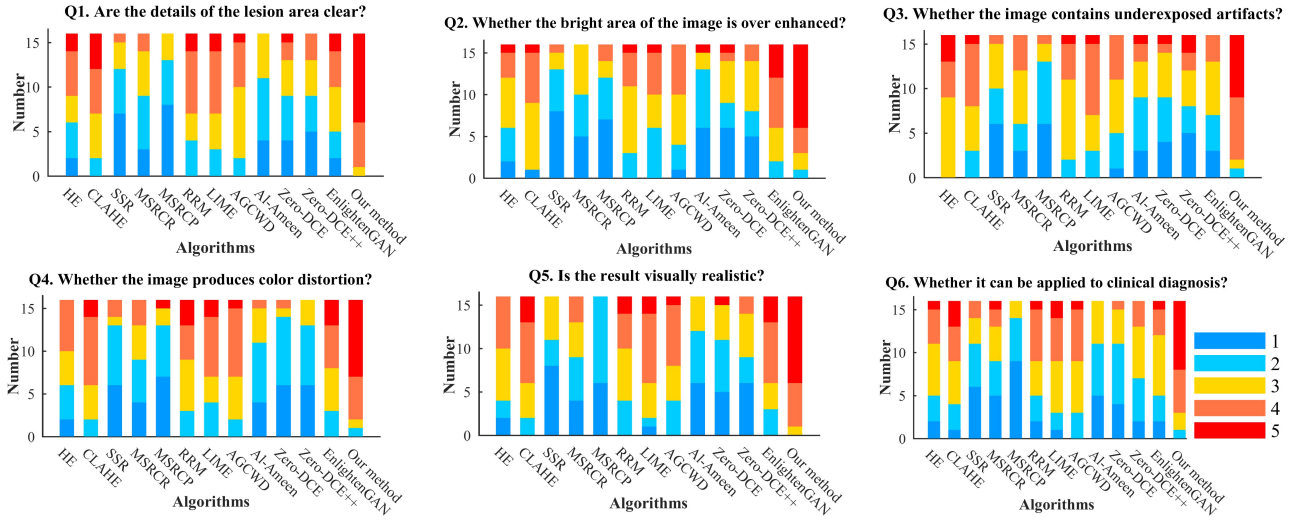


Fig. 10 Results of our user study. Each color bar shows the number of images in each score index.

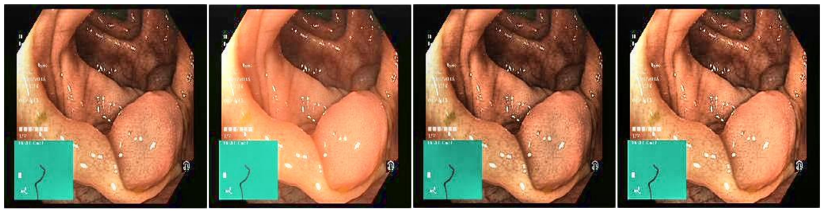
The results of our user study are shown in Fig. 10. Among them, red, orange, yellow, cyan and blue represent scores of 5, 4, 3, 2 and 1 respectively. The distribution of different methods shows that surgeons are favorably on the results of our method. Compared with other methods, our method gets more "red" and "orange" ratings, and much less "blue" and "cyan" ratings. The scores of SSR and MSRCP were very low. This is because they cause color distortion and over-enhancement, which is not conducive to people to observe the details of polyp image. Obviously, our method is superior to other methods, and these statistics are consistent with our analysis of Visual Quality Comparison and No-Referenced Image Quality Assessment. From the feedback of Q6, we can see that the algorithm proposed in this paper can be applied to clinical diagnosis.

5.5 Ablation Study

We perform ablation studies to demonstrate the effectiveness of each component of our method on the same polyp images. The evaluation is performed on the testing set A.

