

# Modeling and Enhancing Low-quality Retinal Fundus Images

Ziyi Shen, *Member, IEEE*, Huazhu Fu, *Senior Member, IEEE*, Jianbing Shen, *Senior Member, IEEE*, and Ling Shao, *Senior Member, IEEE*

arXiv:2005.05594v2 [eess.IV] 23 Jul 2020

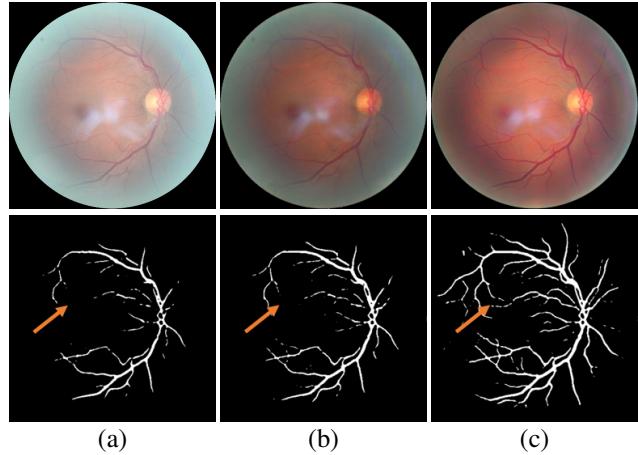
**Abstract**— Retinal fundus images are widely used for clinical screening and diagnosis of eye diseases. However, fundus images captured by operators with various levels of experiences have a large variation in quality. Low-quality fundus images increase the uncertainty in clinical observation and lead to a risk of misdiagnosis. Due to the special optical beam of fundus imaging and retinal structure, the natural image enhancement methods cannot be utilized directly. In this paper, we first analyze the ophthalmoscope imaging system and model the reliable degradation of major inferior-quality factors, including uneven illumination, blur, and artifacts. Then, based on the degradation model, a clinical-oriented fundus enhancement network (cofe-Net) is proposed to suppress the global degradation factors, and simultaneously preserve anatomical retinal structures and pathological characteristics for clinical observation and analysis. Experiments on both synthetic and real fundus 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, medical image synthesis, deep neural network.

## I. INTRODUCTION

Due to the safety and cost-effectiveness at observing eye abnormalities, retinal fundus images are widely used for clinical screening and diagnosis of ocular diseases by the ophthalmologists and computer-aided diagnosis systems [1]–[4]. However, fundus images tend to experience large variations in quality. As they are often acquired under different lighting environments, using various cameras, and by distinct operators with varying levels of experience. A screening study of 5575 patients found that about 12% of fundus images are not of adequate quality to be readable by the ophthalmologists [5]. Major causes of low quality in retinal fundus images include uneven illumination, blurs, and artifacts, which not only prevent reliable diagnosis by ophthalmologists, but also disturb the performance of automated image analyzing systems [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]).

Co-corresponding author: Jianbing Shen and Huazhu Fu



**Fig. 1.** The illustration of fundus image correction. (a) The low-quality fundus image. (b) Enhanced result by [8]. (c) Enhanced result by our method. The first row is 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 and disc region).

Recently, general image enhancement have achieved state-of-the-art performances, especially with the development of deep learning techniques [9]–[11]. However, different from general images, retinal fundus images have the special ophthalmoscope imaging process and retinal anatomical structures for clinical diagnosis, which introduce various additional challenges. First, the retina can not be illuminated internally, both 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 [12]. 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, cause them to be easily confused with artifacts and noises. These issues mean that fundus image correction method must to be able to both suppress the undesired low-quality factors and preserve the pathological characteristics simultaneously, which general enhancement techniques could not be satisfied. 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 the artifacts, and some vessels are miss-segmented by automated system.

To address these issues, in this paper, we design a degradation model that approximates major factors of low-quality fundus images, including light transmission disturbance, imaging blurry, and retinal artifacts. Then, a clinical-oriented fundus enhancement method is proposed to suppress the local outliers and undesired artifacts, while at the same time preserving the atomical 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 capable of correcting fundus images with more accurate structures and suppressing local defects. A corrected fundus example of our method is shown in Fig. 1 (c). The main contributions of this paper are summarized as follows:

- 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 in a real clinic. All degradation models are designed based on several imaging stages.
- A novel clinical-oriented fundus enhancement network (cofe-Net) is developed to correct the low-quality fundus images for clinical observation and analysis. Our cofe-Net preserves the atomical retinal structures of fundus image by using RSA module and suppresses the undesired artifacts based on a LQA module.
- We demonstrate that the fundus correction can boost the performances of clinical analysis systems, 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.

