Pattern Recognition 133 (2023) 108968

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Retinal image enhancement with artifact reduction and structure retention

Bingyu Yang[a], He Zhao[a,*], Lvchen Cao[c], Hanruo Liu[b], Ningli Wang[b], Huiqi Li[a,*]

[a] Beijing Institute of Technology, Beijing, 100081, China
[b] Beijing Institute of Ophthalmology, Beijing Tongren Hospital, Capital Medical University, Beijing, 100730, China
[c] School of Artificial Intelligence, Henan University, Zhengzhou, 450046, China

ARTICLE INFO

Article history:
Received 21 April 2022
Revised 14 June 2022
Accepted 9 August 2022
Available online 10 August 2022

Keywords:
Retinal image enhancement
Generative adversarial networks
High frequency

ABSTRACT

Enhancement of low-quality retinal fundus images is beneficial to clinical diagnosis of ophthalmic diseases and computer-aided analysis. Enhancement accuracy is a challenge for image generation models, especially when there is no supervision by paired images. To reduce artifacts and retain structural consistency for accuracy improvement, we develop an unpaired image generation method for fundus image enhancement with the proposed high-frequency extractor and feature descriptor. Specifically, we summarize three causes of tiny vessel-like artifacts which always appear in other image generation methods. A high frequency prior is incorporated into our model to reduce artifacts by the proposed high-frequency extractor. In addition, the feature descriptor is trained alternately with the generator using segmentation datasets and generated image pairs to ensure the fidelity of the image structure. Pseudo-label loss is proposed to improve the performance of the feature descriptor. Experimental results show that the proposed method performs better than other methods both qualitatively and quantitatively. The enhancement can improve the performance of segmentation and classification in retinal images.

1. Introduction

In recent years a lot of disease detection and segmentation algorithms for fundus images have been proposed to assist clinical diagnosis and automatic retinal image analysis [1]. Both clinical disease screening and standard image methods require high-quality fundus images, while the high quality cannot always be guaranteed due to reasons like noisy image capture, sample/patient variability [2]. The unsatisfied quality of retinal images is usually shown as poor illuminance, low contrast, and blurriness, which makes it hard for ophthalmologists to distinguish different diseases so as to decrease the accuracy of diagnosis [3]. Meanwhile, poor-quality images sometimes lead to unsatisfied results in automatic image processing (e.g. segmentation, tracking), which may affect further disease analysis.

The blurriness of fundus image can be classified into four grades [4], which is shown in Fig. 1(a)–(d). The blurriness is caused by different degrees of cataracts. As cataracts worsen, the blurriness of the image increases and the visibility of the fundus structure decreases. It is difficult to accurately extract the fundus structure of blurred images, especially heavily blurred images. Fig. 1(e) is an underlit image and Fig. 1(f) is a light leak image where the edge part is brighter due to the imperfect imaging process. The structure in the excessive dark or bright area is not easily recognizable.

Many excellent retinal image enhancement methods have been proposed recently. For classical methods, complex algorithms are designed to adjust the contrast and brightness of blurred images according to the prior characteristics [5]. For DL methods, some methods first generate paired datasets, in which clear and blurry images correspond pixel by pixel, and then train the enhancement networks with the generated data [6]. The enhancement results usually depend on the generated paired data. Other methods use the cycle consistency [7] to loosen the data constraints. Enhancement accuracy is the main challenge of these methods because unpaired datasets have less supervision than paired datasets, especially for details like blood vessels.

For unpaired image translation methods, we find that tiny vessel-like artifacts always appear in some hard cases. In addition, we should also pay more attention to the anatomical structures in the blurry image, such as blood vessels. So we propose a fundus image enhancement method with artifact reduction and structure retention. Our method has an enhancement generator, a blur generator, a high frequency (HF) extractor, a feature descriptor (FD)

* Corresponding authors.
E-mail addresses: zhaohe@bit.edu.cn (H. Zhao), huiqili@bit.edu.cn (H. Li).

https://doi.org/10.1016/j.patcog.2022.108968
0031-3203/

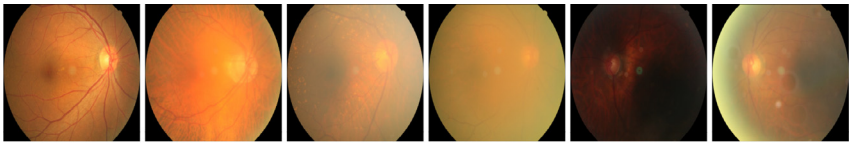

Fig. 1. Retinal images. (a) Clear image. (b) Slightly blurred image. (c) Moderately blurred image. (d) Severely blurred image. (e) Underlit image. (f) Edge light leak image.

and two discriminators. Our contributions can be mainly summarized in the following four aspects:

- We develop an unpaired fundus image enhancement method, which can effectively reduce artifacts and ensure structural consistency.
- We summarize three causes of artifacts — severe blurriness, imperfect illumination, and misleading information. High frequency prior is incorporated into our generative networks to reduce the artifacts by the proposed high-frequency extractor.
- A feature descriptor is trained alternately with the generator to ensure the fidelity of image structure. Pseudo-label loss is proposed to extract a better vessel feature in blurry images.
- Both visual comparison and quantitative evaluation prove the superiority of this method. And the enhancement can improve retinal image processing such as vessel segmentation, disease classification.

2. Related works

2.1. Classical retinal image enhancement

Classical fundus image enhancement methods usually use the prior information of the image to design complex algorithms manually. These methods usually focus on the improvement of contrast and the adjustment of brightness [8]. They can be divided into three categories.

Methods of Histogram Adjustment Contrast limited adaptive histogram equalization (CLAHE) [9] is wildly used in retinal image enhancement for contrast improvement. In [10], Setiawan et al. use CLAHE in the green channel to improve the quality of color retinal images. Zhou et al. [5] first leverage a luminance matrix obtained by gamma correction and then use CLAHE in the luminosity channel of L*a*b* [11] color space to adjust the luminosity and contrast. Gupta et al. [3] employ a quantile-based histogram equalization method after using adaptive gamma correction.

Methods Based on Image Formation Model Image formation model is usually described as:

$$\mathbf{I}(\mathbf{x}) = \mathbf{J}(\mathbf{x})t(\mathbf{x}) + \mathbf{A}(1 - t(\mathbf{x})), \tag{1}$$

where $\mathbf{x}$ is the input point, $\mathbf{I}$ is the observed intensity, $\mathbf{J}$ is the scene radiance, $\mathbf{A}$ is the global atmospheric light, and t is the medium transmission [12]. Xiong et al. [13] make use of the image formation model and estimate background illuminance and transmission map by extracting background and foreground. Then the blurry fundus image is divided into high-intensity areas and low-intensity areas to be processed separately. Gaudio et al. [14] re-interpret the distortion model for image dehazing. Based on dark channel prior [12], a family of brightening, darkening and sharpening methods are developed. These methods can be combined for retinal image enhancement. In [15], a double pass fundus reflection (DPFR) model for retinal image enhancement is proposed based on the image formation model. Retinex theory [16] and dark channel prior [12] are used to correct illumination and dehaze in this method.

Methods with Separated High and Low Frequencies The low-frequency component of a fundus image contains local brightness information, while the high-frequency component contains structural information such as blood vessels, lesions, optic cup and disk. Cao et al. [8] also use Gaussian filtering to remove the influence of base-intensity and then non-uniform addition is used to enhance the contrast. In [17], retina cortex theory [16] is employed to remove low frequency in the root domain. Then grayscale adjustment and refinement are used to further enhance the contrast and adjust the image color. The idea of removing low-frequency components can be used for reference because the structure and details of an image are more concentrated in high frequencies.