(1)Ablation study on derived images

To demonstrate the effect of each derived image proposed in Section 4.1, we carry out the ablation experiment on derived images. Fig. 11 shows the enhancement results of removing the three derived images respectively and the enhancement results of our method. Compared with Fig. 11(d), the overall brightness of the image is reduced as shown in Fig. 11(a), which indicates that the GC derived image can significantly boost the performance of the overall brightness. According to the lining texture of the large intestine in Fig. 11(b) and Fig. 11(d), it can be seen that the contrast of the bright part of the image is reduced when removing the CLAHE derived image, which demonstrate the importance of the CLAHE derived image. Fig. 11(c) shows the enhancement results after removing the LIME derived image. After the removal, the image details in the dark region are not clear, which indicates the use of the LIME derived image improves the texture details in the dark region. Average Entropy, Contrast, and Sharpness scores of testing set A generated by DerivedFuses trained with different derived images are shown in Table 5. Results illustrate that adding three derived images from GC, CLAHE and LIME yield better performance.



(a) Without GC (b) Without CLAHE (c) Without LIME (d) Enhanced result

Fig. 11 Ablation experiment results on derived images.

Table 5 Objective index results in terms of Entropy, Contrast, and Sharpness. These comparisons are carried out on testing set A. ↑ means that the larger the value of the corresponding objective index, the better the enhancement result.

Method	Entropy↑	CII↑	AG↑
Without GC	7.3894	1.0027	7.0882
Without CLAHE	7.2753	0.8310	5.9575
Without LIME	7.3696	1.2034	7.5019
Enhanced result	**7.6314**	**1.3095**	**8.5074**

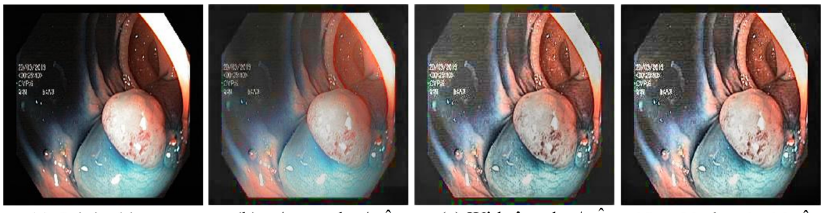
Table 6 Entropy, Contrast, and Sharpness for DerivedFuse trained with different combinations of losses. These comparisons are carried out on testing set A. ↑ means that the larger the value of the corresponding objective index, the better the enhancement result. "w/o" means without.

Loss	Entropy↑	CII↑	AG↑
w/o $\widehat{s}$, w/o $\widehat{l}$	7.3524	0.6458	4.3380
with $\widehat{s}$, w/o $\widehat{l}$	7.4731	0.8132	7.9844
with $\widehat{s}$, with $\widehat{l}$	**7.6314**	**1.3095**	**8.5074**

(2)Ablation study on loss functions

To validate the effect of the MDF SSIM proposed in Section 4.3, we compare the performance of DerivedFuse trained by various combinations of losses. As shown in Fig. 12, the first column shows the original image. The second column shows the images generated by MEF SSIM without the local structure $\widehat{s}$ of Equation (4) and the local brightness $\widehat{l}$ of Equation (5). MEF SSIM fuses structural features of all derived images. The third column shows the generated images with the local structure $\widehat{s}$ of Equation (4) but without the local brightness $\widehat{l}$ of Equation (5). $\widehat{s}$ of Equation (4) is obtained by extracting structural features of CLAHE and LIME derived images. The fourth column shows our proposed MDF SSIM, which with the local structure $\widehat{s}$ of Equation (4) and the local brightness $\widehat{l}$ of Equation (5). Observing the results in Fig. 12(b), we can found that the MEF SSIM loss function does not consider different image feature structures and fuses all of the structure components of the input patches, that leads to the problems of low brightness, low contrast and losing the internal texture details of the original image after enhancement. In Fig. 12(c), it only fuses the structure components of CLAHE derived image and LIME derived image, and does not consider the brightness of local patches. That could improve the contrast and clarity of the image, but reduce the brightness of the highlighted regions on the polyps.

And our method MDF SSIM not only fuses the structure components of CLAHE derived image and LIME derived image, but also considers the brightness of local patches. That significantly improves the brightness, contrast and clarity of the image and enriches the detailed texture information of the original image, as shown in Fig. 12(d). Average Entropy, Contrast, and Sharpness scores of testing set A generated by DerivedFuses trained with different loss functions are shown in Table 6. The Entropy, Contrast, and Sharpness values of the MDF SSIM are optimal. These results demonstrate that MDF SSIM plays a key role in preserving the detailed information of the original image.