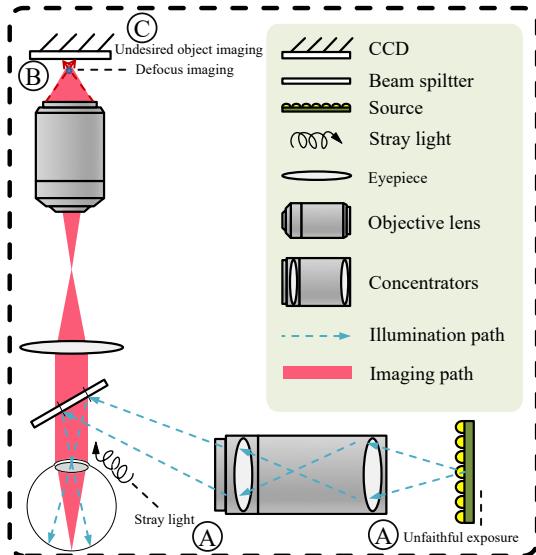
## II. RELATED WORK

In this section, we summarize current image correction techniques and discuss the algorithms that have been especially applied to fundus images. Several methods [13]–[15] exploit the image contrast normalization and contrast limited adaptive histogram equalization (CLAHE) techniques to restore an image. For example, Setiawan et al. [16] applied CLAHE to fundus image enhancement specifically. Instead of simply considering the color and texture information, some algorithms [17]–[19] decompose the reflection and illumination, achieving image enhancement and correction by estimating the solution in an alternate minimization scheme. Guo et al. [20] further proposed to refine the illumination map and enhance low-light images. These models have been extended to an integrated scheme incorporating gamma correction [21] and CLAHE

for fundus image luminance and contrast adaption [22], [23]. While these algorithms based on a bottom-up framework are effective, the optimal solution relies heavily on global image statistics and mapping functions, ignores discriminative features, which may introduce undesired artifacts and distortion.

Along another line, knowledge learning based methods have also been developed for the image correction task, which utilize 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 [24], [25], distribution fitting algorithms [26], [27], and latent structure-driven methods [28]–[30], which have also been specifically applied to fundus images [31]–[33]. Cheng et al. [34] proposed a structure-preserving guided filter to achieve the image enhancement. These algorithms commonly constrain the optimal solution using a regularization scheme to solve the non-convex problem. This may incur a heavy computational cost, limiting their applicability to clinical applications.

Recently, due to the powerful image representation ability, deep learning techniques have been widely used in computer vision. This has enabled the rapid advancement of reconstruction techniques, making them much better equipped to address various challenging scenarios, such as low-light image enhancement [11], [35], [36], dehazing and deraining [37], [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. Eilertsen et al. [8] proposed to solve the High Dynamic Range (HDR) 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. The reciprocating transformation pipeline [9] was introduced to preserve more details from the degraded images. 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 [43] 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 solution by simply minimizing the content loss. As the latent prior/feature-driven schemes enable the generative model to directly express the latent details, in allusion to this superiority, focusing on the images stocked with explicit prior information, Liu et al. [44] proposed a deep prior ensemble and integrated knowledge-driven cues for natural image enhancement. Different from the non-convex optimization framework, the deep learning method is a heuristic pattern that relies heavily on a great deal of training data. Due to the particular pathological characteristics and disease markers, the general CNN-based models often do not perform well on medical samples, especially fundus images. To deblur and enhance clinical images, Zhao et al. [45] and Liu et al. [46] applied the adversarial loss. However, despite their computational efficiency, 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.

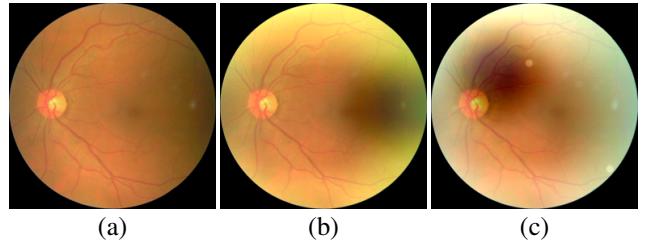


**Fig. 2. Schematic diagram of ophthalmoscope imaging system.** It contains a illumination system and an imaging system. The clinic imaging process in the real scene undergoes several interference such as exposure, imaging outlier which introduce unexpected drawbacks on the retinal fundus images and result in the image degradation. Besides, the undesired objects imaging also is a crucial factor which interferes with the image quality and impacts the diagnostic process.

### III. RETINAL FUNDUS IMAGE DEGRADATION MODEL

Clinical image collection in a complex environment using an ophthalmoscope often encounters several types of interference, which is introduced in the optical feed-forward system. As shown in Fig. 2, light transmission disturbance is often caused by exposure issue. Due to the interspace between the eye and camera, stray light may enter into the ophthalmoscope, mix with the lighting source and result in a uneven exposure. This also affects the faithful setting of the programmed exposure, which causes global over-/under-exposure. In addition, imaging blurry simulating the 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 reformulation representation for the interference that occurs during the collection of fundus images. Our degradation model could be used to not only support the typical scheme in fundus propagation model, but also to synthesize a pairwise fundus quality dataset for subsequent research. We summarize the interference in terms of three factors, including *light transmission disturbance, imaging blur, and retinal artifacts*.