However, classical methods are usually proposed with multiple steps while each step often has a lot of artificially designed parameters. The design process is complicated and time-consuming, especially when considering different kinds of degradation like insufficient illuminance or overexposure. Besides, model design and model parameter selection are empirically set based on the data set, which may lead to data sensitivity. Furthermore, image enhancement is an ill-posed problem, since one input can produce several outputs. The selection of the final enhanced result depends on the designers experience and visual evaluation.

2.2. Deep-learning-based retinal image enhancement

Recently DL-based methods get more attention and they can be roughly divided into three categories according to the data sets.

Data Pre-generation Methods It is hard to obtain a sufficient amount of paired high-quality and low-quality fundus images in the real imaging process for model training. But image degradation is easier to simulate compared with image enhancement. Therefore, these methods degrade high-quality images to generate paired data. Luo et al. [18] propose CataractSimGAN to synthesize cataract-like images. They use these generated paired images to train the CataractDehazeNet for enhancement. Shen et al. [6] model the interference in terms of three factors. The high-quality images can be processed by the model with randomly perturbed variables to obtain their degraded counterparts. Based on the degraded data set, Cofe-Net is proposed to suppress the degradation factors. These methods are usually multi-step and the enhancement results depend on their degeneration algorithm.

Single-Shot Image Reconstruction Methods Qayyum et al. [19] put forward a single-shot deep image prior (DIP) based [20] approach which does not require any training data. They leverage the image formation model [12] to decompose the retinal image into the transmission map, atmospheric light and enhanced image via coupled DIPs. Dark channel prior [12] is incorporated when training the models. Blood vessels and other structures are enhanced in the image reconstruction.

Unpaired Image Translation Methods Unpaired image-to-image translation generally refers to image translation from the source domain to the target domain without paired data [7]. You et al. [21] propose a retinal image enhancement method called Cycle-CBAM, which adopts Convolutional Block Attention Module (CBAM) [22] to improve the baseline of CycleGAN [7]. They also prove that the enhanced results are beneficial to diabetic retinopathy classification. Zhao et al. [23] leverage cyclic consistency at the feature level with a dynamic feature descriptor for retinal image enhance-

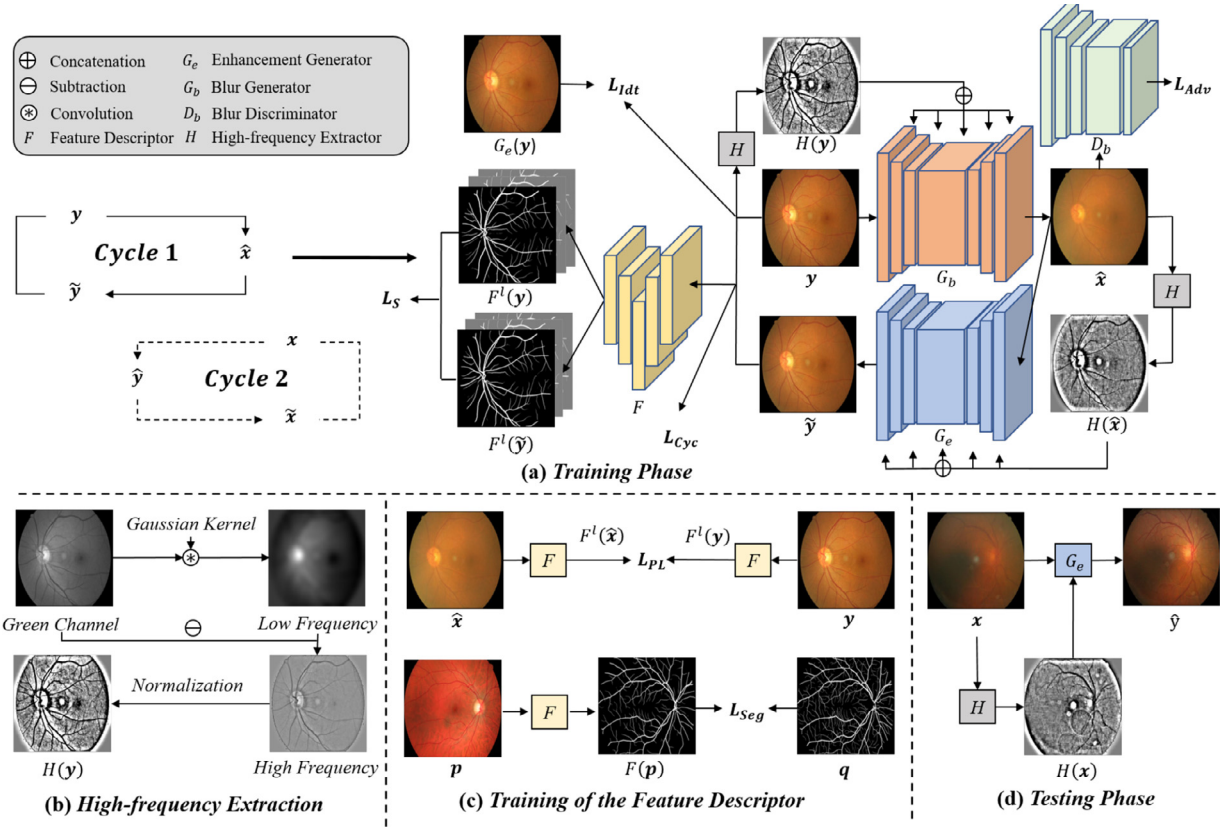

Fig. 2. Flowchart of our approach. There are two cycles in the training phase and their training process is the same. Only Cycle 1 is shown in (a). The high-frequency extractor H extracts the high frequency of the green channel as the side input for the generator to reduce artifacts. The feature descriptor F is trained alternately with generators and used to ensure the consistency of vessel structure. G_e is used for enhancement in the testing phase. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ment. This method reduces the artifacts in the generated images and the enhanced results are conducive to subsequent segmentation, tracking tasks. These methods are end-to-end and the enhancement process is learned from real data to avoid designers subjective choice.

However, due to the lack of pixel-by-pixel supervision of the paired data, the challenge of this method lies in the accuracy of the enhanced results. Small vessel-like artifacts sometimes accompany the image generation. In addition, extracting the structure of blurred images is a challenge for the generator. Our method is based on the unpaired image translation but we introduce the idea of high and low frequency separation in classical methods to reduce the artifacts. On the other hand, we develop a special feature descriptor with the pseudo-label loss to assist the training of the generator.