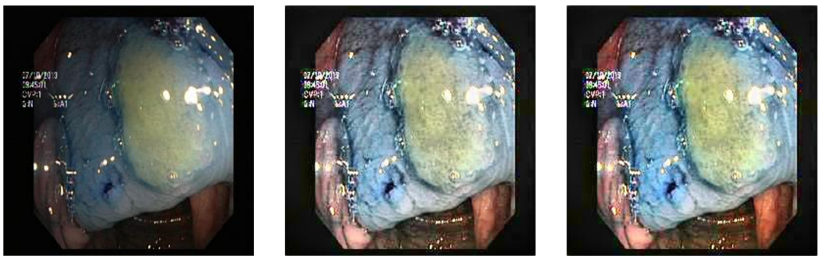


(a) Original image (b) w/o $\hat{s}$ and w/o $\hat{l}$ (c) With $\hat{s}$ and w/o $\hat{l}$ (d) With $\hat{s}$ and with $\hat{l}$

Fig. 12 Ablation experiment results on loss functions."w/o" means without.

(3)Ablation study on saturation adjustment function

In order to validate the effect of the saturation adjustment function proposed in Section 4.4, we also conduct an ablation experiment on Saturation adjustment function. Fig. 13(a) is the original image. Fig. 13(b) shows the enhancement effect of endoscope image without using saturation adjustment function. Fig. 13(c) shows the effect of endoscopic image enhancement using our proposed saturation adjustment function. We can see that it enriches the color information of the original endoscopic image by using our saturation adjustment function. In order to objectively evaluate the quality of the enhanced image by the proposed saturation adjustment function. We use the average saturation (Saturation, SAT) to evaluate the image processing effect from the perspective of color. It can be seen from Table 7 that the saturation value of the image processed by the proposed saturation adjustment function is the most considerable.



(a) Original image (b) Without saturation adjustment (c) Saturation adjustment

Fig. 13 Ablation experiment results on saturation adjustment function.

Table 7 Objective index results in terms of Saturation. These comparisons are carried out on testing set A.

Method	SAT↑
Without saturation adjustment	0.3115
Saturation adjustment	**0.3945**

6 Conclusion

This paper proposes a new deep unsupervised endoscopic images enhancement method based on multiple derived image fusion. Fourteen state-of-the-art algorithms have been compared with our proposed method. The experimental results on public endoscopic image datasets show that the Entropy value of our method is 3.27% higher than the optimal entropy value of comparison algorithms. The CII of our proposed method is 6.19% higher than that of comparison algorithms. The AG of our method is 7.83% higher than the optimal AG of comparison algorithms. The experimental results on public natural image datasets show that the Entropy value of our method is 1.91% higher than the optimal entropy value of comparison supervised learning algorithms. The CII value of our proposed method is 118.27% higher than the optimal CII value of comparison supervised learning algorithms. The AG value of our method is 14.69% higher than the optimal AG value of comparison supervised learning algorithms. In addition, sixteen medical doctors and students have evaluated our proposed algorithm for clinical endoscopic diagnosis and treatment.

In the future work, we will improve the method and apply it to the segmentation of lesions in endoscopic images and the detection of lesions in endoscopic images.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by the Shanghai Natural Science Foundation of China under Grant No.19ZR1419100 and the National Natural Science Foundation of China under Grant No.61402278.

