

Modeling and Enhancing Low-Quality Retinal Fundus Images

Ziyi Shen^{ID}, Member, IEEE, Huazhu Fu^{ID}, Senior Member, IEEE,
Jianbing Shen^{ID}, Senior Member, IEEE, and Ling Shao, Fellow, IEEE

Abstract—Retinal fundus images are widely used for the clinical screening and diagnosis of eye diseases. However, fundus images captured by operators with various levels of experience have a large variation in quality. Low-quality fundus images increase uncertainty in clinical observation and lead to the risk of misdiagnosis. However, due to the special optical beam of fundus imaging and structure of the retina, natural image enhancement methods cannot be utilized directly to address this. In this article, we first analyze the ophthalmoscope imaging system and simulate a reliable degradation of major inferior-quality factors, including uneven illumination, image blurring, and artifacts. Then, based on the degradation model, a clinically oriented fundus enhancement network (cofe-Net) is proposed to suppress global degradation factors, while simultaneously preserving anatomical retinal structures and pathological characteristics for clinical observation and analysis. Experiments on both synthetic and real images demonstrate that our algorithm effectively corrects low-quality fundus images without losing retinal details. Moreover, we also show that the fundus correction method can benefit medical image analysis applications, e.g., retinal vessel segmentation and optic disc/cup detection.

Index Terms—Medical image enhancement, retinal fundus image, deep neural network.

I. INTRODUCTION

DUCE to their safety and cost-effectiveness in acquiring, retinal fundus images are widely used by both ophthalmologists and computer-aided diagnosis systems for the clinical screening and diagnosis of ocular diseases [1], [2]. However, fundus images tend to experience large variations in quality. A screening study of 5,575 patients found that about 12% of fundus images are not of adequate quality to be readable by ophthalmologists [3]. In some cases, when the degradation is caused by the images being obtained through *internal* cataractous turbid media, enhancement methods, such as [4], can be used to restore ‘high quality’. Then, the corrected

Manuscript received September 20, 2020; revised November 18, 2020 and December 1, 2020; accepted December 5, 2020. Date of publication December 9, 2020; date of current version March 2, 2021. (Corresponding authors: Huazhu Fu; Jianbing Shen.)

The authors are with the Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates (e-mail: ziyi.shen@inceptioniai.org; huazhu.fu@inceptioniai.org; jianbing.shen@inceptioniai.org; ling.shao@inceptioniai.org).

Digital Object Identifier 10.1109/TMI.2020.3043495

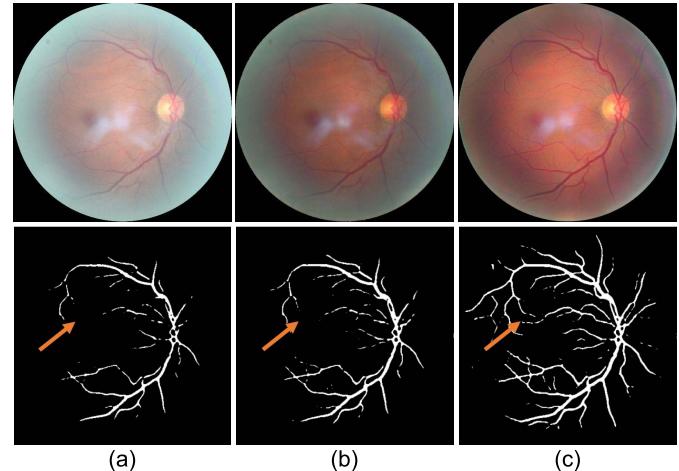


Fig. 1. Illustration of fundus image correction. (a) The low-quality fundus image. (b) Result enhanced by [8]. (c) Result enhanced by our method. The first row shows the fundus images, and bottom row is the corresponding vessel segmentation results using an automated method [7]. Our method corrects the low-quality fundus image while enhancing the clinical structures (e.g., vessel, optic disc and cup regions).

images can be used to support the observation of other diseases (e.g., age-related maculopathy, diabetic retinopathy, and glaucoma). However, in addition to this pathogenic degradation, in real applications, *external* interference factors caused by handcrafted imaging equipment and poor environmental conditions are also common. For instance, images are often taken under different lighting environments, using various cameras, and by distinct operators with varying levels of experience. Common examples of low-quality factors in retinal fundus images thus include uneven illumination, image blurring, and artifacts, which not only prevent reliable diagnosis by ophthalmologists, but also affect the performance of automated image analyzing systems [5], [6]. An example is shown in Fig. 1 (a), where uneven illumination and artifacts prevent the vessel and disc region from being fully/clearly observed, and affect the performance of the automated vessel segmentation method (i.e., [7]).

Recently, general image enhancement methods have achieved state-of-the-art performances, especially with the development of deep learning techniques [9], [10]. However, different from general images, retinal fundus images are

acquired through a special ophthalmoscope imaging process to capture anatomical retinal structure for clinical diagnosis, which introduces various additional challenges. First, the retina cannot be illuminated internally; both the incident and reflected imaging beams have to traverse the pupil. Moreover, the spherical geometry of the eye creates significant inter-reflection, resulting in shading artifacts [11]. Second, anatomical retinal structures (e.g., vessel, optic disc and cup) in fundus images are limited but highly important for clinical diagnosis, and should thus be enhanced in the correcting process. Third, some pathological characteristics (e.g., hemorrhages, microaneurysms, and drusen) are usually only a few pixels wide and appear as circular shapes, causing them to be easily confused with artifacts and noise. These issues mean that fundus image correction methods need to be able to both suppress the undesired low-quality factors and preserve the pathological characteristics simultaneously, which general enhancement techniques cannot do to a satisfactory level. For example, Fig. 1 (b) shows the enhanced result of Fig. 1 (a) when using the general image enhancement method [8], where the disc region still suffers from artifacts, and vessels are miss-segmented by the automated system.

To address these issues, in this article, we design a degradation model that simulates major factors of low-quality fundus images, including light transmission disturbance, image blurring, and retinal artifacts. Then, a clinically oriented fundus enhancement method (cofe-Net) is proposed to suppress the local outliers and undesired artifacts, while at the same time preserving the anatomical retinal structures, e.g., vessel and optic disc/cup regions. To this end, two new modules, the retinal structure activation (RSA) and clinical low-quality activation (LQA), are introduced. The RSA module is used to preserve the retinal structure, while the LQA is employed to remove the low-quality factors. Based on the human perception mechanism, the proposed network with additional error metrics for perceiving artifacts, is able to correct fundus images with more accurate structures and suppress local defects. An example of a fundus image corrected by our method is shown in Fig. 1 (c). The main contributions of this article are summarized as follows:

- 1) A fundus degradation model based on the retinal ophthalmoscope imaging system is designed to simulate low-quality fundus images. It can be widely utilized to support the typical propagation scheme in fundus image generation models. To the best of our knowledge, this is the first work to model the optical ophthalmoscope. All degradation models are designed based on several imaging stages.
- 2) A novel clinically oriented fundus enhancement network (cofe-Net) is developed to correct low-quality fundus images for clinical observation and analysis. Our cofe-Net preserves the anatomical retinal structures of fundus image using the RSA module and suppresses undesired artifacts with the LQA module.
- 3) We show that fundus correction can boost the performances of clinical analysis tasks, *e.g.*, vessel segmentation and disc/cup detection, on poor-quality images. Experimental results on both synthetic and real fundus

images demonstrate that our algorithm performs favorably against state-of-the-art approaches.¹

II. RELATED WORK

In this section, we summarize current image correction techniques and discuss the algorithms that have been specifically applied to fundus images. Several methods [11], [12] exploit image contrast normalization and contrast limited adaptive histogram equalization (CLAHE) techniques to restore an image. For example, Setiawan *et al.* [13] applied CLAHE to fundus image enhancement specifically. Instead of simply considering the color and texture information, some algorithms [14]–[16] decompose the reflection and illumination, achieving image enhancement and correction by estimating the solution in an alternate minimization scheme. Guo *et al.* [17] further proposed to refine the illumination map and enhance low-light images. These models have been extended to an integrated scheme incorporating gamma correction [18] and CLAHE for fundus image luminance and contrast adaption [19], [20]. While these algorithms based on a bottom-up framework are effective, their optimal solution relies heavily on global image statistics and mapping functions, ignoring discriminative features, which may introduce undesired artifacts and distortion.