**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 is still an element that impacts the illumination. Unstable stray light which affects the camera configuration (light source), may result in under-/over- exposure during the image collection. Furthermore, the personal situation and manually mydriasis also cause disturbance during the imaging procedure. For local factors, the sensitivity of specific regions



**Fig. 3. Synthetic degraded fundus images.** (a) Clear image. (b) Simulated sample undergoing light transmission disturbance. (c) Simulated sample undergoing light transmission disturbance, imaging blurry, and retinal artifacts. Zoom-in for details.

on the image plane exist difference, the variant sensitization ability often cause uneven illumination in an image. It is due to the clearance between fundus and ophthalmoscope, an initiative light leak phenomenon caused by the patient, an inappropriate exposure imported to the imaging plan with a variant distance to optical axis will result in an uneven illumination. In addition, according to the equipment design of the ophthalmoscope system, the diverse lens apertures and the embedded optical compensation mechanism limit the amount of light as well as affect the dynamic range of the fundus images. Here, we simulate the degradation as an aggregation model. The global factor is modeled by contrast, brightness and saturation interference, while the local factor is defined as an extra non-uniform illumination on the fundus imaging. To formulate the light degradation model, given a ground truth image  $\mathbf{x}$ , a degraded image  $\mathbf{x}'$  is defined as:

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

where  $\alpha$ ,  $\beta$  and  $s$  refer to the contrast factor, brightness and saturation, respectively. We use a clipping function  $\text{clip}(\cdot; s)$  to mimic 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  $J_{ij} = n_l|_{(i-a)^2+(j-b)^2 < r_L^2}$ . A Gaussian kernel  $G_L$  is then applied to ensure luminance smoothness. Fig. 3 (b) gives a synthetic example with over-exposure and uneven illumination by our degradation model.

**Imaging Blurry:** During fundus imaging procedure, in addition to the program setting, human operator error can also be introduced. Since the ophthalmoscope is applied on patients, the distance between fundus and object plane is a random variable; a wrong setting of the focal length of the optical system during the funduscopic, 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 the imaging blurry, we define the defocus blur:

$$\mathbf{x}' = \mathbf{x} * G_B(r_B, \sigma_B) + n, \quad (2)$$

where  $G_B$  is a Gaussian filter with a radius of  $r_B$  and spatial constant  $\sigma_B$ , and  $n$  denotes the additive noise. The degraded fundus image  $\mathbf{x}'$  is generated via a convolution operation between the ground truth image  $\mathbf{x}$  and a Gaussian filter.

Especially, the blur radius  $r$  can be defined by:

$$r = \frac{F\nu_0 - D(F - \nu_0)}{Df}, \quad (3)$$

where  $F$  is the focal length of the optical lens system, and  $f$  is its f-number.  $D$  and  $\nu_0$  denote the object and image distance.

**Retinal Artifacts:** The above formulated models are similar to an internal based, which is caused by the imaging system itself. However, additional degradation may be introduced if the imaging is conducted under poor conditions. Unlike the generic additive and multiplicative noise, dust and grains attaching on the lens as well image on the plane and yield undesired blurred images that affecting the fundus image quality and following diagnosis. We define the speckles interfering as a retinal artifact, which can be modeled by a multi-step imaging procedure:

$$\mathbf{x}' = \mathbf{x} + \sum_i G_R(r_i/4, \sigma_i) * \mathbf{o}_i. \quad (4)$$

In the context of random light fields,  $r_i$  and  $\sigma_i$  are defined as the radius and variance of a undesired object  $i$  imaged on the plane, which can be calculate using (3).  $\mathbf{o}_i$  is the luminance bias. We again define a Gaussian filter  $G_R$  to simulate the defocused imaging of an undesired object, due to its specific imaging distance (attaching on lens). In this way, retinal artifacts is defined as undesired imaging items.

To generate the real clinical fundus images, we provide a specific example containing various artifacts as shown in Fig. 3 (c). The proposed degradation model is in particularly well-designed to simulate the real funduscopic examination procedure. Numerous data-dependent reconstruction algorithms attempt to learn a degrade-latent mapping between image pairs using supervised learning approaches, especially for clinical imaging tasks which rely heavily on the retinal image quality. For fundus image correction, the high-quality images can be processed by the above-mentioned models with randomly perturbed variables to obtain their degraded counterparts.

#### IV. CLINICAL-ORIENTED FUNDUS ENHANCEMENT NETWORK

For medical image processing, there are certain crucial cues that require special attention. Thus not only is it necessary to preserve the fine structures in 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 clinical-oriented fundus enhancement network (cofe-Net) to solve the fundus image correction problem, which contains three modules as shown in Fig. 4. Given a clinical 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 a correct network is well designed to fuse these information with the latent features of 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 model  $\mathcal{G}$ . Let  $\mathbf{I} \in \mathbb{R}^{w \times h \times c}$  and  $\mathbf{L} \in \mathbb{R}^{w \times h \times c}$  denote the input low-quality image and the latent image, respectively. The correction procedure can be formulated as:

$$\mathbf{L} = \mathcal{G}(\mathbf{I}; \mathbf{W}_{\mathcal{G}}), \quad (5)$$

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

$$\mathcal{L}_c = \|\mathbf{G}_{\mathcal{L}} - \mathcal{G}(\mathbf{I}, \mathbf{W}_{\mathcal{G}})\|_2^2, \quad (6)$$

where  $\mathbf{G}_{\mathcal{L}}$  is the ground truth 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 an equal pixel-wise framework. Therefore, the vanilla model simply pays equal attention to global image, without focusing on meaningful regions (e.g., vessels, disc/cup regions) into special consideration, which may fail to remove the local artifacts that interfere, or preserve real structures.