References

1. D.F. Shen, J.J. Guo, G.S. Lin, J.Y. Lin, Content-aware specular reflection suppression based on adaptive image inpainting and neural network for endoscopic images, Comput. Methods Programs Biomed. 192(2020) 105414, doi: 10.1016/j.cmpb.2020.105414.
2. J. Zhang, M. Han, Y. Dai, Three-dimensional porous structure reconstruction for low-resolution monocular endoscopic images, Optics and Precision Engineering 28(9)(2020) 2085-2095.
3. H. Farid, Blind inverse gamma correction, IEEE Trans. Image Process. 10(10)(2001) 1428-1433.
4. K. Zuiderveld, Contrast limited adaptive histogram equalization, Graphics gems (1994) 474-485.
5. X. Guo, Y. Li, H. Ling, LIME: Low-light image enhancement via illumination map estimation, IEEE Trans. Image Process. 26(2)(2017) 982-993, doi: 10.1109/TIP.2016.2639450.
6. W. Xia, E.S. Chen, T. Peters, Endoscopic image enhancement with noise suppression, Healthc. Technol. Lett. 5(5)(2018) 154-157, doi: 10.1049/htl.2018.5067.
7. M. Long, X. Xie, G. Li, Z. Wang, Wireless Capsule Endoscopic Image Enhancement Method Based on Histogram Correction and Unsharp Masking in Wavelet Domain, 2019 17th IEEE International New Circuits and Systems Conference (NEWCAS), 2019, pp. 1-4, doi: 10.1109/NEWCAS44328.2019.8961243.
8. W.S. Liew, T.B. Tang, C.H. Lin, C.K. Lu, P.i. Biomedicine, Automatic Colonic Polyp Detection Using Integration of Modified Deep Residual Convolutional Neural Network and Ensemble Learning Approaches, Comput. Methods Programs Biomed. 206 (2021) 106114, doi: 10.1016/j.cmpb.2021.106114.
9. T. Shi, H. Jiang, B. Zheng, A Stacked Generalization U-shape network based on zoom strategy and its application in biomedical image segmentation, Comput. Methods Programs Biomed. 197(2020) 105678, doi: 10.1016/j.cmpb.2020.105678.
10. K.G. Lore, A. Akintayo, S. Sarkar, LLNet: A deep autoencoder approach to natural low-light image enhancement, Pattern Recognit. 61(2017) 650-662, doi: 10.1016/j.patcog.2016.06.008.
11. C. Li, J. Guo, F. Porikli, Y. Pang, LightenNet: A convolutional neural network for weakly illuminated image enhancement, Pattern Recognit. Lett. 104(2018) 15-22, doi: 10.1016/j.patrec.2018.01.010.