3. Methodology

3.1. Model structure

3.1.1. Overview

We regard fundus image enhancement as a special image-to-image translation problem and our baseline is CycleGAN [7]. The flowchart of our approach is displayed in Fig. 2. Denote low-quality retinal fundus images as $\{\boldsymbol{x}_n\}_{n=1}^{N} \in \mathbb{R}^{W\times H\times 3}$ and the unpaired high-quality images as $\{\boldsymbol{y}_m\}_{m=1}^{M} \in \mathbb{R}^{W\times H\times 3}$. Our goal is to enhance the low-quality image $\boldsymbol{x}$ to generate the corresponding enhanced image $\hat{\boldsymbol{y}}$ as clear as the high-quality image $\boldsymbol{y}$. At the same time, $\hat{\boldsymbol{y}}$ should keep the content consistent with $\boldsymbol{x}$. Our method has one high-frequency extractor, one feature descriptor, two generators, and two discriminators. The high-frequency extractor H is used to extract the high frequency of color retinal images. The two generators have the same model structure. The enhancement generator $G_e : (\boldsymbol{x}, H(\boldsymbol{x})) \rightarrow \hat{\boldsymbol{y}}$ represents enhancing a blurry image into a clear image, where $H(\boldsymbol{x})$ is sent into G_e from the side to assist enhancement. The blur generator $G_b : (\boldsymbol{y}, H(\boldsymbol{y})) \rightarrow \hat{\boldsymbol{x}}$ denotes blurring the input clear image. The two discriminators are denoted as D_e and D_b respectively. D_e is used to distinguish real high-quality images from images generated by G_e, and D_b is used to distinguish between the real blurred images and the images generated by G_b. The feature descriptor F is employed to extract structural information of real fundus images $\boldsymbol{x}$, $\boldsymbol{y}$ and reconstructed images $\tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{y}}$. After training, we use G_e to enhance the low-quality images with the help of the high-frequency extractor. The model structures will be discussed in detail in the following sections.

3.1.2. High-frequency extractor

The process of high-frequency extraction is shown in the left bottom of Fig. 2. The low frequency of the green channel is separated by the Gaussian kernel. High frequency is obtained by subtraction. We normalize the high frequency to correct the extreme value. High-frequency images are concatenated with input images and feature maps of the generator to help the image generation. The removal of low-frequency components means that the effects of uneven lighting are eliminated. The stretching of contrast amplifies structural information, and the selection of the green channel reduces misleading information.

We find that artifacts sometimes accompany the enhancement when we utilize generative adversarial networks without the constraint of paired data such as CycleGAN [7]. The artifacts usually emerge as small blood vessels in the enhancement results, which

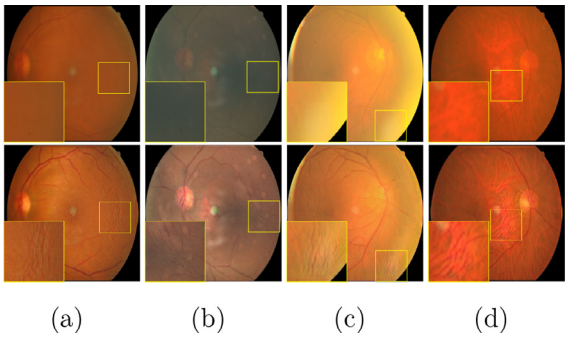

Fig. 3. Images with artifacts. The first row shows the low-quality image and the second row is the enhanced results by CycleGAN [7]. (a) Heavily blurred image. (b) Underlit image. (c) Light leak image. (d) Choroidal misleading image.

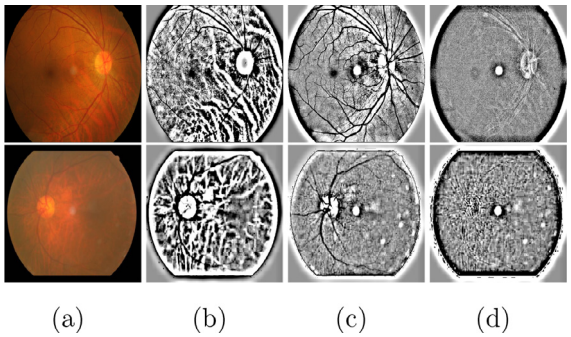

Fig. 4. High frequency of RGB. The first row is a clear image and the second row is a blurred image. (a) Original image. (b) HF of the red channel. (c) HF of the green channel. (d) HF of the blue channel. The green channel has the clearest structural information. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

actually does not exist in the real blur images. We summarize several scenarios where artifacts easily appear. Examples are shown in Fig. 3. First, heavily blurry images are prone to have artifacts after enhancement. The contrast in heavily blurry areas is too low to identify the fundus structure. The generator easily treats plausible perturbations as blood vessels and thus produces artifacts. Second, images with imperfect illumination such as insufficient light or light leak are more likely to produce artifacts. These areas are usually highly blurry so that the fundus structure is very difficult to recognize. In addition, it needs two steps for the enhancement of images with imperfect illumination – brightness adjustment and contrast stretching, which is more complex than just contrast stretch in other low-quality images. Third, some textures similar to retinal vessels, such as choroidal vessels, may mislead the generator to produce artifacts.

In order to reduce the artifacts, uneven illumination and misleading information should be eliminated and fundus structure information needs to be increased. In fundus images, the uneven light intensity is in the low-frequency component, and the fundus structure information is always in the high-frequency component. Inspired by retina cortex theory [16] and frequency separation methods [8], we develop a high-frequency extractor to reduce the influence of illumination and amplify the fundus structure. A two-dimensional Gaussian kernel is employed to extract the low-frequency

$$Gauss(p,q) = \frac{1}{2\pi\sigma^2} e^{-\frac{(p-\mu)^2+(q-\mu)^2}{2\sigma^2}}, \tag{2}$$

where p and q are the coordinates of pixels, μ denotes the central point and σ determines the falling gradient of the gaussian kernel. The size of the Gaussian kernel is about $\frac{1}{6}$ of the retinal image in order to cover the optic disc. σ is set as the kernel size divided by π.

High-frequency extraction can be summarized as the following three steps. First, the input image is convolved with a Gaussian kernel to get the low-frequency image. We pad the mask with the mean value to reduce its impact on the boundary region before convolution. Second, low frequency is subtracted from the input image to get the high frequency. Third, the high frequency is normalized with the threshold T and some extremums are abandoned using the truncation function f. Extreme values that deviated from the histogram are corrected and the contrast of the high-frequency image is stretched during the normalization. Denote the input image as $\boldsymbol{x}$, its high-frequency image $H(\boldsymbol{x})$ can be expressed as:

$$H(\boldsymbol{x}) = f\left(\frac{(\boldsymbol{x} - \boldsymbol{x} * Gauss) + T}{2T}\right), \tag{3}$$

where *Gauss* means the pre-defined Gaussian kernel and $*$ denotes convolution. f is the truncation function which is defined as:

$$f(a) = \begin{cases} 0, (a < 0) \\ a, (0 \le a < 1) \\ 1, (1 \le a) \end{cases} \tag{4}$$

We select the green channel because it has the best contrast, the least noise, and the richest retinal structures in retinal images [24]. High-frequency images of the three channels are shown in Fig. 4. In both clear and blurred images, the high frequency of the red channel has a lot of choroidal texture, which will interfere with the enhancement process. The blue channel has the lowest contrast, unobvious structural information, and large noise. The high frequency of the green channel has the most obvious blood vessels and the highest image quality in the three channels. It can provide a lot of useful structure information in blurred images, so we only use the green channel to assist the image enhancement.

3.1.3. Feature descriptor

In order to retain the consistency of structural information, blood vessels should be paid more attention to during enhancement. A specialized feature descriptor F is developed to assist image generation. Commonly used feature extractors are usually trained on clear images. They are not very good at extracting information from blurred images, because the contrast of blurred images is low and the gradient information is not obvious. In our approach, clear and blurred image pairs are generated during the training of the generator (e.g. $\boldsymbol{y}$ and $\hat{\boldsymbol{x}}$). The feature descriptor is dynamically trained by the retinal vessel segmentation dataset and generated image pairs. We send clear images to the feature descriptor to get pseudo labels of the paired blurred images. Then the feature descriptor is used to extract structure information of real images $\boldsymbol{x}$, $\boldsymbol{y}$ and reconstructed images $\tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{y}}$. The training process of the feature descriptor is intuitively similar to generative adversarial nets, where the generator and the feature descriptor update the parameters alternately. Compared with unsupervised enhancement tasks, segmentation is much easier to optimize with the high-quality retinal images and corresponding ground truths. Furthermore, the feature descriptor F is also trained with paired clear and blurred images and gradually improves the performance on blurred images.