To address this issue, we introduce a supervised activation mechanism, LQA module, to perceive the undesired low-quality factors in clinical images. Our LQA module  $\mathcal{M}$  is built upon a convolutional encoder-decoder network architecture, as shown in Fig. 4, and produces a vector  $\mathbf{m} \in [0, 1]$ . This differentiable neural activation operator can be defined as:

$$\mathbf{m} = \sigma(\mathcal{M}(\mathbf{I}, \mathbf{W}_{\mathcal{M}})), \quad (7)$$

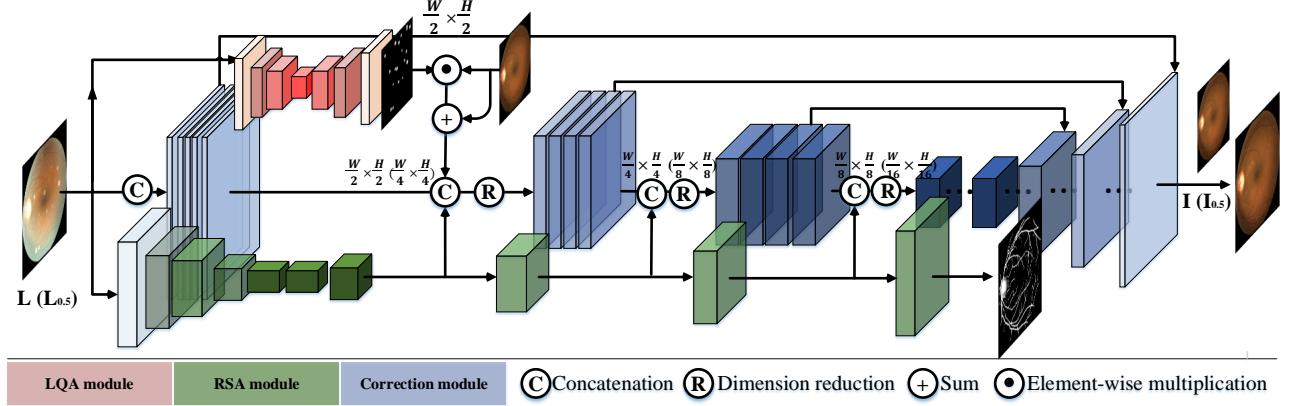
where  $\sigma$  denotes the non-linear activation function that normalizes the feature. The proposed correction framework is designed to perceive local artifacts via a LQA branch with parameters  $\mathbf{W}_{\mathcal{M}}$ , and embed the corresponding map to guide the correction. Let  $\mathbf{F}_{\mathcal{L}}$  denote the low-level features extracted from the correction network. We define this LQA map as:  $\sigma(\mathcal{M}(\mathbf{L}, \mathbf{W}_{\mathcal{M}})) \odot \mathbf{F}_{\mathcal{L}}$ , where  $\odot$  is an element-wise multiplication. The high-level features extracted from LQA are capable of fusing the low-features into the correction model. Therefore, the embedded framework is able to encourage a more efficient representation and achieve a higher learning ability. Here, we define the pixel-wise  $\mathcal{L}_2$  loss as a constraint for training the low-quality activation neural network:

$$\mathcal{L}_m = \|\mathbf{G}_m - \mathbf{m}\|_2^2, \quad (8)$$

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

$$\mathcal{L}_p = \sigma(\mathcal{M}(\mathbf{I}, \mathbf{W}_{\mathcal{M}})) \odot \|\mathbf{G}_m - \mathcal{G}(\mathbf{I}; \mathbf{W}_{\mathcal{G}})\|_2^2. \quad (9)$$

The LQA module can assist the correction by focusing on the undesired artifacts.



**Fig. 4. Architecture of the proposed structure perceiving enhancement neural 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 correction module. The LQA module activates and embeds low quality image regions into the correction module at an early stage. The RSA module continuously injects retinal structural information to the correction network by feature map concatenation at each scale. Leveraging these additional information, the correction module predicts high-quality images in an encoder-decoder fashion.

For this clinical fundus correction task, the proposed LQA module aims to identify retinal artifacts in the fundus image. Since undesired objects (e.g., dust and grain) on the object lens often present analogous images due to their similar shape and the inherent object/image distance, 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 of our LQA module as a  $0.5 \times$  image to reduce the computational load. As shown in Fig. 4, the encoder-decoder network contains three convolutional layers with  $3 \times 3$  kernels followed by a  $2 \times$  max-pooling layer for downsampling. Then, the perceived information is integrated and represented by the three corresponding convolutional layers. We embed a transposed convolutional layer with  $4 \times 4$  kernels for each scale to upsample the feature maps. Each filter is followed by a ReLU function as the non-linear activation. Finally, another  $1 \times 1$  convolutional layer is utilized to reshape the feature maps and predict the LQA map.

### B. Retinal Structure Activation module

Natural image processing tends to reconstruct images with more diverse context in terms of human perception, often containing exaggerated or fake details. However, in some cases, this may not have a positive impact especially in fundus images, which are used as the foundations for clinical diagnosis, since incorrectly generated components can heavily skew the pathological features. It is essential to develop a new framework that can preserve more realistic and convincing content in these images.