Along another line, learning based methods have also been developed for the image correction task, utilizing the various features extracted from the images to learn directional filters. For instance, latent image priors have been adopted for correction and restoration in sparsity-based models [21], [22], distribution fitting algorithms [23]–[25], variational frameworks [26], and latent structure-driven methods [27]–[29], all of which have also been specifically applied to fundus images [30]–[33]. These algorithms typically constrain the optimal solution using a regularization scheme to solve the non-convex problem. This may incur a heavy computational cost, limiting their applicability in clinical settings.

Recently, due to the powerful image representation ability, deep learning techniques have been widely used in computer vision and medical image analysis. This has enabled the rapid advancement of reconstruction techniques, making them much better equipped to address various challenging tasks, such as low-light image enhancement [10], [34], [35], dehazing and deraining [36]–[38], and deblurring [39], [40]. For image correction, convolutional neural network (CNN) based approaches attempt to learn a mapping operator between the ground truth and low-quality images [8], [41]–[43]. Eilertsen *et al.* [8] proposed to solve the high dynamic range task in an end-to-end fashion, under the constraint of a pixel-wise loss. This method produces accurate global tone-mapping but results in an over-smooth solution. Ren *et al.* [41] and Lv *et al.* [42] aimed to enhance images in a coarse-to-fine manner, using a multi-scale framework and a feature fusion mechanism, respectively. In addition, Talebi and Milanfar [44] introduced a deep neural image assessment model and applied it to restore more content under extreme conditions (*e.g.*, dark and bright areas). These end-to-end methods aim to learn an optimal

¹The source code of our method is available at https://github.com/HzFu/EyeQ_enhancement

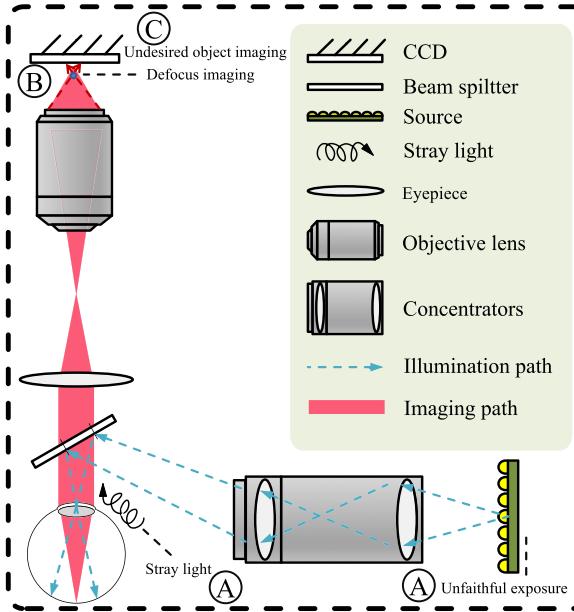


Fig. 2. Schematic diagram of ophthalmoscope imaging system. It contains an illumination system and an imaging system. The clinical imaging process in practice undergoes several types of interference, such as exposure and imaging outliers, which introduce unexpected drawbacks to the retinal fundus images and result in image degradation. Besides, the capturing of undesired objects is also a crucial factor that interferes with the image quality and impacts the diagnostic process.

solution by simply minimizing the content loss. Focusing on images stocked with explicit prior information, Liu *et al.* [45] proposed a deep prior ensemble and integrated knowledge-driven cues for natural image enhancement. Different from the non-convex optimization framework, deep learning methods follow a heuristic pattern that relies heavily on a great deal of training data. Due to the particular pathological characteristics and disease markers, general CNN-based models often do not perform well on medical samples, especially fundus images. To deblur and enhance clinical images, Zhao *et al.* [46] and Liu *et al.* [47] applied the adversarial loss. However, while computationally efficient, these methods only focus on generating photo-realistic images, ignoring the lesion areas significant to clinical applications. Therefore, designing an effective deep learning model for fundus image correction is the focus of this work.

III. FUNDUS IMAGE DEGRADATION MODEL

Clinical image collection in a complex environment using an ophthalmoscope often encounters several types of interference, introduced in the optical feed-forward system. For instance, as shown in Fig. 2, light transmission disturbance is often caused by exposure issues. Due to the interspace between the eye and camera, stray light may enter into the ophthalmoscope, mix with the lighting source and result in uneven exposure. This also affects the tuning setting of the programmed exposure, causing global over-/under-exposure. In addition, image blurring caused by human factors (such as eyeball movement, fluttering, and defocus) results in low-quality images. Besides, the capturing of undesired objects (e.g., dust) during imaging

is also a crucial factor that reduces image quality and impedes subsequent diagnosis. In this section, we propose a reformulated representation of the interference that occurs during the collection of fundus images. Our degradation model could be used to not only support current fundus propagation models, but also synthesize a high-quality pairwise fundus dataset for subsequent research. We summarize the interference in terms of three factors, including *light transmission disturbance*, *image blurring*, and *retinal artifacts*.

A. Light Transmission Disturbance

We first introduce the interference caused by light, which can be categorized into two types, global and local factors. For global factor, since existing fundus cameras are programmed with auto-exposure, ambient light will impact the illumination. Unstable stray light, which affects the camera configuration (light source), may result in under-/over-exposure during image collection. Furthermore, the subjective situation and manual mydriasis also cause disturbance during the imaging procedure. For local factor, the sensitivities of specific regions on the image plane may differ, causing uneven illumination in an image. This can be due to the interspace between the fundus and ophthalmoscope. An initiative light leak phenomenon caused by the patient or an inappropriate exposure imported to the imaging plane with a variant distance to the optical axis will result in uneven illumination. In addition, depending on the equipment design of the ophthalmoscope system, the diverse lens apertures and embedded optical compensation mechanism can also limit the amount of light, as well as affect the dynamic range of the fundus images. Here, we simulate the degradation using an aggregation model. The global factors are modeled by contrast, brightness and saturation interference, while the local factors are defined as additional non-uniform illumination on the fundus image. To formulate the light degradation model, given a ground truth (clean) image \mathbf{x} , its degraded image \mathbf{x}' with light transmission disturbance is defined as:

$$\mathbf{x}' = \text{clip}(\alpha(\mathbf{J} \cdot G_L(r_L, \sigma_L) + \mathbf{x}) + \beta; s), \quad (1)$$

where α , β and s refer to the contrast factor, brightness and saturation, respectively. We use a clipping function $\text{clip}(\cdot; s)$ to model the global degradation as a saturation process. For local light leak/lack, \mathbf{J} is defined as an illumination bias to be over-/under-illuminated at a panel centered at (a, b) with a radius of r_L . Therefore, each of its entries is defined as:

$$J_{ij} = n_l |_{(i-a)^2 + (j-b)^2 < r_L^2}, \quad (2)$$

where n_l is the over-/under-intensity weight. A Gaussian kernel G_L is then applied to ensure luminance smoothness. Fig. 3 (b) gives an example with over-exposure and uneven illumination synthesized by our degradation model.

B. Image Blurring

During the fundus imaging procedure, in addition to issues caused by the program settings, human operator error can also be introduced. Since the ophthalmoscope is applied externally on patients, the distance between the fundus and object plane

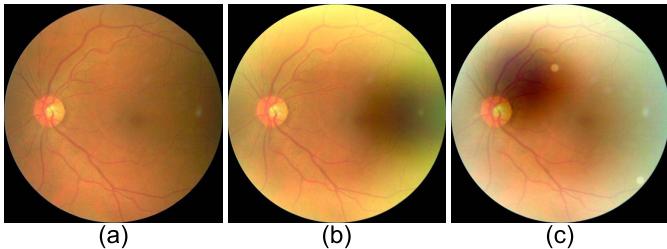


Fig. 3. Synthetic degraded fundus images. (a) Clear image. (b) Simulated sample undergoing light transmission disturbance. (c) Simulated sample undergoing light transmission disturbance, image blurring, and retinal artifacts.

is a random variable; a wrong setting for the focal length of the optical system during the funduscopy, or a dilated pupil caused by light simulation, can cause an undesired object distance between the image plane and lens, and result in image blurring. To simulate this, we define the defocus blurring as:

$$\mathbf{x}' = \mathbf{x} \cdot G_B(r_B, \sigma_B) + n, \quad (3)$$

where G_B is a Gaussian filter with a radius r_B and spatial constant σ_B , and n denotes the additive random Gaussian noise. The degraded fundus image \mathbf{x}' is generated via a convolution operation between the ground truth image \mathbf{x} and a Gaussian filter. Specifically, the blur radius r_B is defined as:

$$r_B = \frac{F \cdot v_0 - D(F - v_0)}{D \cdot f}, \quad (4)$$

where F is the focal length of the optical lens system, and f is its f-number. D and v_0 denote the object distance and image distance, respectively.