U-Net [25] structure is employed in our feature descriptor. The model has 4 down-sampling, 4 up-sampling layers, and a convolutional layer with stride 1 at the outermost to expand the channel to 64. The number of channels is doubled every time it is down-sampled. Corresponding feature maps are concatenated before being sent to the up-sampling layers.

3.1.4. Generator and discriminator

There are two generators in our method. G_e is designed to enhance blurred images, while G_b is used to blur clear images. Retinal images and their high-frequency images of the green channel are concatenated together to work as input to the generator. Besides, high-frequency images are downsampled and sent to concatenate with feature maps of different layers in the generator. They will be concatenated every time the feature maps are downsampled because receptive fields differ in the feature maps of different sizes. Normalization and activation are performed before concatenation. Then the two cycles in our approach can be expressed as:

$$G_e(\boldsymbol{x}, H(\boldsymbol{x})) = \hat{\boldsymbol{y}}, G_b(\hat{\boldsymbol{y}}, H(\hat{\boldsymbol{y}})) = \tilde{\boldsymbol{x}}, \tag{5}$$

$$G_b(\boldsymbol{y}, H(\boldsymbol{y})) = \hat{\boldsymbol{x}}, G_e(\hat{\boldsymbol{x}}, H(\hat{\boldsymbol{x}})) = \tilde{\boldsymbol{y}}, \tag{6}$$

where $\tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{y}}$ denote the reconstructed images from $\boldsymbol{x}$ and $\boldsymbol{y}$.

The discriminator D_e is employed to distinguish the enhanced images $\hat{\boldsymbol{y}}$ from the real high-quality images $\boldsymbol{y}$, while D_b is used to distinguish between the real blurred images and the images generated by G_b. The function of the discriminator is defined as $D : \boldsymbol{X} \rightarrow d \in [0, 1]$. When the input is a real image, d approaches 1, and when the input image is from the generator, d is close to 0.

Both generators have the same model structure. They consist of 3 convolutional layers, a residual bottleneck module, and 3 corresponding deconvolutional layers. The bottleneck module has 9 residual blocks, in which two convolutional layers and one skip connection are utilized. The deconvolutional layers have the same kernel sizes and stride as the corresponding convolutional layers. The two discriminators also have the same structure. It consists of 5 convolutional layers and the input is downsampled four times.

3.2. Objective function

Three kinds of neural networks need to be trained in our method: generator, discriminator, and feature descriptor. When one of the networks is trained, the parameters of the other two models are not updated. In each iteration, we first update the feature descriptor, then the generator, and finally the discriminator. We minimize the losses to optimize network parameters.

3.2.1. Generator

Adversarial loss $\mathcal{L}_{Adv}^{G}$ can constrain the network to learn the mapping from blur to clear or from clear to blur. We use the least square loss to train the generator because of its stability.

$$\mathcal{L}_{Adv}^{G} = \left(D_e(\hat{\boldsymbol{y}}) - 1\right)^2 + \left(D_b(\hat{\boldsymbol{x}}) - 1\right)^2. \tag{7}$$

Cycle consistency loss can reduce the space of possible mapping and keep the content consistent during the training. $\mathcal{L}_1$ loss is calculated between real images $\boldsymbol{x}$, $\boldsymbol{y}$ and reconstructed images $\tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{y}}$. The cycle consistency loss $\mathcal{L}_{Cyc}$ at image level can be expressed as:

$$\mathcal{L}_{Cyc} = \|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_1 + \|\tilde{\boldsymbol{y}} - \boldsymbol{y}\|_1, \tag{8}$$

where $\|\cdot\|_1$ is the absolute-value norm.

Identity loss $\mathcal{L}_{Idt}$ is used to help preserve the image color.

$$\mathcal{L}_{Idt} = \|G_e(\boldsymbol{y}, H(\boldsymbol{y})) - \boldsymbol{y}\|_1 + \|G_b(\boldsymbol{x}, H(\boldsymbol{x})) - \boldsymbol{x}\|_1. \tag{9}$$

Structure consistency loss $\mathcal{L}_S$ uses the feature descriptor F to extract feature maps of real images and reconstructed images. It allows the generator to pay more attention to fundus structure information. $F^l(\cdot)$ refers to extracting the feature maps of the lth layer and λ_l is the weight. If l is the outermost layer of the feature descriptor, the output is the segmentation map.

$$\mathcal{L}_S = \sum_l \lambda_l \left(\left\| F^l(\tilde{\boldsymbol{x}}) - F^l(\boldsymbol{x}) \right\|_1 + \left\| F^l(\tilde{\boldsymbol{y}}) - F^l(\boldsymbol{y}) \right\|_1 \right). \tag{10}$$

Full objective can be summarized as follows:

$$\mathcal{L}_{Total}^{G} = \mathcal{L}_{Adv}^{G} + \lambda_{Cyc}\mathcal{L}_{Cyc} + \lambda_{Idt}\mathcal{L}_{Idt} + \lambda_S\mathcal{L}_S. \tag{11}$$

3.2.2. Discriminator

Adversarial loss is utilized to train the discriminators. We average the output of the discriminator to get the final prediction. The output of a real image tends to be 1, and a fake image tends to be 0.

$$\mathcal{L}_{Adv}^{D} = \left(D_e(\hat{\boldsymbol{y}})\right)^2 + (D_e(\boldsymbol{y}) - 1)^2 + \left(D_b(\hat{\boldsymbol{x}})\right)^2 + (D_b(\boldsymbol{x}) - 1)^2. \tag{12}$$

3.2.3. Feature descriptor

Segmentation loss $\mathcal{L}_{Seg}$ is used to train the feature descriptor to capture the structure information of retinal images. Denote vessel segmentation dataset as $\{\boldsymbol{p}_k, \boldsymbol{q}_k\}_{k=1}^{K}$, where $\boldsymbol{p}_k \in \mathbb{R}^{W\times H\times 3}$ is the color retinal image and $\boldsymbol{q}_k \in \mathbb{R}^{W\times H\times 1}$ is the corresponding segmentation map.

$$\mathcal{L}_{Seg} = \|F(\boldsymbol{p}) - \boldsymbol{q}\|_1. \tag{13}$$

Pseudo label loss $\mathcal{L}_{PL}$ can improve the capability of the feature descriptor when the input is blurry images. We make use of the image pairs produced during the training of the generator. The segmentation result of the clear image is used as the pseudo label of the blurred image to train the F. $\bar{F}$ means only forward propagation and parameters are not updated.

$$\mathcal{L}_{PL} = \sum_l \lambda_l \left(\left\| F^l(\boldsymbol{x}) - \bar{F}^l(\hat{\boldsymbol{y}}) \right\|_1 + \left\| F^l(\hat{\boldsymbol{x}}) - \bar{F}^l(\boldsymbol{y}) \right\|_1 \right). \tag{14}$$

The total loss of feature descriptor is expressed as:

$$\mathcal{L}_{Total}^{F} = \mathcal{L}_{Seg} + \lambda_{PL}\mathcal{L}_{PL}. \tag{15}$$

4. Experiments and results

4.1. Datasets and implementation details

We experiment on public and private datasets. For experiments with public datasets, Eye-Quality (EyeQ) Assessment Dataset [26] annotates the quality of retinal images, where “good”, “usable” and “reject” labels correspond to image qualities ranging from good to poor. *EyeQ training set:* We randomly select 600 images marked as “usable” or “reject” as low-quality images, and 600 images with good annotations as high-quality images. *Segmentation training set:* In order to train the feature descriptor, 140 fundus vessel segmentation images are selected from CHASEDB1 [27], DRIVE [28], DRHAGIS [29], HRF [30], and IOSTAR [31] datasets. *EyeQ test set:* 200 images from Eye-Quality Assessment Dataset [26] with “usable” or “reject” labels are employed. All images are resized to 512 × 512 before being sent to the model.