To address this issue, we propose to extract the deep clinical features from fundus images, and inspire the generative procedure with a directional guidance. A retinal structure activation (RSA) module is utilized to cover features of the retinal structure from a lower level to a deeper level, which is equipped with another encoder-decoder fashion as shown in Fig. 4, where a MSE loss is utilized to constrain this structure activation module. Then, the non-linear RSA maps obtained

from the activation function in multiple depths are explicitly utilized to accelerate the main correction model via dense connections. It can be formulated as:

$$\mathbf{S}^l = \sigma^n (\mathcal{T}_S^n ([\mathbf{F}_S^n(\mathbf{I}), \mathbf{F}_G^n(\mathbf{I})])), \quad (10)$$

where  $\mathbf{F}_S^n(\mathbf{I})$  and  $\mathbf{F}_G^n(\mathbf{I})$  denote the features extracted from the  $n^{th}$  scale of the RSA module and correction network, respectively.  $[.,.]$  is a concatenation operator, and the RSA map is fused via a learnable non-linear transformation filter  $\sigma^n (\mathcal{T}_S^n (.) )$ . 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 correction framework. Thus, they are able to provide the network with a stronger representational ability. In this way, the key components of the fundus image are encoded in the hidden states of the correcting network, considerably mitigating the problems of content vanishing and singular positioning.

As illustrated in Fig. 4, the encoder-decode RSA module first encodes the input image using a pre-trained ResNet-34. The convincing and efficient 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.

### C. Image Correction Module

With the assistance of the LQA and RSA modules, we design the correction network  $\mathcal{G}$  as a multi-scale framework to achieve the correction of clinical fundus images, which progressively generates images in a coarse-to-fine fashion. In this way, we propose to train a single 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, each scale consists of three residual blocks in each level of the encoder-decoder network. We first encode the image, inferring

features by identifying the critical information by densely embedding (concatenating) the knowledge learned from the RSA module. Here note that the features aggregated at each level are integrated and their dimensions are reduced by a convolutional layer. Then, a symmetric decoder module 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. In this correction module, we again apply a stride of 2 to the convolutional layer and introduce a transposed convolutional layer to achieve upsampling and downsampling throughout the whole model. Specifically, since the input of each scale with six channels, comparing with the input of the second scale which is comprised of upsampled results from the lower scale and the original image, we in particular copy the input image for the first scale with a factor of  $0.5 \times$  to satisfy the same channels which are fed into two scales. Finally, the overall loss function for the proposed architecture can be defined as:

$$\mathcal{L} = \mathcal{L}_m + \sum_{k=1}^K (\lambda_p \mathcal{L}_p^k + \lambda_c \mathcal{L}_c^k), \quad (11)$$

where  $\lambda_c$  and  $\lambda_p$  are the weights that balance the generation loss and LQA loss during training.  $K$  denotes the index of each scale. Here we set  $\lambda_c = 1$ ,  $\lambda_p = 10$ . These penalty terms are employed to stabilize the training process and ensure the corrected image is visually reasonable.

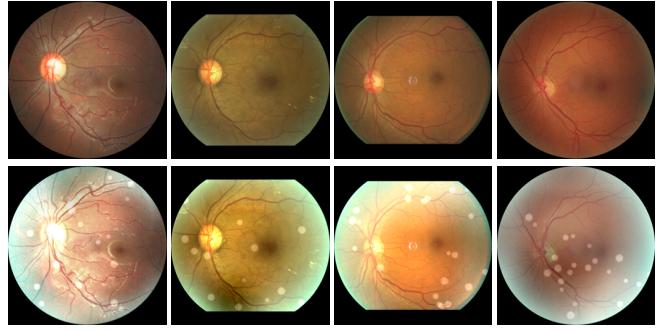
## V. EXPERIMENTS AND EVALUATION

### A. Datasets

To train our cofe-Net, we manually select 13,000 high-quality images, free from interference factors, from the EyeQ dataset [6], which is based on Kaggle Diabetic Retinopathy Detection dataset [47]. We randomly choose the degradation factors (e.g., light transmission disturbance, imaging blurry, and retinal artifacts), and process images using the proposed fundus image degradation formulation to online generate low-quality image, for generating the training set. Note that we also process the images by blending all factors to simulate a complicated real-world situation. The source code of our degradation algorithm and pre-trained correction model will be available. For testing, we also utilize the proposed degradation model to randomly generate degraded images for quantitative evaluation. We totally synthesize 500 testing images using the Kaggle [47] dataset and DRIVE [48] 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 also randomly choose 50 images from the Kaggle dataset [47], and build a benchmark to support a user study task.

### B. Implementation

During training, we resize the input degraded images to  $512 \times 512$  with a batch size of 16 for each iteration. To accelerate the training process, we first train the RSA and LQA module, and then apply a end-to-end training strategy to the whole correction framework. The proposed cofe-Net is



**Fig. 5. Synthetic samples in low-quality fundus image dataset.**

implemented in PyTorch, using stochastic gradient descent (SGD) for optimization. The learning rate is initialized to  $1 \times 10^{-4}$  for the first 150 epochs and then gradually decays to zero over the next 150 epochs.

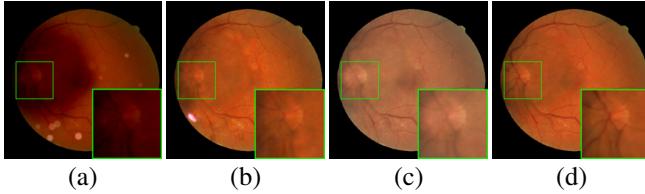
### C. Degradation Model Setting

In the proposed retinal fundus images degradation model, we summarize the interference in terms of three factors, including light transmission disturbance, imaging blur and retinal artifacts. To better simulate the real clinical scenes, we analysis the images in Kaggle [47] dataset, and give a setting description of each items.