C. Retinal Artifact

The above formulated models simulate internal factors, which are caused by the imaging system itself. However, additional degradation may be introduced if the imaging is conducted under a poor condition. Although not additive or multiplicative noise, dust and grains attaching on the lens of the imaging plane can also yield blurred images that affect the fundus image quality and following diagnosis. We define this type of interference as retinal artifact, which is modeled by a multi-step imaging procedure:

$$\mathbf{x}' = \mathbf{x} + \sum_k^K G_R(r_k/4, \sigma_k) \cdot \mathbf{o}_k. \quad (5)$$

In the context of random light fields, r_k and σ_k are defined as the radius and variance of an undesired object k imaged on the plane, where r_k can be calculated as the same way in Eq. 4. K is the undesired object number. Parameter \mathbf{o}_k is the luminance bias. We again define a Gaussian filter G_R to simulate the defocused imaging of an undesired object.

D. Degradation

Totally, we propose three different degradation models in Eqs. (1, 3, and 5), which simulate light transmission disturbance, image blurring, and retinal artifacts, respectively.

These models can be utilized independently to synthesize individual low-quality factor, or randomly combined for the more complex degradation. An example containing various types of degradation is shown in Fig. 3 (c), where our proposed degradation model is particularly well-designed to simulate the real funduscopic examination procedure.

IV. CLINICALLY ORIENTED FUNDUS ENHANCEMENT NETWORK

For medical image analysis, there are certain crucial cues that require special attention. Thus, not only is it necessary to preserve the fine structures in a clinical image, but also the pathological characteristics. In other words, it is essential to correct the images while simultaneously guaranteeing they remain clinically significant for disease diagnosis and analysis. Therefore, we propose a clinically oriented fundus enhancement network (cofe-Net) to correct the low-quality fundus image, as shown in Fig. 4. Given a low-quality fundus image, a low-quality activation (LQA) module is employed to identify the retinal artifacts, while a retinal structure activation (RSA) module is used to perceive the latent retinal structure. Then, an image correction module is introduced to fuse these prior information with the latent features of the image to produce the final corrected image.

A. Low-Quality Activation Module

To enhance fundus images and encourage valid and accurate pathological analysis of ocular diseases, it is necessary to exploit a learning-based solution that extracts latent features from the low-quality fundus image and learns a good representation to reconstruct it. Thus, we first define a generative network \mathcal{G} . Let \mathbf{I} and $\hat{\mathbf{I}}$ denote the low-quality input image and the corrected image, respectively, with size $w \times h \times c$, where w and h denote the width and height of the image, and c is the channel number (e.g., $c = 3$ for an RGB image). The correction procedure can be formulated as:

$$\hat{\mathbf{I}} = \mathcal{G}(\mathbf{I}; \mathbf{W}_{\mathcal{G}}), \quad (6)$$

where $\mathbf{W}_{\mathcal{G}}$ denotes the learnable parameters of network \mathcal{G} . To progressively generate the corrected images, a robust \mathcal{L}_2 loss is used as the object function:

$$\mathcal{L}_c = \|\mathbf{I}_G - \mathcal{G}(\mathbf{I}, \mathbf{W}_{\mathcal{G}})\|_2^2, \quad (7)$$

where \mathbf{I}_G is the ground truth (clean) image. However, only using the \mathcal{L}_2 loss to guide the convergence of the neural network often results in an over-smooth solution due to the similarity between the input image and the restored image in terms of their pixel-wise appearances. Therefore, the vanilla model simply pays equal attention to all areas of the global image, without taking clinical regions (e.g., vessels, disc/cup regions) into special consideration. The model may thus fail to remove local artifacts, or preserve real structures.

To address this issue, we introduce a supervised activation mechanism, the LQA module, to perceive the undesired low-quality factors in clinical images. Our LQA module \mathcal{M} is built upon a convolutional encoder-decoder architecture, as shown in red part of Fig. 4, and produces artifact attention map

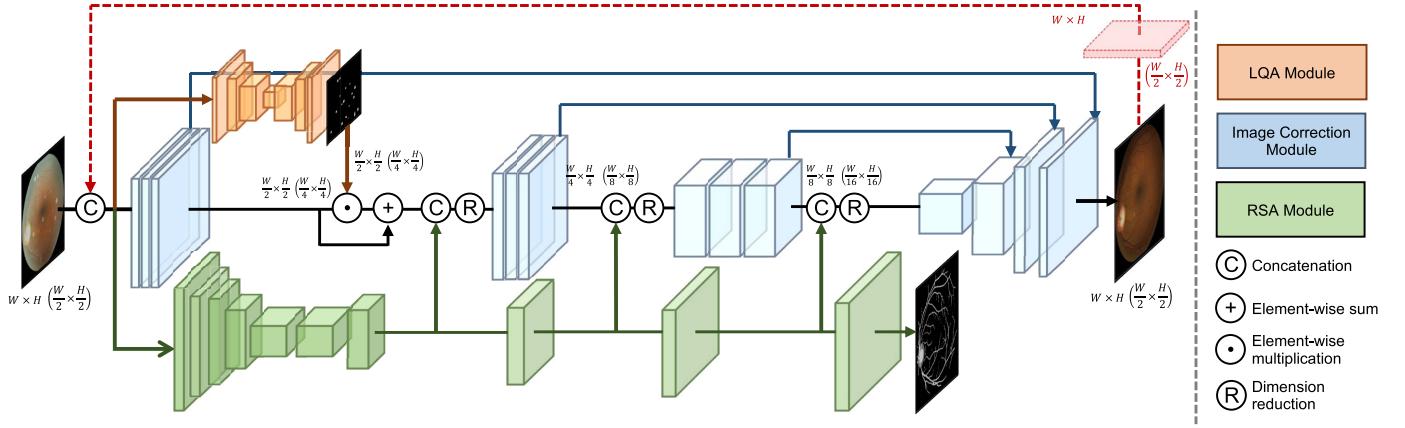


Fig. 4. Architecture of the proposed network. A low-quality image is simultaneously rendered to three sub-modules, i.e., the low quality activation (LQA) module, the retinal structure activation (RSA) module and the image correction module. The LQA module activates and embeds the representation of the low-quality region into the correction module at an early stage, while the RSA module continuously injects retinal structural information to the correction network by feature map concatenation at each scale. Leveraging this additional information, the image correction module enhances high-quality images based on an encoder-decoder network. Note that we cascade the network for two scales ($0.5\times$ and $1\times$) using shared weights. The red dotted line denotes the upsampled result of $0.5\times$ scale, which is only used in the first $0.5\times$ loop.

m corresponding the low-quality region. This differentiable neural activation operator can be defined as:

$$\mathbf{m} = \sigma(\mathcal{M}(\mathbf{I}, \mathbf{W}_\mathcal{M})), \quad (8)$$

where σ denotes the non-linear activation function that normalizes the feature. The proposed correction framework is designed to perceive local artifacts via LQA branch with parameters $\mathbf{W}_\mathcal{M}$, and embed the corresponding map to guide the correction. Let $\mathbf{F}_\mathcal{L}$ denote the low-level features extracted by the image correction network. We combine the artifact attention map \mathbf{m} of LQA and $\mathbf{F}_\mathcal{L}$ as: $\mathbf{F}_\mathcal{L} \odot \mathbf{m} + \mathbf{F}_\mathcal{L}$, where \odot is an element-wise multiplication. The artifact aware features extracted from the LQA are integrated with the low-level features into the correction model. Therefore, the embedded framework encourages a more efficient representation and achieves a higher ability for distinguishing the low-quality region. Here, we utilize a pixel-wise \mathcal{L}_2 loss to constrain the LQA module training:

$$\mathcal{L}_m = \|\mathbf{m}_G - \mathbf{m}\|_2^2, \quad (9)$$

where \mathbf{m}_G denotes the ground truth artifact mask. The supervision enables the LQA branch to efficiently learn an activation representation. Beside the \mathcal{L}_2 loss function, we also apply another relevant loss function to identify the undesired artifacts and train the correction network by minimizing the low-quality perceiving loss:

$$\mathcal{L}_p = \|\mathbf{m} \odot (\mathbf{I}_G - \mathcal{G}(\mathbf{I}; \mathbf{W}_G))\|_2^2, \quad (10)$$

where \mathbf{I}_G is the the ground truth (clean) image. This low-quality perceiving loss encourages to highlight the correction of artifact regions. The LQA module can assist the correction by focusing on the undesired artifacts.