For experiments with private datasets, images collected clinically from Anzhen Hospital and Tongren Hospital are applied. *Clinical training set:* We employ 550 low-quality and 550 high-quality fundus images to train our networks. The image quality (high quality or low quality) of the dataset is labeled by ophthalmologists. The segmentation training set is also used for training the feature descriptor. There are three private test sets in our experiments. *Clinical test set:* We use 50 clinically collected low-quality fundus images from the same source as the clinical training set for no-reference assessment. *Cataract surgery test set:* Sixteen pairs of images before and after cataract surgery are tested for full-reference assessment. Image after surgery is the high-quality ground truth of the image before surgery. Angle and position offsets between image pairs will affect the accuracy of the evaluation. So these image pairs are registered using PIIFD [32], and only the overlapping

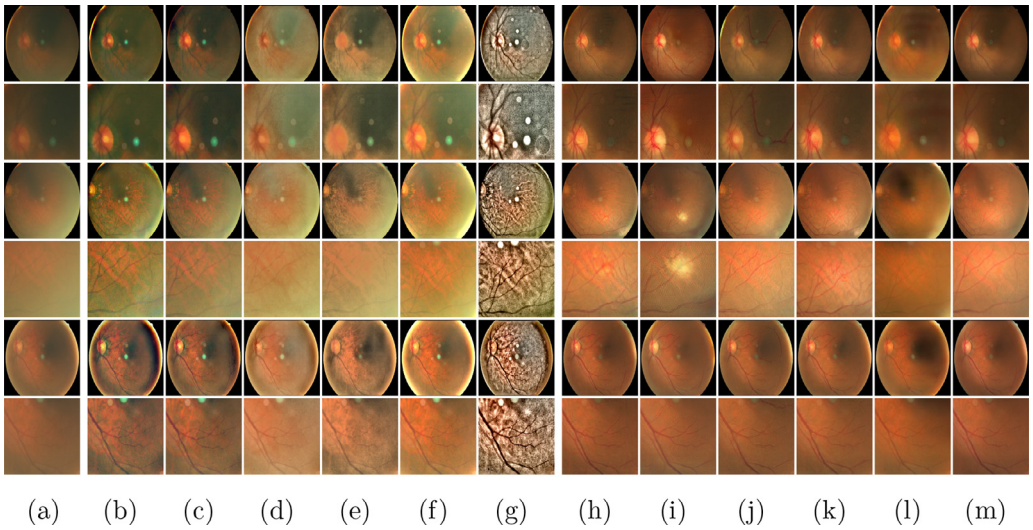

Fig. 5. Visual comparison with state-of-the-art methods. (a) Low-quality image. (b) Cao and Li [8] (c) Xiong et al. [13] (d) Zhou et al. [5] (e) Gupta and Tiwari [3] (f) Gaudio et al. [14] (g) Zhang et al. [15] (h) CycleGAN. [7] (i) CUT. [33] (j) You et al. [21] (k) Zhao et al. [23] (l) Cofe-Net [6] (m) Ours. (b)–(g) are classical methods and (h)–(m) are deep learning methods.

area is counted during the assessment. *Degradation test set:* Another data set for full-reference assessment is the images got from the fundus image degradation algorithm in [6]. We sent 100 clear images to the degradation model to get the corresponding blur images.

Similarly as CycleGAN [7], we set $\lambda_{Cyc} = 10$, $\lambda_{Idt} = 5$ in the training phase. The first layer and four down-sampling layers of the feature descriptor are selected to calculate $\mathcal{L}_S$ and $\mathcal{L}_{PL}$, while λ_l is set as $[\frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1]$ following the settings in [18]. λ_S is set as 0.5 and λ_{PL} is set as 0.025 according to the ablation study. There are hundreds of feature layers involved in the calculation of $\mathcal{L}_S$ and $\mathcal{L}_{PL}$, while only several channels for other losses like $\mathcal{L}_{Cyc}$. The small value of λ_S and λ_{PL} can make these losses in the same order of magnitude. The networks are trained for 200 epochs.

We adopt no-reference and full-reference image quality assessment methods to evaluate the enhancement results. For no-reference assessment, BRISQUE [34] and entropy are adopted. We retrain the classifier of BRISQUE with fundus images to better evaluate the image quality for the specific application. The training set contains 1400 fundus images with different qualities labeled by ophthalmologists. PSNR and SSIM are utilized for full-reference assessment. Lower values of BRISQUE indicate better image quality, while higher entropy, PSNR and SSIM scores mean better enhancement results.

4.2. Comparison with other methods

In this section, we compare our approach with other state-of-the-art color retinal enhancement methods. These methods can be divided into two categories: classical methods [3,5,8,13–15] and DL-based methods [6,7,21,23,33]. Methods [7,21,23,33] are retrained using their original codes with parameter refinement to achieve the best performance while Cofe-Net [6] is tested with the pre-trained weights provided by their authors. In addition, the feature descriptor in [23] is trained with the segmentation training set for a fair comparison. Qualitative and quantitative results are provided to illustrate the advantages of our approach.

The visual comparison is shown in Fig. 5. For classical methods, the disadvantage lies in the adjustment of colors, such as the overall greenish result for the light leak image in the third row in Fig. 5. Another disadvantage is that their enhancement results generally have a lot of noise. The advantage of the classical methods is the accuracy for enhancement. For example, the method in Fig. 5(g) is very effective for the enhancement of blood vessels. However, this method has serious color distortion compared with other methods. The difficulty of the deep learning methods is to ensure the accuracy of the enhancement results. For example, images enhanced by other DL-based methods have various degrees of artifacts that dont exist in the real image. Our method combines the high-frequency information commonly used in classical methods. So the enhancement results are accurate without artifacts and the vessel structure is more obvious than other DL-based methods. Besides, the color restoration is better and noise is significantly reduced compared with classical methods.

We further provide the quantitative evaluation compared with other state-of-the-art approaches, which is shown in Table 1. Our method performs the best on BRISQUE in the no-reference assessment on both EyeQ test set and clinical test set. Some classical methods have higher entropy scores than ours because they can easily stretch the contrast. But our method performs best in deep learning methods. In the full-reference evaluation, classical methods perform not as well as DL-based methods on the whole. Our method acquires the highest scores on the cataract surgery test set and the second-highest scores on the degradation test set. Cofe-Net [6] performs best on the degradation test set because it is trained with the data degraded by the same algorithm.

4.3. Ablation studies

To prove the effectiveness of our approach, we add the high-frequency extractor and feature descriptor to the baseline model CycleGAN [7] respectively. Quantitative evaluation is performed on the full-reference cataract surgery dataset because we are more concerned about real low-quality fundus images and full-reference scores are more accurate for these experiments.