**Light Transmission Disturbance:** The light transmission disturbance can be achieved by a convolution operator as shown in (1) in the manuscript. As analyzed in manuscript, we simulated this interference as an aggregation model, which is consist of global and local factors. The global factor is modeled by contrast ( $\alpha$ ), brightness ( $\beta$ ) and saturation ( $s$ ). Thus, we simulate random color jitters with a probability of  $-0.5$  to  $0.5$ . For the local factors, it is defined as an extra non-uniform illumination on the fundus imaging. To define the light leak phenomenon, we randomly define the panel centered at  $c = a, b \in [0.375r_L, 0.625r_L]$ , and  $r_L \in [0.75w, w]$ , where  $w$  denotes image size. The Gaussian filter  $G_L$  is then applied to ensure luminance smoothness. Here we define the  $\sigma_L \in [0.66cr_L, 0.66(w - c)r_L]$ . Especially, to model the usual light leak, we define three kinds of illumination bias as  $[0.63, 0.80, 0.35]$ ,  $[0.56, 0.93, 0.93]$ , and  $[1, 1, 1]$ , it is able to well simulate the real light leak in imaging procedure. Similar to the degradation model of light leak, for the uneven exposure problem, we further define the  $r_L \in [0.3w, 0.5w]$ , and the illumination bias for underexposure model is simulated as a local brightness jitters with a probability of  $-0.5$  to  $-0.1$ . The  $\sigma_L \in [0.55r_L, 0.75r_L]$  of Gaussian filter is defined for the following smooth operator.

**Imaging Blurry:** The defocus imaging blurry can be achieved by a convolution operator as shown in (2) in the manuscript. We introduce a defocus degradation model to simulate the image blurring factor, which is caused by an undesired object distance setting during funduscopy. Here we set  $\sigma_B = 0.03w$ , and  $r_B \in [0.01w, 0.015w]$ .

**Retinal Artifacts:** The retinal artifacts can be achieved by a convolution operator as shown in (4) in the manuscript. To simulate this retinal artifacts problem, due to attaching



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

on the lens, we define the undesired objects imaging procedure undergoes defocus blur. In this task, we define the object  $i$  with a radius  $r_i \in [0.025w, 0.05w]$ , and the defocus 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 undesired objects imaging, we especially randomly define 10 to 25 objects on each fundus image, it is able to simulate this interference in the real clinical scenes. In Fig. 5, we provide such generated samples undergoing the simulated degradation factors, the constructed low-quality image dataset can be explored to train our clinical-oriented fundus enhancement network for retinal fundus image correction.

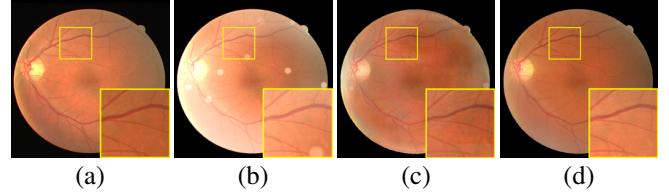
#### D. Ablation Study

To demonstrate the effectiveness of cofe-Net, ablation study results are reported. 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  $L_2$  loss. As shown in Fig. 6 (a), unlike natural images 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  $L_2$  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 and presented the explicit visual information. 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 the generic factors affecting the imaging quality, such as the exposure, and defocus, which are introduced based on the passive incidents, the undesired matter images in initiative situations is 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 artifacts. We provide examples in Fig. 6 (d). It can be seen, the proposed network including both the LQA module and robust structure-perceiving loss function is able to better remove outlier arti-

**TABLE I**  
**ABLATION STUDY.** OUR FULL MODEL PRODUCES COMPARABLE RESULTS ON THE TESTING DATASETS BASED ON VARIOUS METRICS.

Multi-scale	$\mathcal{L}_c$	RSA	LQA	DRIVE [48]		Kaggle [47]	
				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
				<b>21.24</b>	<b>0.758</b>	<b>20.51</b>	<b>0.885</b>



**Fig. 7. Effect of multi-scale framework on enhancement.** (a) Ground truth image. (b) Low-quality image. (c) w/o multi-scale fashions. (d) w/ multi-scale fashions. The proposed multi-scale fashion is able to correct fundus images with more details.