In summary, the proposed LQA module aims to highlight the representation of retinal artifact in the fundus image. Since undesired objects (e.g., dust and grains) on the object lens often have similar appearances, we tend to merely perceive the shape and position of this kind of artifact. Therefore, we

apply a small-scale network for the LQA module and set the input to be a $0.5\times$ image to reduce the computational load. As shown in Fig. 4, the LQA module contains three encoder for extracting the features, and the perceived information is integrated and represented by the three corresponding encoder blocks. Each encoder block contains three convolutional layers with 3×3 kernels followed by a $2\times$ max-pooling layer for downsampling, while each decoder block includes one 3×3 convolutional layer and one 4×4 transposed convolutional layer for upsampling the feature maps. Each convolutional layer is followed by a ReLU function as the non-linear activation. Finally, a 1×1 convolutional layer is utilized to predict the artifact attention map.

B. Retinal Structure Activation Module

Natural image enhancement method tends to reconstruct images with more diverse context in order to please human perception, often containing exaggerated details. However, in some cases, this may not have positive impact. This is especially true for medical images, such as fundus images, which are used as the foundations for clinical diagnosis. The incorrectly synthetic content may heavily skew the pathological description.

To address this issue, we introduce a retinal structure activation (RSA) module to preserve features of the retinal structure and guide the reconstruction processing. As illustrated in Fig. 4, the RSA module first encodes the input image using a pre-trained ResNet-34. The features are directly extracted by the decoding operator for the following structure extraction. Then, the decoding residual layers are used to scale up the bottleneck features (by a factor of 16) to the original image size. An additive skip connection from the encoder is applied to the corresponding decoding layers. Our RSA module predicts a segmentation mask map for main retinal structures, e.g., vessel, optic disc and cup regions. We utilize a mean squared error (MSE) loss as the segmentation loss to constrain the prediction and simultaneously, the non-linear RSA feature

maps obtained from different depths are explicitly utilized to assist the main correction model. This is formulated as:

$$\mathbf{S}^l = \mathcal{T}_S^n (\text{Concat} [\mathbf{F}_S^n(\mathbf{I}), \mathbf{F}_G^n(\mathbf{I})]), \quad (11)$$

where $\mathbf{F}_S^n(\mathbf{I})$ and $\mathbf{F}_G^n(\mathbf{I})$ denote the features extracted from the n^{th} -level of the RSA module and image correction network, respectively. $\text{Concat}[\cdot, \cdot]$ is a concatenation operator, and the RSA map is fused via a learnable non-linear transformation filter $\mathcal{T}_S^n(\cdot)$. The feature maps from different levels are stacked and embedded with directional information to enable an accurate mapping between the features of the RSA module and our image correction module. Thus, they are able to provide the network with a stronger representational ability, and preserve more realistic and meaningful content. In this way, the key components of the fundus image are encoded in the hidden states of the image correction network, considerably reducing the problems of content vanishing and singular positioning.

C. Image Correction Module

With the assistance of the LQA and RSA modules, we design the image correction network \mathcal{G} as a two-scale cascaded framework to progressively enhance images in a coarse-to-fine manner. Specifically, we propose to train a single encoder-decoder network for two scales ($0.5 \times$ and $1 \times$) by sharing weights to reduce the number of trainable parameters. As shown in Fig. 4, we first encode the image, inferring features by identifying the critical information. This is done by densely embedding (concatenating) the knowledge learned from the RSA module. The features aggregated at each level are integrated and their dimensions are reduced by a convolutional layer. Then, a symmetric decoder is applied to generate the corrected image. We also add skip connections between the encoder and corresponding decoder to avoid gradient explosion or vanishing during training. For the two-scale cascaded framework, we first feed the input image with a downsampled size (i.e., $0.5 \times$ scale) to the network, and then utilize a transposed convolutional layer to produce the corrected result with the original size (i.e., $1 \times$ scale). After that, the input image is concatenated with the upsampled corrected result and fed into the image correction network again to be processed at the $1 \times$ scale. Note that the input of the $1 \times$ scale is comprised of the upsampled result from the $0.5 \times$ scale and the original image, while the input of our image correction network has six channels. We thus duplicate the input image of the first scale (i.e., $0.5 \times$ scale) to provide the same number of channels. Finally, the overall loss function for our whole network is defined as:

$$\mathcal{L} = \mathcal{L}_m + \sum_s (\lambda_p \mathcal{L}_p^s + \lambda_c \mathcal{L}_c^s) + \lambda_v \mathcal{L}_v, \quad (12)$$

where \mathcal{L}_v is the MSE loss of the RSA module, and s denotes the scale index (e.g., $s = 0.5, 1$). λ_p , λ_c and λ_v are weights that balance the different losses. Note that \mathcal{L}_p and \mathcal{L}_c are calculated at both scales to control the image reconstruction. \mathcal{L}_v and \mathcal{L}_m are simultaneously computed only at the original scale (i.e., $1 \times$). Here we set $\lambda_p = 10$, $\lambda_c = 1$, $\lambda_v = 0.1$. These penalty terms are employed to stabilize the training process and ensure the corrected image is visually reasonable.

V. EXPERIMENTS

A. Datasets

To train our cofe-Net, we manually select 13,000 high-quality images, free from interference factors, from the EyeQ dataset [5], which is based on the Kaggle Diabetic Retinopathy Detection dataset [48]. We then randomly choose degradation factors (e.g., light transmission disturbance, image blurring, and retinal artifacts), and process the images using the proposed degradation formulation to generate low-quality images, which make up the training set. Note that we also process the images by combining all factors to simulate a complicated real-world situation. For testing, we also utilize the proposed degradation model to randomly generate degraded images for quantitative evaluation. We synthesize 500 testing images in total using the Kaggle [48] dataset and DRIVE [49] dataset. Here peak signal to noise ratio (PSNR) and structural similarity (SSIM) are utilized to evaluate the image quality. In addition, for qualitative analysis, we randomly choose 50 images from the Kaggle dataset [48], and build a benchmark to support a user study task.

B. Implementation

During training, we resize the degraded input images to 512×512 with a batch size of 16 for each iteration. To accelerate the training process, we first train the RSA and LQA modules, and then apply an end-to-end training strategy to the whole correction framework. The proposed cofe-Net is implemented in PyTorch, using stochastic gradient descent (SGD) for optimization. The learning rate is initialized to 1×10^{-4} for the first 150 epochs, and then gradually decayed to zero over the next 150 epochs. The network is trained on an NVIDIA Tesla V100 GPU with 32GB of memory per card. The number of model parameters is roughly 41M, and the computational cost is 270 GMac.

C. Degradation Model Settings

In the proposed retinal fundus image degradation model, we summarize the interference in terms of three factors, including light transmission disturbance, image blurring and retinal artifacts. To better simulate real clinical scenarios, we collect the images from the Kaggle [48] dataset, and use different settings to produce degradation cases.

1) Light Transmission Disturbance: The light transmission disturbance is achieved using Eq. 1. We simulate this interference using an aggregation model, which consists of global and local factors. The global factors include contrast (α), brightness (β) and saturation (s). Thus, we simulate random color jitters with a probability of -0.5 to 0.5 . The local factors are defined as additional non-uniform illumination on the fundus image. To simulate the light leak phenomenon, we randomly define the panel centered at $c = (a, b) \in [0.375 r_L, 0.625 r_L]$, and $r_L \in [0.75 w, w]$, where w denotes the image size. The Gaussian filter G_L is then applied to ensure luminance smoothness. Here we define $\sigma_L \in [0.66 cr_L, 0.66(w - c)r_L]$. Specifically, to model the typical light leak, we define three kinds of illumination bias as $[0.63, 0.80, 0.35]$, $[0.56, 0.93, 0.93]$,

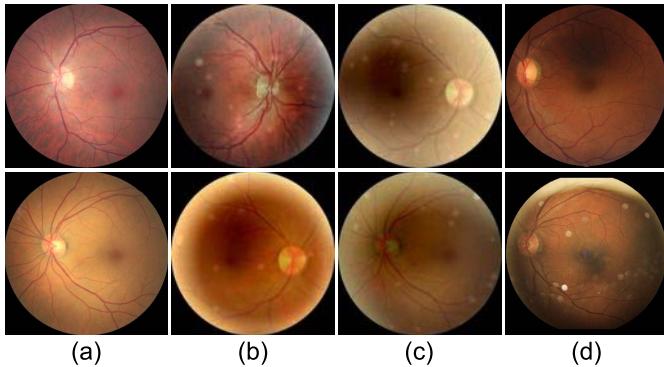


Fig. 5. Synthetic samples in low-quality fundus image dataset. (a) Ground truth images. (b-c) Synthesized low-quality images. (d) Real low-quality images.