4.3.1. High-frequency extractor for artifact reduction

As shown in Table 2, high-frequency information is helpful for the retinal image enhancement task. We add the high-frequency images into different feature layers of the generator to explore the best way for concatenation. Features of the generator are down-sampled twice in total before being sent to the residual blocks. Layer 0 means the high-frequency images are concatenated with the input images and Layer 1 and 2 indicate the combination with

Table 1
Quantitative evaluation compared with other state-of-the-art methods.

		No-reference				Full-reference			
		EyeQ		Clinical		Surgery		Degradation	
		BRISQUE	Entropy	BRISQUE	Entropy	PSNR	SSIM	PSNR	SSIM
Classical Method	Cao et al. [8]	52.09	6.271	54.29	6.283	17.91	0.8451	17.33	0.8163
	Xiong et al. [13]	49.12	6.404	52.64	6.669	19.03	0.8773	16.55	0.8094
	Zhou et al. [5]	51.43	6.630	49.23	6.740	15.46	0.8293	16.53	0.8672
	Gupta and Tiwari [3]	52.43	6.676	51.37	6.768	17.07	0.8561	17.27	0.8553
	Gaudio et al. [14]	54.15	**6.747**	56.73	6.821	14.92	0.8041	15.15	0.8469
	Zhang et al. [15]	46.92	6.170	54.59	**6.846**	12.07	0.6664	11.25	0.6754
DL- Based Method	CycleGAN [7]	46.96	6.640	54.21	6.672	19.19	0.8896	18.69	0.8897
	CUT [33]	52.29	6.659	53.63	6.653	19.03	0.8773	18.65	0.8896
	You et al. [21]	45.16	6.669	49.46	6.661	18.95	0.8853	18.98	0.8928
	Zhao et al. [23]	45.28	6.621	53.73	6.680	19.19	0.8879	19.02	0.8937
	Cofe-Net [6]	55.87	6.672	55.02	6.653	18.58	0.8880	**20.72**	**0.9218**
	Ours	**43.75**	6.695	**46.63**	6.710	**20.01**	**0.9012**	20.18	0.9006

Table 2
Ablation study on the cataract surgery test set.

	Model component		PSNR	SSIM
CycleGAN	–		19.19	0.8896
CycleGAN	Layer 0		19.65	0.8929
+	Layer 0, 1		19.76	0.8949
HF	Layer 0, 1, 2		**19.86**	**0.8966**
	All Layers		19.67	0.8934
CycleGAN	$\lambda_S = 0.5$,	$\lambda_{PL} = 0.0125$	19.59	0.8924
+	$\lambda_S = 0.5$,	$\lambda_{PL} = 0.025$	**19.70**	**0.8975**
FD	$\lambda_S = 0.5$,	$\lambda_{PL} = 0.125$	19.55	0.8961
	$\lambda_S = 0.25$,	$\lambda_{PL} = 0.025$	19.57	0.8967
	$\lambda_S = 2.5$,	$\lambda_{PL} = 0.025$	19.69	0.8974
Ours	Layer 0, 1, 2			
			20.01	**0.9012**
(CycleGAN + HF + FD)	$\lambda_S = 0.5$,	$\lambda_{PL} = 0.025$		

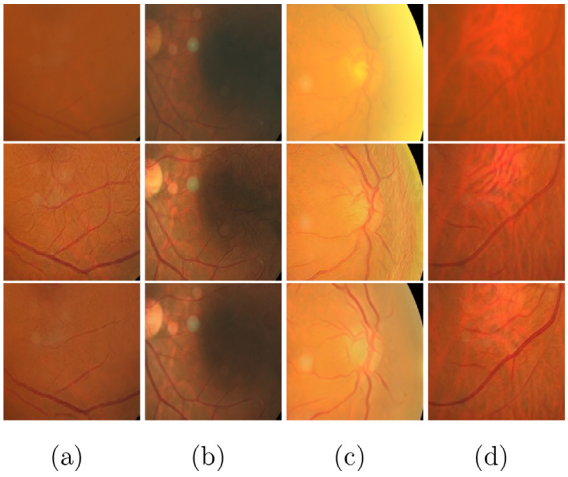

Fig. 6. Effectiveness of HF extractor. From top to bottom are low-quality images, image enhanced by CycleGAN and CycleGAN + HF. (a) Heavily blurred image. (b) Underlit image. (c) Light leak image. (d) Choroidal misleading image.

features downsampled once and twice. The best result is that high-frequency images are concatenated with the input image and features after each downsampling. If we add them to all feature maps in the generator, the scores drop instead. This is because some high-frequency images are too close to the output and the remaining convolutional layers cannot handle the high-frequency images well.

Fig. 6 demonstrates the effect of the high frequency on the removal of artifacts. There are some flocculent artifacts in the enhanced result of CycleGAN in Fig. 6(a) because of the severe blur.

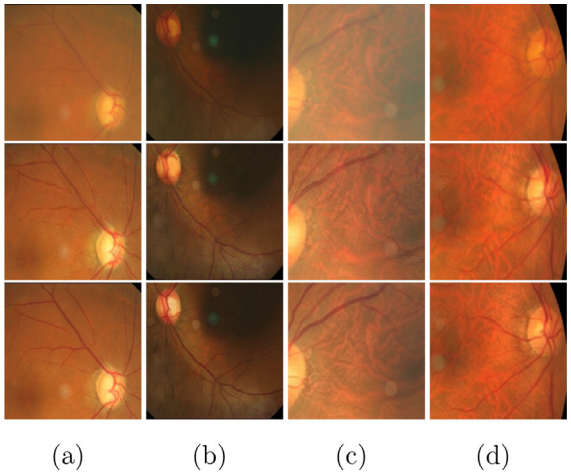

Fig. 7. Effectiveness of feature descriptor. From top to bottom are low-quality images, image enhanced by CycleGAN and CycleGAN + FD. (a)-(d) are different fundus images.

After adding the high-frequency information of the green channel, the artifacts disappear and the image becomes smoother. In Fig. 6(b), the result of CycleGAN is very noisy in the dark area and some blood vessels that do not exist are generated. In Fig. 6(c), there is some red texture in the light leak area. High frequency reduces the impact of uneven illumination, so these two artifacts also disappear in our method. Fig. 6(d) shows the artifacts caused by the choroid. Our method reduces artifacts and the blood vessel is more obvious than the baseline method at the same time.

4.3.2. Feature descriptor for structure retention

We select the best weights $\lambda_S = 0.5$, $\lambda_{PL} = 0.025$ in the training phase, which is shown in Table 2. Adding the feature descriptor can improve the quality of generated blood vessels, which is shown in Fig. 7. After adding the feature descriptor, the contrast of the main blood vessels in Fig. 7(a) becomes larger compared with the baseline method. In Fig. 7(b), some small blood vessels and image details are improved after adding the feature descriptor. In Fig. 7(c) and (d), our approach can recognize and enhance some small vessels ignored by the baseline model. Figure 8 displays the segmentation results of the feature descriptor with and without pseudo label loss. Pseudo label loss plays an important role in blurred image feature extraction. Many blood vessels cannot be segmented without pseudo label loss. As shown in Fig. 8(c), the segmentation result is greatly improved, where blood vessels are more accurate and continuous.

Table 3
Computational burden analysis on deep learning methods. G, D, F denote generators, discriminators and feature descriptors.