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

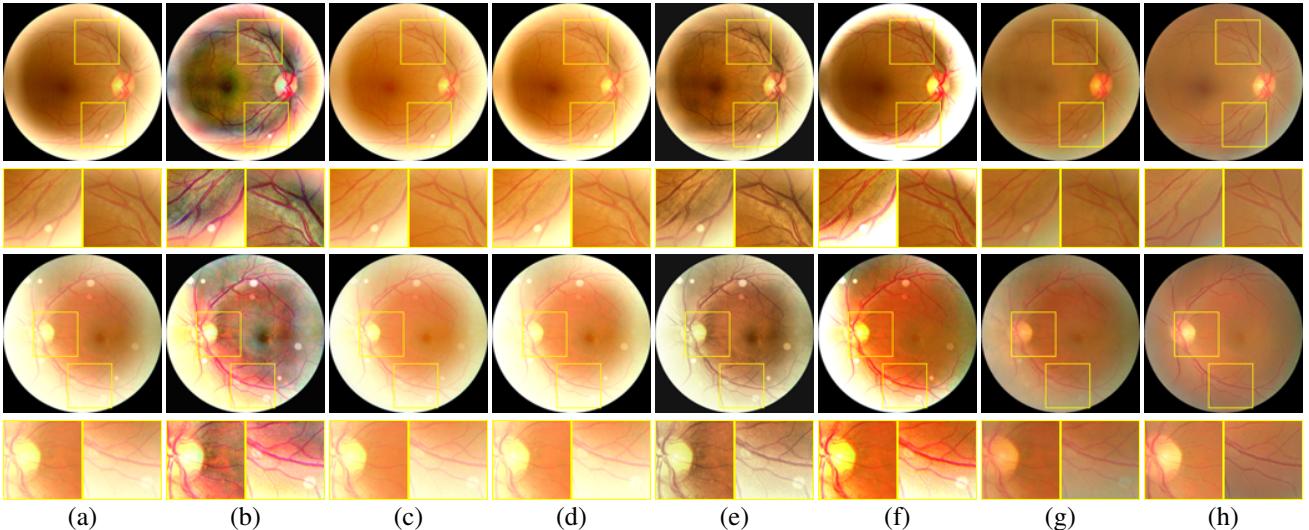
Dataset	DRIVE [48]		Kaggle [47]	
	PSNR	SSIM	PSNR	SSIM
Cheng et al. [34]	14.97	0.648	15.02	0.845
Guo et al. [20]	14.10	0.703	13.54	0.868
Fu et al. [19]	15.56	0.722	14.66	0.882
Tian et al. [23]	15.42	0.721	14.71	0.664
Zuiderveld [49]	15.93	0.740	14.05	0.716
Eilertsen et al. [8]	19.01	0.755	18.40	0.841
Our cofe-Net	<b>21.24</b>	<b>0.758</b>	<b>20.51</b>	<b>0.885</b>

facts 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, 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 produce considerable performance promotion.

#### E. Comparison with State-of-the-art Methods

In this section, we provide qualitative and quantitative comparisons with six correction methods [8], [19], [20], [23], [34], [49] 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 applied for real clinical image correction tasks. As shown in Fig. 8 and Table II, the proposed



**Fig. 8.** Visual comparison of synthesized clinical retinal images. The results of proposed algorithm is able to correct images with preserving the clinical features and suppressing the retinal artifacts. (a) real images. (b) Setiawan et al. [16]. (c) Fu et al. [19]. (d) Guo et al. [20]. (e) Tian et al. [23]. (f) Cheng et al. [34]. (g) Eilertsen et al. [8]. (h) Ours.

cofe-Net achieves comparable performance, performing particularly favorably against the existing methods on the blended degraded images. We also provide a visualized 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 [20], [34], [49] fail to remove the noticeable undesired retinal artifacts. Specifically, medical image correction aims to restore a high-quality image by suppressing noise while at the same time preserving essential pathological characteristics. The hand-crafted feature-based methods consider a publicly distributed image character which often results in aliasing artifacts in specific medical images, such as aggravating an uneven illumination and a tessellated fundus. In contrast, our method is capable of enhancing fundus images as well as removing these additional artifacts, which accelerates the subsequent diagnosis task.

We further provide comparisons on real fundus images, employing user studies to quantitatively evaluate the state-of-the-art correcting methods and our approach. Our paired comparison strategy evaluates the quality of the medical images. For each test, we provide pairs of restored fundus images generated by the compared correction methods and our approach. We display these results in random order and ask the participants to vote for one result based on the instructions. Note that both the suppression of artifacts and preservation of lesions should be taken into account. Finally, we collect the feedback from 87 participants with 79 valid votes. The percentage 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 better-corrected results.

#### F. Clinical Image Analysis and Application

Since the medical image correction models must be effectively applied to real clinical tasks, to demonstrate the effectiveness of the proposed correction method, we conduct additional experiments on clinical image analysis, involving

**TABLE III**  
PERCENTAGE SCALE OF ENHANCEMENT METHOD IN USER STUDY.

method	Image Quality	Lesions Quality
Zuiderveld [49]	2.53	3.80
Guo et al. [20]	5.06	6.33
Fu et al. [19]	6.33	2.53
Tian et al. [23]	8.86	17.72
Cheng et al. [34]	10.13	12.66
Eilertsen et al. [8]	22.78	15.19
Our cofe-Net	<b>44.31</b>	<b>41.77</b>