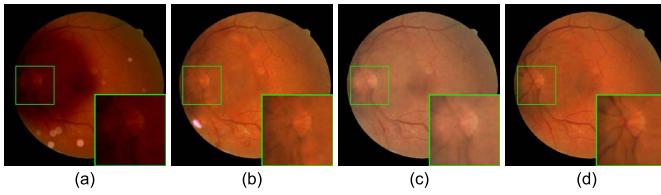


Fig. 6. Effect of structure perceiving activation model. (a) Low-quality image. (b) Backbone model. (c) With RSA module. (d) With RSA and LQA modules. The proposed method is capable of correcting the clinical fundus images with more structures and less artifacts.

and [1, 1, 1]. With this, we effectively simulate the real light leak in the imaging procedure. Similar to the degradation model for light leak, for the uneven exposure problem, we further define $r_L \in [0.3w, 0.5w]$, and the illumination bias for the underexposure model is simulated as local brightness jitters with a probability of -0.5 to -0.1 . The $\sigma_L \in [0.55r_L, 0.75r_L]$ of the Gaussian filter is defined for the following smooth operator.

2) Image Blurring: The image blurring is simulated by using Eq. 3, which is caused by an undesired object distance setting during funduscopic imaging. Here we set $\sigma_B = 0.03 w$, and $r_B \in [0.01 w, 0.015 w]$.

3) Retinal Artifact: The retinal artifacts are achieved using Eq. 5. We define object i as having a radius $r_i \in [0.025w, 0.05w]$, while the defocused imaging is defined by a Gaussian operator G_R with $\sigma_i = 5 + 0.8r_i$. The illumination bias in this item is defined as $1 - e^{-(0.5+0.04r_i) \times (0.012r_i)}$. To model the imaging of undesired objects, we randomly add 10 to 25 artifacts to each fundus image, which can simulate the interference in real clinical scenarios. In Fig. 5, we provide examples of images generated with the different degradation factors. The constructed low-quality image dataset can be used to train our model for retinal fundus image correction.

D. Ablation Study

To demonstrate the effectiveness of cofe-Net, we conduct several ablation studies. By removing the RSA and LQA modules from the proposed architecture, the model becomes a multi-scale baseline generation network, which is constrained by an \mathcal{L}_2 loss. As shown in Fig. 6 (a), unlike natural images

TABLE I
ABLATION STUDY. OUR FULL MODEL PRODUCES COMPARABLE RESULTS ON THE TESTING DATASETS IN TERMS OF VARIOUS METRICS.

Multi-scale	\mathcal{L}_c	RSA	LQA	DRIVE [49]		Kaggle [48]	
				PSNR	SSIM	PSNR	SSIM
	✓			20.07	0.733	16.86	0.798
	✓	✓		21.06	0.748	17.04	0.807
	✓	✓	✓	21.24	0.740	18.48	0.877
	✓	✓	✓	21.24	0.758	20.51	0.885

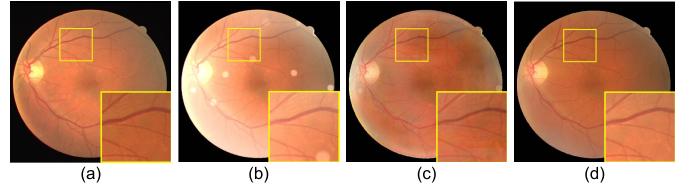


Fig. 7. Effect of multi-scale framework on enhancement. (a) Ground truth image. (b) Low-quality image. (c) Without multi-scale framework. (d) With multi-scale framework. The proposed multi-scale framework corrects fundus images with more details.

which contain abundant mid-frequency structures and high-frequency textural details, the features represented in fundus images often appear as small vessel branches and their corresponding lesions. The network optimized solely from the \mathcal{L}_2 loss generates relatively smooth images, as shown in Fig. 6 (b), preventing the tiny vessels from being reconstructed with sharp details. The features and morphology of the lesions, such as hemorrhages, are also unable to be preserved. To address this issue, we embed the RSA module as a strong prior to guide the correction procedure. In Fig. 6 (c), our method accurately generates noticeable features e.g., vessels, and can also more accurately distinguish the disc/cup from the fundus region.

Instead of simply considering generic factors affecting the imaging quality, such as exposure and defocus, some undesired matters in initiative situations are also taken into account. We train the network by embedding the above architecture as well as the proposed LQA module. A trainable network is used to identify the noise region, then the specific undesired activation map is embedded with the latent features of the fundus image. The lateral inhibition accelerates the differentiation ability of the correction network to focus on the artifact. We provide examples in Fig. 6 (d). As can be seen, the proposed network including both the LQA module and robust structure-perceiving loss function is able to better remove outlier artifacts and improve the generation performance. The results in Table I also show that the proposed LQA and RSA modules can achieve reasonable and obvious improvements.

Furthermore, to demonstrate the validity of the multi-scale framework, we also investigate the effect of removing the coarse-to-fine supervision without embedding intermediate results into the model. In Fig. 7, the fundus images contain very tiny vessels and lesions. Since the multi-scale network is capable of providing an incremental training strategy that encourages a robust and stable convergence procedure,

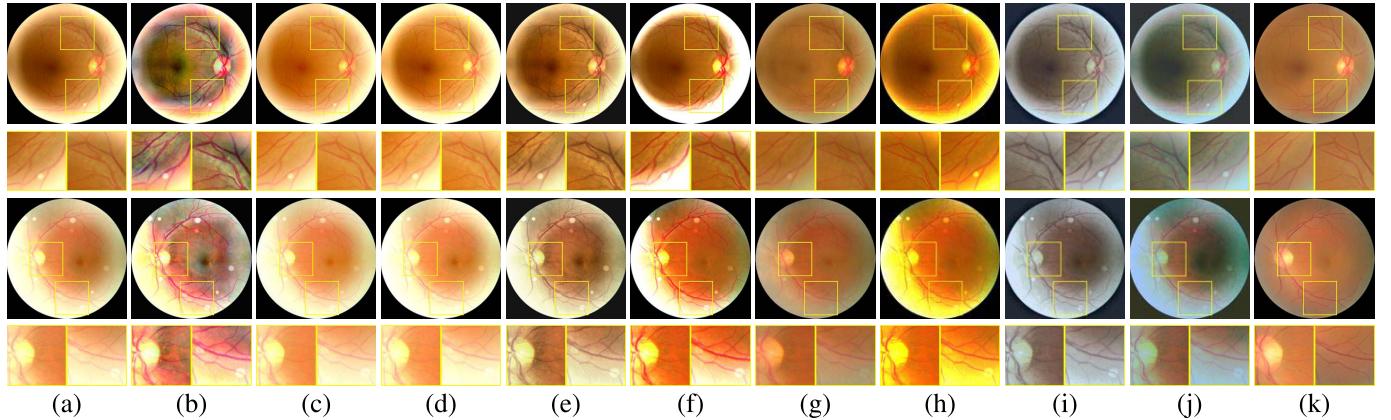


Fig. 8. Visual comparison on synthesized retinal images. The results of proposed algorithm corrects images with preserving the clinical features and suppressing the retinal artifacts. (a) Real images. (b) Setiawan *et al.* [13]. (c) Fu *et al.* [16]. (d) Guo *et al.* [17]. (e) Tian *et al.* [20]. (f) Cheng *et al.* [33]. (g) Eilertsen *et al.* [8]. (h) He *et al.* [28]. (i) Fu *et al.* [26]. (j) Li *et al.* [25]. (k) Ours.