12. C. Wei, W. Wang, W. Yang, J. Liu, Deep retinex decomposition for low-light enhancement, in: British Machine Vision Conference, 2018, arXiv: 1808.04560.
13. L. Chen, W. Tang, N.W. John, T.R. Wan, J. Zhang, De-smokeGCN: Generative Cooperative Networks for Joint Surgical Smoke Detection and Removal, IEEE Trans. Medical Imaging 39(5)(2020) 1615-1625, doi: 10.1109/TMI.2019.2953717.
14. K.R. Prabhakar, V.S. Srikar, R.V. Babu, DeepFuse: A Deep Unsupervised Approach for Exposure Fusion with Extreme Exposure Image Pairs, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 4724-4732, doi: 10.1109/ICCV.2017.505.
15. 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 (BIBM), 2020, pp. 2403-2408, doi: 10.1109/BIBM49941.2020.9313443.
16. T. Hiroyasu, K. Hayashinuma, H. Ichikawa, N. Yagi, Preprocessing with image denoising and histogram equalization for endoscopy image analysis using texture analysis, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015, pp. 789-792, doi: 10.1109/EMBC.2015.7318480.
17. M.S. Imtiaz, S.K. Mohammed, F. Deeba, K.A. Wahid, Tri-Scan: A three stage color enhancement tool for endoscopic images, J. Medical Syst. 41(6)(2017) 102, doi: 10.1007/s10916-017-0738-z.
18. B. Sdiri, M. Kaaniche, F.A. Cheikh, A. Beghdadi, O.J. Elle, Efficient enhancement of stereo endoscopic images based on joint wavelet decomposition and binocular combination, IEEE Trans. Medical Imaging 38(1)(2018) 33-45, doi: 10.1109/TMI.2018.2853808.
19. X. Luo, A.J. McLeod, S.E. Pautler, C.M. Schlachta, T.M. Peters, Vision-based surgical field defogging, IEEE Trans. Medical Imaging 36(10)(2017)2021-2030, doi: 10.1109/TMI.2017.2701861.
20. R.C. Gonzalez, R.E. Woods, Digital image processing. in. Prentice hall Upper Saddle River, NJ, (2002).
21. X. Qiao, J. Bao, H. Zhang, L. Zeng, D. Li, Underwater image quality enhancement of sea cucumbers based on improved histogram equalization and wavelet transform, Information Processing in Agriculture 4(3)(2017) 206-213, doi: 10.1016/j.inpa.2017.06.001.
22. S.C. Huang, F.C. Cheng, Y.S. Chiu, Efficient Contrast Enhancement Using Adaptive Gamma Correction With Weighting Distribution, IEEE Trans. Image Process. 22(3)(2013) 1032-1041, doi:10.1109/TIP.2012.2226047.
23. Z. Al-Ameen, Nighttime image enhancement using a new illumination boost algorithm, IET Image Process. 13(8)(2019) 1314-1320, doi:10.1049/iet-ipr.2018.6585.
24. Z. 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, doi: 10.1109/ICIP.1996.560995.
25. D.J. Jobson, Z. Rahman, G.A. Woodell, A multiscale retinex for bridging the gap between color images and the human observation of scenes, IEEE Trans. Image Process. 6(7)(1997) 965-976, doi: 10.1109/83.597272.
26. D.J. Jobson, Z. Rahman, G.A. Woodell, Properties and performance of a center/surround retinex, IEEE Trans. Image Process. 6(3)(1997) 451-462, doi: 10.1109/83.557356.
27. A.B. Petro, C. Sbert, J.M. Morel, Multiscale retinex, Image Process. Line (2014) 71-88, doi: 10.5201/ipol.2014.107.
28. 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(6)(2018) 2828-2841, doi: 10.1109/TIP.2018.2810539.
29. F. Lv, F. Lu, J. Wu, C. Lim, MBLLEN: Low-Light Image/Video Enhancement Using CNNs, in: BMVC, 2018, pp. 1-13.
30. K. Fukushima, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biol Cybern 36(4)(1980) 193-202.
31. H. Ma, S. Ma, Y. Xu, M. Zhu, Low-light image enhancement based on deep convolutional neural network, Acta Optica Sinica 39(2)(2019) 99-108, doi: 10.3788/AOS201939.0210004.
32. I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Advances in Neural Information Processing Systems 3(11)(2014) 2672-2680, doi: 10.1145/3422622.
33. J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE international conference on computer vision (ICCV), 2017, pp. 2223-2232.
34. 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 (CVPR), 2020, pp. 1780-1789.
35. C. Li, C. Guo, C.L. Chen, Learning to enhance low-light image via zero-reference deep curve estimation, IEEE Trans. Pattern Anal. Mach. Intell. (2021), doi:10.1109/TPAMI.2021.3063604.
36. 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, doi: 10.1109/TIP.2021.3051462.
37. P. Singh, A neutrosophic-entropy based clustering algorithm (NEBCA) with HSV color system: A special application in segmentation of Parkinson's disease (PD) MR images, Comput. Methods Programs Biomed. 189(2020) 105317, doi: 10.1016/j.cmpb.2020.105317.
38. 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(4)(2004) 600-612, doi: 10.1109/TIP.2003.819861.
39. 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, 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, doi: 10.1145/3083187.3083212.
40. 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.
41. J. Bernal, F.J. Sanchez, G. Fernandez-Esparrach, D. Gil, C. Rodriguez, F.J. Vilarino, WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians, Comput. Medical Imaging Graph. 43(2015) 99-111, doi: 10.1016/j.compmedimag.2015.02.007.
42. J. Silva, A. Histace, O. Romain, X. Dray, B. Granado, surgery, Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer, Int. J. Comput. Assist. Radiol. Surg. 9(2)(2014) 283-293.
43. D. Vazquez, J. Bernal, F.J. Sanchez, G. Fernandez-Esparrach, A.M. Lopez, A. Romero, M. Drozdzal, A. Courville, A benchmark for endoluminal scene segmentation of colonoscopy images, Computer Vision in Healthcare Applications 2017(2017), doi: 10.1155/2017/4037190.
44. F.J. Sanchez, J. Bernal, C. Sanchez-Montes, C.R. de Miguel, G. Fernandez-Esparrach, Applications, Bright spot regions segmentation and classification for specular highlights detection in colonoscopy videos, Mach. Vis. Appl. 28(8)(2017) 917-936.
45. 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.
46. C. Lee, C. Lee, C. Kim, Contrast enhancement based on layered difference representation, in: 2012 19th IEEE International Conference on Image Processing, 2012, pp. 965-968, doi: 10.1109/ICIP.2012.6467022.