	Modules	Parameters (M)	Training Time /Epoch(s)	Inference Time /Image (s)
CycleGAN [7]	2G 2D	28.29	308	0.1483
CUT [33]	1G 1D 1F	14.70	478	0.3470
You et al. [21]	2G 2D	14.93	293	0.1491
Zhao et al. [23]	2G 2D 1F	45.55	354	0.1451
Cofe-Net [6]	3 Branches	41.22	–	0.2528
CycleGAN+HF	2G 2D	28.29	344	0.1498
CycleGAN+FD	2G 2D 1F	45.55	381	0.1486
Ours	2G 2D 1F	45.55	432	0.1529

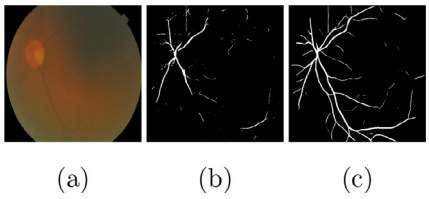

Fig. 8. Effectiveness of the pseudo label loss. (a) Low-quality image. (b) Segmentation result by the feature descriptor without pseudo label loss. (c) Segmentation result with pseudo label loss.

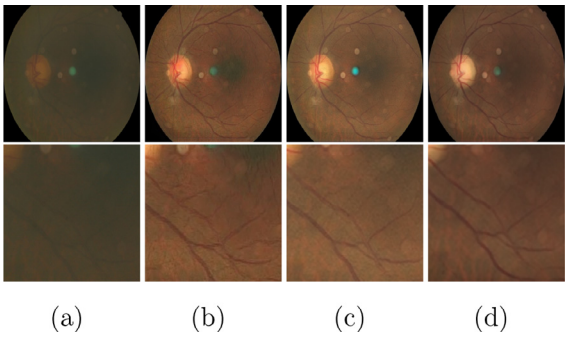

Fig. 9. Visual comparison for ablation study. The second row is the enlarged image of the first row. (a) Low-quality image. (b) Image enhanced by CycleGAN. (c) CycleGAN + FD. (d) Ours (CycleGAN + FD + HF).

4.3.3. Qualitative and quantitative analysis

Visual comparison for ablation study is shown in Fig. 9. On the whole, the enhancement result of the baseline method is noisy, and there are many flocculent artifacts around the heavily blurred and dark area. After adding the feature descriptor, the blood vessels are much clearer and the artifacts are reduced. After adding the high-frequency extractor, the contrast of blood vessels gets bigger. The image is smoother overall and artifacts are eliminated. Moreover, by adding FD and HF extractor, the quality of blood vessels is gradually improved. The color of blood vessels is darkened, and the graininess of the image is reduced, which makes the image more realistic. The quantitative evaluation of the surgery dataset is shown in Table 2. After adding these two modules, PSNR and SSIM are gradually improved. In addition, our method (CycleGAN + FD + HF) has a statistical improvement from CycleGAN with the p-value $0.034 < 0.05$.

4.4. Computational burden analysis

For classical methods, no training process is required. It takes a few seconds for the enhancement of each image. For example, it requires 8.661 s per image on average for Xiong et al. [13] and 4.208 s for Zhang et al. [15]. For the deep-learning methods, the number of parameters, training time and inference time are summarized in Table 3. The training time increases slightly after adding HF to CycleGAN [7], while both training time and the number of parameters increase after using the feature descriptor. After training, the inference efficiency of the DL-based methods is similar and faster than classical methods. The inference time of our method doesn't increase significantly compared to CycleGAN.

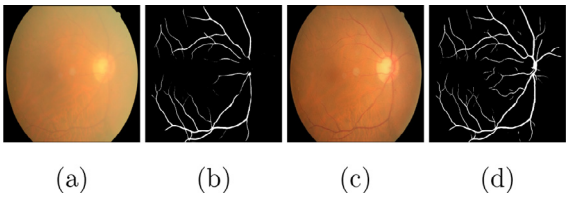

Fig. 10. Application on vessel segmentation. (a) Low-quality image. (b) Segmentation result of the low-quality image. (c) Enhancement result. (d) Segmentation result of the enhanced image.

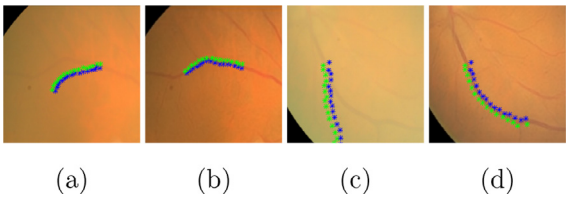

Fig. 11. Application on vessel tracking. (b) and (d) are the enhancement results of (a) and (c). The tracking trajectory is corrected after enhancement.

4.5. Applications

Our enhancement method can be used as the preprocessing of blurred retinal images for vessel segmentation and tracking. We use the segmentation training set to train a U-Net [25] for vessel segmentation. The results are shown in Fig. 10. Before enhancement, many blood vessel structures are blurry, so the segmentation results are not very satisfactory. Our method can improve the visibility of blood vessels so the segmentation results are significantly improved. We also test the enhancement method with a blood vessel tracking algorithm [35]. As shown in Fig. 11, our enhancement method corrects the tracking curve.

Our approach is also helpful for automatic disease diagnosis. The retinal fundus multi-disease image dataset (RFMiD) [36] is used to train an Efficient-Net [37] for automatic detection. There are 317 fundus images with the label of media haze in this dataset. And in these media haze images, there are 137 images with other diseases. The haze may hinder disease detection. We classify whether the images have other diseases except for media haze. The diagnostic accuracy is improved after enhancement. The results are shown in Table 4. For images without the label of media haze, there are 401 healthy images and 1202 diseased images. Some images also have slight blur or light leak. After enhancement, the slight turbidity is eliminated and pathological structures become more obvious. Therefore, the detection accuracy gets better. Our enhancement method not only improves the detection accuracy of

Table 4
Accuracy of automatic disease diagnosis. Image with media haze is blurred. Image without meadia haze is clear or slightly blurred.

Dataset	Media Haze	w/o Media Haze
Original Image	0.8095	0.9281
Enhanced Image	**0.8571**	**0.9468**

blurry images but also benefits clear images or slightly blurred images.

5. Conclusion

An end-to-end fundus image enhancement method is proposed in this paper. We use cycle-consistency constraints in the feature and image level with unpaired training images. In addition, when deep learning methods deal with difficult images, such as severely blurred or dark images, it is easy to produce small vessel-like artifacts. The proposed high frequency extractor can effectively reduce the artifacts. The feature descriptor trained with pseudo labels can also improve the accuracy of enhancement.

This method has better color and less noise compared with traditional methods. And it is more accurate and does not require paired data compared with deep learning methods. This method can improve the results of blood vessel segmentation and tracking. It is also helpful for automatic disease detection tasks and can improve the accuracy of classification. So it can be used as a preprocessing of these computer-aided algorithms. As the proposed algorithm can enhance retinal images with fewer artifacts, it can be employed in clinics to facilitate the diagnosis of ocular diseases. It can also be combined with retinal imaging devices to obtain retinal images with better quality.

Our current work pays more attention to blood vessels. In clinical diagnosis, ophthalmologists care more about specific areas, such as disease or optic disc besides blood vessels. A very important task in our future work is to enhance the feature of diseases to make it easier for ophthalmologists to distinguish. It is also a challenge because different diseases vary in color, shape, and size.

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.