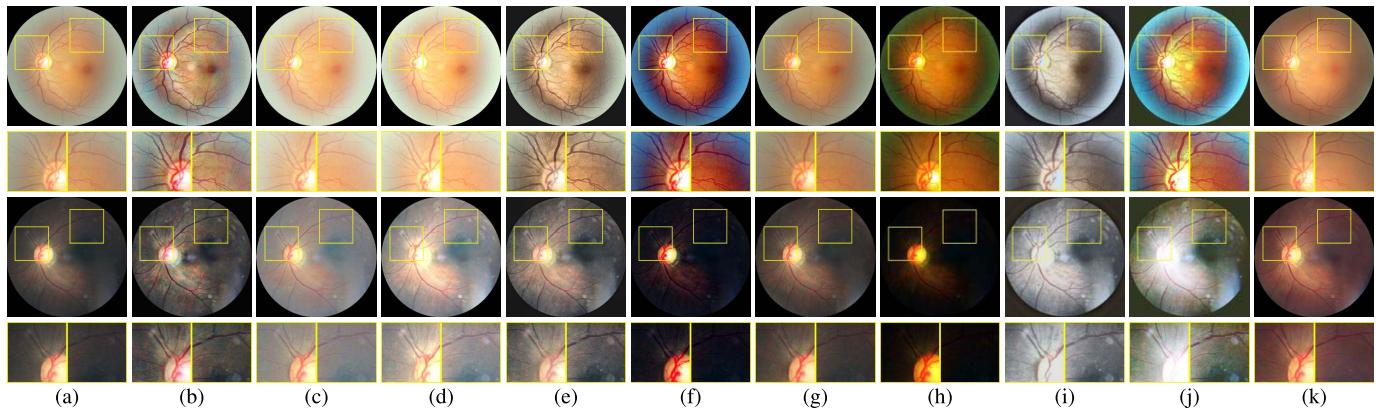


Fig. 9. Visual comparison on real clinical retinal images. The proposed algorithm corrects images while preserving clinical features and suppressing retinal artifacts. (a) Real images. (b) Setiawan *et al.* [13]. (c) Fu *et al.* [16]. (d) Guo *et al.* [17]. (e) Tian *et al.* [20]. (f) Cheng *et al.* [33]. (g) Eilertsen *et al.* [8]. (h) He *et al.* [28]. (i) Fu *et al.* [26]. (j) Li *et al.* [25]. (k) Ours.

it provides sharper correction results than the single-scale model. We also provide a quantitative comparison in Table I, which demonstrates that the multi-scale framework is able to provide considerable performance promotion.

E. Comparison With State-of-the-Art Methods

In this section, we provide qualitative and quantitative comparisons with nine state-of-the-art methods [8], [16], [17], [20], [25], [26], [28], [33], [50] to demonstrate the advantages of the proposed cofe-Net. We first conduct an experiment on the degraded fundus images for quantitative analysis. Then, to demonstrate that our algorithm is also equally applicable to real clinical images, these methods are also used in real clinical image correction tasks. As shown in Fig. 8 and Table II, the proposed cofe-Net achieves comparable performance, performing particularly favorably against the existing methods on the images with multiple degradation. We also provide a visual comparison on real images in Fig. 9. The proposed approach effectively corrects the fundus images with relatively sharp and clean details. We observe that the methods [17], [33], [50] fail to remove the noticeable undesired retinal artifacts. Specifically, medical image correction aims to restore a high-quality

TABLE II
QUANTITATIVE COMPARISON WITH STATE-OF-THE-ART METHODS.
WE COMPUTE THE AVERAGE PSNR AND SSIM
ON TWO TEST SETS.

Dataset	DRIVE [49]		Kaggle [48]	
	PSNR	SSIM	PSNR	SSIM
Cheng <i>et al.</i> [33]	14.97	0.648	15.02	0.845
Guo <i>et al.</i> [17]	14.10	0.703	13.54	0.868
Fu <i>et al.</i> [16]	15.56	0.722	14.66	0.882
Tian <i>et al.</i> [20]	15.42	0.721	14.71	0.664
Zuiderveld [50]	15.93	0.740	14.05	0.716
Eilertsen <i>et al.</i> [8]	19.01	0.755	18.40	0.841
He <i>et al.</i> [28]	15.78	0.559	15.56	0.749
Fu <i>et al.</i> [25]	10.19	0.580	9.76	0.564
Li <i>et al.</i> [26]	9.51	0.543	9.47	0.547
Our cofe-Net	21.24	0.758	20.51	0.885

image by suppressing noise while at the same time preserving essential pathological characteristics. The handcrafted feature-based methods only consider the general image characteristics, which easily results in ignoring specific artifacts in medical images. In contrast, our method is capable of enhancing fundus images as well as removing these additional artifacts, which accelerates the subsequent diagnosis task.

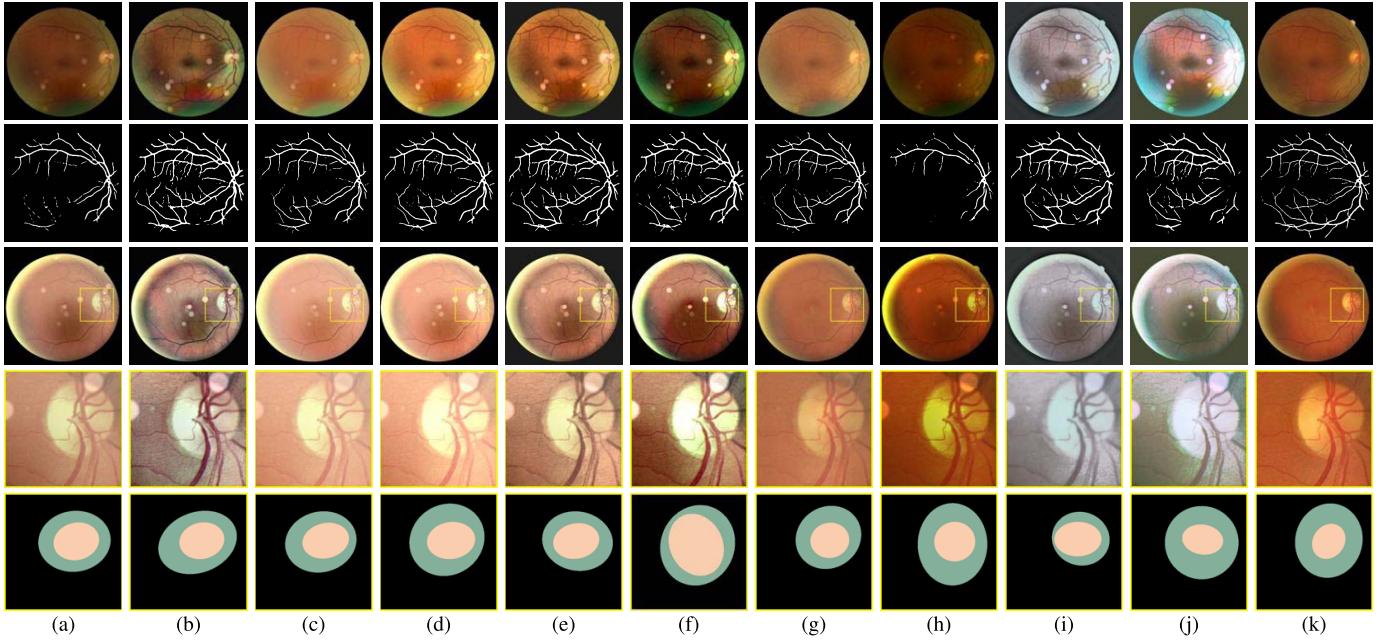


Fig. 10. Visual comparison of vessel segmentation on synthesized images. The proposed method can improve the clinical analysis on vessel segmentation and disc/cup detection. (a) Low quality images. (b) Setiawan *et al.* [13]. (c) Fu *et al.* [16]. (d) Guo *et al.* [17]. (e) Tian *et al.* [20]. (f) Cheng *et al.* [33]. (g) Eilertsen *et al.* [8]. (h) He *et al.* [28]. (i) Fu *et al.* [26]. (j) Li *et al.* [25]. (k) Ours.

TABLE III

PERCENTAGE SCALE OF ENHANCEMENT METHOD IN USER STUDY.

Method	Image Quality	Lesions Quality
Setiawan [13]	2.17	2.74
Guo <i>et al.</i> [17]	7.41	5.86
Fu <i>et al.</i> [16]	5.48	2.04
Tian <i>et al.</i> [20]	8.12	16.88
Cheng <i>et al.</i> [33]	9.44	11.93
Eilertsen <i>et al.</i> [8]	21.42	14.82
He <i>et al.</i> [28]	1.73	1.16
Fu <i>et al.</i> [26]	2.25	2.58
Li <i>et al.</i> [25]	1.53	2.17
Our cofe-Net	40.45	39.82

We further provide comparisons on real fundus images, employing user studies to quantitatively evaluate the state-of-the-art methods and our method. We use a paired comparison strategy to evaluate the quality of the medical images. For each test, we provide 200 pairs of fundus images corrected using different enhancement methods. We display these results in random order and ask the participants (including ophthalmologists/clinicians, and students with previous fundus image analysis experience) to rank the results based on the instructions. Note that both the suppression of artifacts and preservation of lesions are taken into account. Finally, we collect 65 valid responses. The percentages of votes for each method are shown in Table III. As can be seen, the proposed fundus image enhancement method receives the most votes for best corrected results.

F. Clinical Image Analysis and Applications

Since medical image correction models should be applied to real clinical tasks, to demonstrate the effectiveness of the

TABLE IV

VESSEL SEGMENTATION AND DISC/CUP DETECTION EVALUATION.

	Vessel Seg.			Disc/Cup Seg.	
	AUC	Acc	Sen	F-score	J-score
Without enhance	0.924	0.943	0.532	0.735	0.746
Cheng <i>et al.</i> [33]	0.950	0.940	0.586	0.713	0.705
Guo <i>et al.</i> [17]	0.953	0.951	0.541	0.829	0.835
Fu <i>et al.</i> [16]	0.954	0.952	0.561	0.845	0.852
Tian <i>et al.</i> [20]	0.941	0.945	0.575	0.822	0.803
Setiawan [13]	0.957	0.943	0.591	0.861	0.858
Eilertsen <i>et al.</i> [8]	0.954	0.950	0.585	0.884	0.852
He <i>et al.</i> [28]	0.955	0.949	0.521	0.872	0.861
Fu <i>et al.</i> [26]	0.943	0.945	0.524	0.724	0.701
Li <i>et al.</i> [25]	0.940	0.945	0.534	0.839	0.796
with enhance	0.957	0.951	0.600	0.890	0.863

proposed method, we conduct additional experiments on clinical image analysis tasks, including the vessel segmentation and optic disc/cup detection. The DRIVE [49] and REFUGE [51] datasets are selected for evaluation, which provide the annotations of vessel and disc/cup regions. CE-Net [7] and M-Net [52] are employed as segmentation baselines.

1) *Vessel Segmentation*: For the vessel segmentation task, the images from DRIVE [49] are used to generate 100 low-quality images for quantitative assessment. The vessel segmentation results are shown in Fig. 10 and Table IV. As can be seen, CE-Net fails to obtain high performance on low-quality clinical images. In contrast, a strong vessel structure can be extracted after applying the proposed correction method. It obtains better performance on corrected images.

2) *Optic Disc/Cup Detection*: We also conduct an experiment on disc and cup detection. We simulate 400 degraded images with different settings from the REFUGE test set [51]. M-Net [52] is used to segment the optic disc/cup. We report

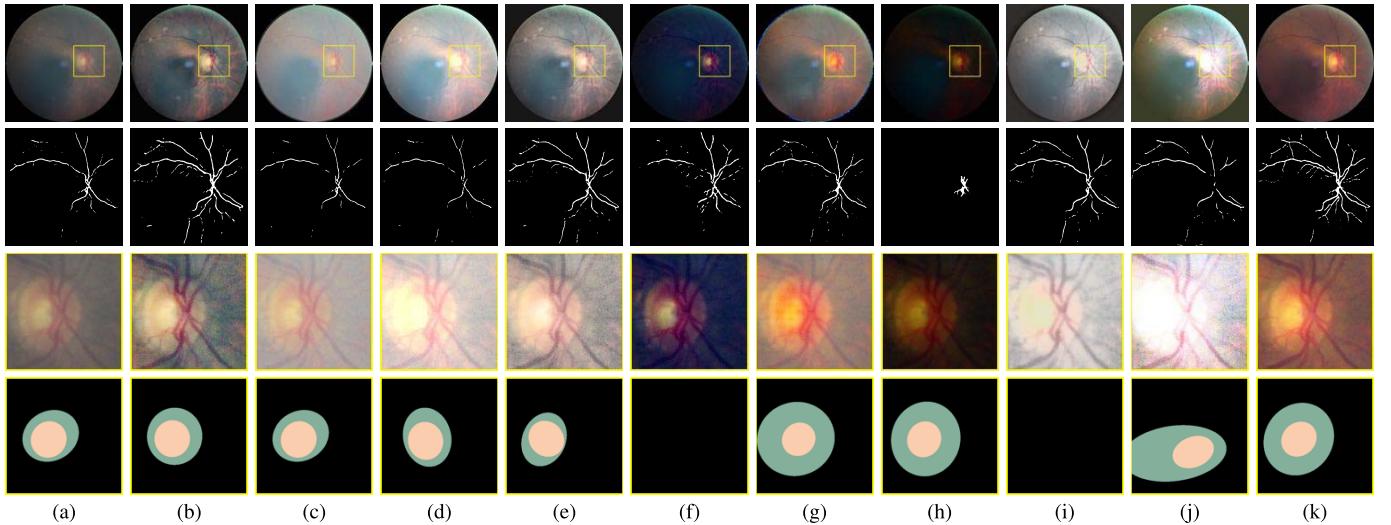


Fig. 11. Vessel segmentation and optic disc/cup detection on real images. (a) Low-quality images. (b) Setiawan *et al.* [13]. (c) Fu *et al.* [16]. (d) Guo *et al.* [17]. (e) Tian *et al.* [20]. (f) Cheng *et al.* [33]. (g) Eilertsen *et al.* [8]. (h) He *et al.* [28]. (i) Fu *et al.* [26]. (j) Li *et al.* [25]. (k) Ours.

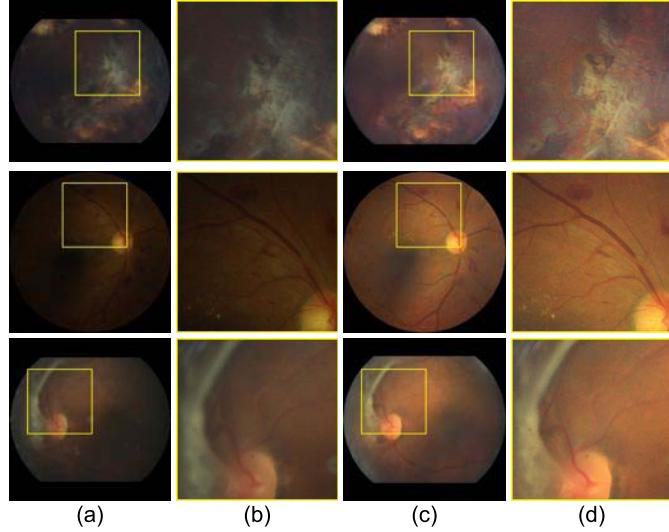


Fig. 12. Lesion analysis on real clinical images. (a) Real image without correction. (b) Zoom-in of (a). (c) The image with our correction. (d) Zoom-in of (c). Our proposed method is capable of correcting images and preserving the lesions for disease analysis.

the F-scores and J-scores in Table IV and quantitative results in Fig. 10. Our method recovers a high dynamic range in the low-quality images, boosting the discriminative representations of the disc/cup.

3) Real Clinical Fundus Image: To demonstrate its clinical value, we also validate the proposed correction method for vessel segmentation and optic disc/cup detection on real fundus images. As shown in Fig. 11, our method produces legitimate corrected real images with clear clinical structures, which can successfully be processed by CE-Net [7]. In addition, the optic disc/cup regions are also enhanced to be more cognizable.

4) Real Lesion Analysis: To further analyze the proposed model on small lesions, we evaluate it on enhanced images from the Kaggle [48] dataset. We design the degradation framework to model the artificial interference caused by

ophthalmoscope imaging procedures. Our algorithm can discriminate between undesired disturbances and specific lesions. It is capable of correcting images and preserving the lesions for subsequent disease diagnosis. In Fig. 12, we further show the enhanced results of retinal fundus images with some lesions, such as hemorrhages and cataracts. Our method is robust to various diseases and retains the lesions, affecting the transparency of the eye's lens.

VI. CONCLUSION

In this article, we have proposed a clinically oriented fundus enhancement network, named cofe-Net, to correct low-quality fundus image while preserving accurate lesion areas and retinal structures. Furthermore, a complete degradation model has also been introduced to generate adequate training image pairs. Experiments support our insight into the problems of fundus image correction and degradation factor modeling. Our cofe-Net can boost the performance for different clinical tasks, such as vessel segmentation and disc/cup detection. Our method can also assist ophthalmologists in ocular disease diagnosis through retinal fundus image observation and analysis, while also being beneficial to automated image analysis systems.