References

[1] R. Zhao, Q. Li, J. Wu, J. You, A nested U-shape network with multi-scale upsample attention for robust retinal vascular segmentation, Pattern Recognit. (2021) 107998.
[2] Ç. Sazak, C.J. Nelson, B. Obara, The multiscale bowler-hat transform for blood vessel enhancement in retinal images, Pattern Recognit. 88 (2019) 739–750.
[3] B. Gupta, M. Tiwari, Color retinal image enhancement using luminosity and quantile based contrast enhancement, Multidimens. Syst. Signal Process. 30 (4) (2019) 1829–1837.
[4] L. Cao, H. Li, Y. Zhang, L. Zhang, L. Xu, Hierarchical method for cataract grading based on retinal images using improved haar wavelet, Inf. Fusion 53 (2020) 196–208.
[5] M. Zhou, K. Jin, S. Wang, J. Ye, D. Qian, Color retinal image enhancement based on luminosity and contrast adjustment, IEEE Trans. Biomed. Eng. 65 (3) (2017) 521–527.
[6] Z. Shen, H. Fu, J. Shen, L. Shao, Modeling and enhancing low-quality retinal fundus images, IEEE Trans. Med. Imaging 40 (3) (2020) 996–1006.
[7] 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, 2017, pp. 2223–2232.
[8] L. Cao, H. Li, Enhancement of blurry retinal image based on non-uniform contrast stretching and intensity transfer, Med. Biol. Eng. Comput. 58 (3) (2020) 483–496.
[9] A.M. Reza, Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement, J. VLSI Signal Process. Syst.Signal Image Video Technol. 38 (1) (2004) 35–44.
[10] A.W. Setiawan, T.R. Mengko, O.S. Santoso, A.B. Suksmono, Color retinal image enhancement using CLAHE, in: International Conference on ICT for Smart Society, IEEE, 2013, pp. 1–3.
[11] R. Korifi, Y. Le Dréau, J.-F. Antinelli, R. Valls, N. Dupuy, CIEL*a*b* color space predictive models for colorimetry devices–analysis of perfume quality, Talanta 104 (2013) 58–66.
[12] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2010) 2341–2353.
[13] L. Xiong, H. Li, L. Xu, An enhancement method for color retinal images based on image formation model, Comput. Methods Programs Biomed. 143 (2017) 137–150.
[14] A. Gaudio, A. Smailagic, A. Campilho, Enhancement of retinal fundus images via pixel color amplification, in: International Conference on Image Analysis and Recognition, Springer, 2020, pp. 299–312.
[15] S. Zhang, C.A. Webers, T.T. Berendschot, A double-pass fundus reflection model for efficient single retinal image enhancement, Signal Process. (2021) 108400.
[16] L. Yao, Y. Lin, S. Muhammad, An improved multi-scale image enhancement method based on retinex theory, J. Med. Imaging Health Inform. 8 (1) (2018) 122–126.
[17] L. Cao, H. Li, Y. Zhang, Retinal image enhancement using low-pass filtering and α-rooting, Signal Process. 170 (2020) 107445.
[18] Y. Luo, K. Chen, L. Liu, J. Liu, J. Mao, G. Ke, M. Sun, Dehaze of cataractous retinal images using an unpaired generative adversarial network, IEEE J. Biomed. Health Inform. 24 (12) (2020) 3374–3383.
[19] A. Qayyum, W. Sultani, F. Shamshad, J. Qadir, R. Tufail, Single-shot retinal image enhancement using deep image priors, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 636–646.
[20] Y. Gandelsman, A. Shocher, M. Irani, "Double-DIP": unsupervised image decomposition via coupled deep-image-priors, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 11026–11035.
[21] Q. You, C. Wan, J. Sun, J. Shen, H. Ye, Q. Yu, Fundus image enhancement method based on CycleGAN, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 4500–4503.
[22] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, CBAM: convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
[23] H. Zhao, B. Yang, L. Cao, H. Li, Data-driven enhancement of blurry retinal images via generative adversarial networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 75–83.
[24] J.-J. Yang, J. Li, R. Shen, Y. Zeng, J. He, J. Bi, Y. Li, Q. Zhang, L. Peng, Q. Wang, Exploiting ensemble learning for automatic cataract detection and grading, Comput. Methods Programs Biomed. 124 (2016) 45–57.
[25] 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, Springer, 2015, pp. 234–241.
[26] H. Fu, B. Wang, J. Shen, S. Cui, Y. Xu, J. Liu, L. Shao, Evaluation of retinal image quality assessment networks in different color-spaces, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 48–56.
[27] M.M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A.R. Rudnicka, C.G. Owen, S.A. Barman, An ensemble classification-based approach applied to retinal blood vessel segmentation, IEEE Trans. Biomed. Eng. 59 (9) (2012) 2538–2548.
[28] J. Staal, M.D. Abràmoff, M. Niemeijer, M.A. Viergever, B. Van Ginneken, Ridge-based vessel segmentation in color images of the retina, IEEE Trans. Med. Imaging 23 (4) (2004) 501–509.
[29] S. Holm, G. Russell, V. Nourrit, N. McLoughlin, DR HAGIS–a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients, J. Med. Imaging 4 (1) (2017) 014503.
[30] A. Budai, R. Bock, A. Maier, J. Hornegger, G. Michelson, Robust vessel segmentation in fundus images, Int. J. Biomed. Imaging 2013 (2013).
[31] J. Zhang, B. Dashtbozorg, E. Bekkers, J.P. Pluim, R. Duits, B.M. ter Haar Romeny, Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores, IEEE Trans. Med. Imaging 35 (12) (2016) 2631–2644.
[32] J. Chen, J. Tian, N. Lee, J. Zheng, R.T. Smith, A.F. Laine, A partial intensity invariant feature descriptor for multimodal retinal image registration, IEEE Trans. Biomed. Eng. 57 (7) (2010) 1707–1718.
[33] T. Park, A.A. Efros, R. Zhang, J.-Y. Zhu, Contrastive learning for unpaired image–to-image translation, in: European Conference on Computer Vision, Springer, 2020, pp. 319–345.
[34] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[35] J. Zhang, H. Li, Q. Nie, L. Cheng, A retinal vessel boundary tracking method based on Bayesian theory and multi-scale line detection, Comput. Med. Imaging Graph. 38 (6) (2014) 517–525.
[36] S. Pachade, P. Porwal, D. Thulkar, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, L. Giancardo, G. Quellec, F. Mériaudeau, Retinal fundus multi-disease image dataset (RFMiD): a dataset for multi-disease detection research, Data 6 (2) (2021) 14.
[37] M. Tan, Q. Le, EfficientNet: rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.

Bingyu Yang is a PhD candidate at the School of Information and Electronics, Beijing Institute of Technology, Beijing, China.

He Zhao received the PhD degree from Beijing Institute of Technology, Beijing, China, in 2020. His research interests are medical image synthesis and disease detection.

Lvchen Cao received the PhD degree from Beijing Institute of Technology, Beijing, China, in 2021. His research interests are classification and enhancement of low-quality medical images.

Hanruo Liu is an associate professor of ophthalmology at Beijing Institute of Ophthalmology, Beijing Tongren Hospital, Capital Medical University. Her research interests are diagnosis of glaucoma.

Ningli Wang is the director of Beijing Institute of Ophthalmology, Beijing Tongren Hospital, Capital Medical University. His contributions include the trans-lamina cribrosa pressure difference theory of open-angle glaucoma.

Huiqi Li received PhD degree from Nanyang Technological University, Singapore in 2003. She is currently a professor at Beijing Institute of Technology. Her research interests are medical image processing and computer-aided diagnosis.