REFERENCES

- [1] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, Dec. 2010.
- [2] U. Schmidt-Erfurth, A. Sadeghipour, B. S. Gerendas, S. M. Waldstein, and H. Bogunović, "Artificial intelligence in retina," *Prog. Retinal Eye Res.*, vol. 67, pp. 1–29, Nov. 2018.
- [3] S. Philip, "The impact of the health technology board for Scotland's grading model on referrals to ophthalmology services," *Brit. J. Ophthalmol.*, vol. 89, no. 7, pp. 891–896, Jul. 2005.
- [4] E. Peli and T. Peli, "Restoration of retinal images obtained through cataracts," *IEEE Trans. Med. Imag.*, vol. 8, no. 4, pp. 401–406, 1989.
- [5] H. Fu *et al.*, "Evaluation of retinal image quality assessment networks in different color-spaces," in *Proc. MICCAI*, 2019, pp. 48–56.
- [6] Y. Cheng *et al.*, "Adversarial exposure attack on diabetic retinopathy imagery," 2020, *arXiv:2009.09231*. [Online]. Available: <http://arxiv.org/abs/2009.09231>

- [7] Z. Gu *et al.*, “CE-net: Context encoder network for 2D medical image segmentation,” *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.
- [8] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, “HDR image reconstruction from a single exposure using deep CNNs,” *ACM Trans. Graph.*, vol. 36, no. 6, p. 178, 2017.
- [9] X. Yang, K. Xu, Y. Song, Q. Zhang, X. Wei, and R. W. H. Lau, “Image correction via deep reciprocating HDR transformation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1798–1807.
- [10] W. Ren *et al.*, “Low-light image enhancement via a deep hybrid network,” *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4364–4375, Sep. 2019.
- [11] M. Foracchia, E. Grisan, and A. Ruggeri, “Luminosity and contrast normalization in retinal images,” *Med. Image Anal.*, vol. 9, no. 3, pp. 179–190, Jun. 2005.
- [12] S. J. Hwang, A. Kapoor, and S. B. Kang, “Context-based automatic local image enhancement,” in *Proc. ECCV*, 2012, pp. 569–582.
- [13] A. W. Setiawan, T. R. Mengko, O. S. Santoso, and A. B. Suksmono, “Color retinal image enhancement using CLAHE,” in *Proc. Int. Conf. ICT Smart Soc.*, Jun. 2013, pp. 1–3.
- [14] M. K. Ng and W. Wang, “A total variation model for retinex,” *SIAM J. Imag. Sci.*, vol. 4, no. 1, pp. 345–365, Jan. 2011.
- [15] W. W. A. M. K. Ng, “A nonlocal total variation model for image decomposition: Illumination and reflectance,” *Numer. Math., Theory, Methods Appl.*, vol. 7, no. 3, pp. 334–355, Jun. 2014.
- [16] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, “A weighted variational model for simultaneous reflectance and illumination estimation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2782–2790.
- [17] X. Guo, Y. Li, and H. Ling, “LIME: Low-light image enhancement via illumination map estimation,” *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.
- [18] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, “Efficient contrast enhancement using adaptive gamma correction with weighting distribution,” *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1032–1041, Mar. 2013.
- [19] M. Zhou, K. Jin, S. Wang, J. Ye, and D. Qian, “Color retinal image enhancement based on luminosity and contrast adjustment,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 3, pp. 521–527, Mar. 2018.
- [20] Q.-C. Tian and L. D. Cohen, “Global and local contrast adaptive enhancement for non-uniform illumination color images,” in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 3023–3030.
- [21] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, “An iterative regularization method for total variation-based image restoration,” *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 460–489, Jan. 2005.
- [22] X. Lu, Y. Yuan, and P. Yan, “Image super-resolution via double sparsity regularized manifold learning,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 12, pp. 2022–2033, Dec. 2013.
- [23] D. Zoran and Y. Weiss, “From learning models of natural image patches to whole image restoration,” in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 479–486.
- [24] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, “Understanding and evaluating blind deconvolution algorithms,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1964–1971.
- [25] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, “Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior,” *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5664–5677, Dec. 2016.
- [26] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding, “A retinex-based enhancing approach for single underwater image,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 4572–4576.
- [27] W. Ren, X. Cao, J. Pan, X. Guo, W. Zuo, and M.-H. Yang, “Image deblurring via enhanced low-rank prior,” *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3426–3437, Jul. 2016.
- [28] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
- [29] S. Wang, J. Zheng, H.-M. Hu, and B. Li, “Naturalness preserved enhancement algorithm for non-uniform illumination images,” *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.
- [30] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, “Multiscale vessel enhancement filtering,” in *Proc. MICCAI*, 1998, pp. 130–137.
- [31] M. Liao, Y.-Q. Zhao, X.-H. Wang, and P.-S. Dai, “Retinal vessel enhancement based on multi-scale top-hat transformation and histogram fitting stretching,” *Opt. Laser Technol.*, vol. 58, pp. 56–62, Jun. 2014.
- [32] A. Baghaie, R. M. D’Souza, and Z. Yu, “Sparse and low rank decomposition based batch image alignment for speckle reduction of retinal OCT images,” in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 226–230.
- [33] J. Cheng, Z. Li, Z. Gu, H. Fu, D. W. K. Wong, and J. Liu, “Structure-preserving guided retinal image filtering and its application for optic disk analysis,” *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2536–2546, Nov. 2018.
- [34] K. G. Lore, A. Akintayo, and S. Sarkar, “LLNet: A deep autoencoder approach to natural low-light image enhancement,” *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.
- [35] K. Wei, J. Yang, Y. Fu, D. Wipf, and H. Huang, “Single image reflection removal exploiting misaligned training data and network enhancements,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8178–8187.
- [36] W. Ren *et al.*, “Gated fusion network for single image dehazing,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3253–3261.
- [37] J. Chen, C.-H. Tan, J. Hou, L.-P. Chau, and H. Li, “Robust video content alignment and compensation for rain removal in a CNN framework,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6286–6295.
- [38] C. Li, C. Guo, J. Guo, P. Han, H. Fu, and R. Cong, “PDR-net: Perception-inspired single image dehazing network with refinement,” *IEEE Trans. Multimedia*, vol. 22, no. 3, pp. 704–716, Mar. 2020.
- [39] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, “Deep semantic face deblurring,” in *Proc. CVPR*, 2018, pp. 8260–8269.
- [40] X. Xu, J. Pan, Y.-J. Zhang, and M.-H. Yang, “Motion blur kernel estimation via deep learning,” *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 194–205, Jan. 2018.
- [41] G. Tang, L. Zhao, R. Jiang, and X. Zhang, “Single image dehazing via lightweight multi-scale networks,” in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 154–169.
- [42] F. Lv, F. Lu, J. Wu, and C. Lim, “MBLLEN: Low-light image/video enhancement using CNNs,” in *Proc. BMVC*, 2018, pp. 220–228.
- [43] C. Li, H. Fu, R. Cong, Z. Li, and Q. Xu, “Nul-Go: Recursive non-local encoder-decoder network for retinal image non-uniform illumination removal,” in *Proc. ACM Multimedia*, 2020, pp. 1478–1487.
- [44] H. Talebi and P. Milanfar, “Learned perceptual image enhancement,” in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, May 2018, pp. 1–13.
- [45] R. Liu, L. Ma, Y. Wang, and L. Zhang, “Learning converged propagations with deep prior ensemble for image enhancement,” *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1528–1543, Mar. 2019.
- [46] H. Zhao, B. Yang, L. Cao, and H. Li, “Data-driven enhancement of blurry retinal images via generative adversarial networks,” in *Proc. MICCAI*, 2019, pp. 75–83.
- [47] U. Upadhyay and S. P. Awate, “A mixed-supervision multilevel GAN framework for image quality enhancement,” in *Proc. MICCAI*, 2019, pp. 556–564.
- [48] Kaggle Diabetic Retinopathy Detection. Accessed: Dec. 8, 2020. [Online]. Available: <https://www.kaggle.com/c/diabetic-retinopathy-detection/data>
- [49] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.
- [50] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” in *Graphics Gems IV*. New York, NY, USA: Academic, 1994, pp. 474–485.
- [51] J. I. Orlando *et al.*, “REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs,” *Med. Image Anal.*, vol. 59, Jan. 2020, Art. no. 101570.
- [52] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, “Joint optic disc and cup segmentation based on multi-label deep network and polar transformation,” *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1597–1605, Jul. 2018.