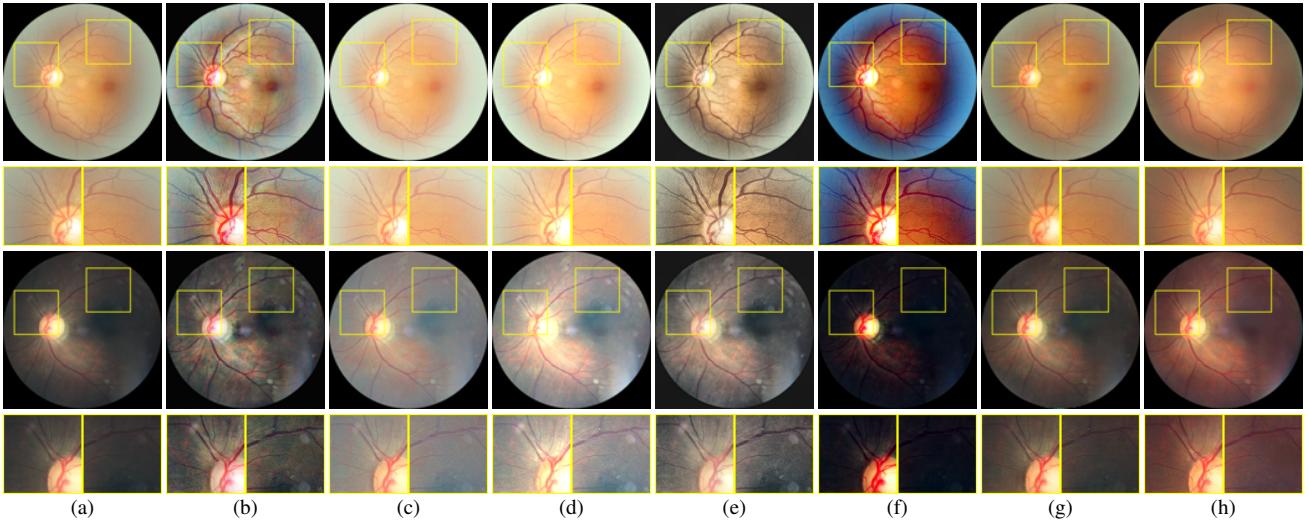
**TABLE IV**  
VESSEL SEGMENTATION AND DISC/CUP DETECTION EVALUATION.

	Vessel Seg.			Disc/Cup Seg.	
	AUC	Acc	Sen	F-score	J-score
w/o enhance	0.924	0.943	0.532	0.735	0.746
w/ enhance	<b>0.957</b>	<b>0.951</b>	<b>0.6</b>	<b>0.890</b>	<b>0.863</b>

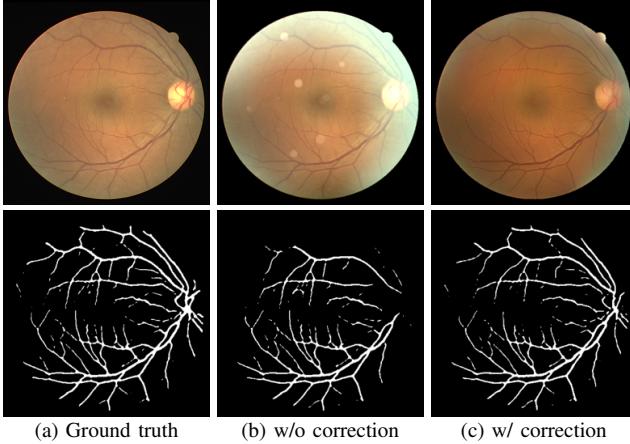
the vessel segmentation and optic disc/cup detection tasks. The DRIVE [48] and REFUGE [50] datasets are selected to evaluate, which provide the annotations of vessel and disc/cup regions. And CE-Net [7] and M-Net [51] are employed for these two tasks as automated segmentation baseline.

**Vessel segmentation:** For the vessel segmentation, based on the degradation model, 100 low-quality images generated from DRIVE [48] are selected for quantitative assessment. The vessel segmentation results are shown in Fig. 10 and Table IV. The CE-Net fails to obtain the high performance on the low-quality clinical images. In contrast, a strong vessel structure can be extracted after applying the proposed correction method. It obtain the better performance on corrected images.

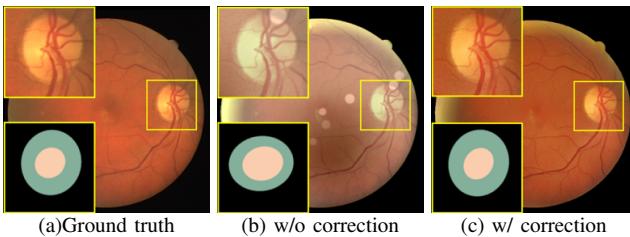
**Optic disc/cup detection:** We also conduct the experiment on disc and cup detection. We simulate 400 degraded images with different setting from the REFUGE test set [50], and use these to validate our method. The M-Net [51] is used to segment disc/cup. We report the F-scores and J-scores in Table IV and



**Fig. 9. Visual comparison of real clinical retinal images.** The results of proposed algorithm is able to correct images with preserving the clinical features and suppressing the retinal artifacts. (a) real images. (b) Setiawan et al. [16]. (c) Fu et al. [19]. (d) Guo et al. [20]. (e) Tian et al. [23]. (f) Cheng et al. [34]. (g) Eilertsen et al. [8]. (h) Ours.



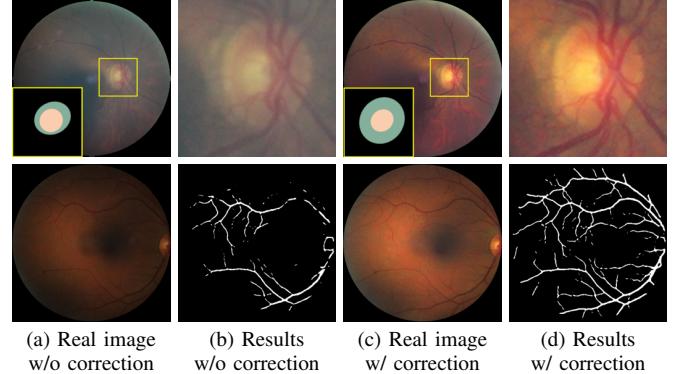
**Fig. 10. Visual comparison of vessel segmentation.** The proposed method can improve the clinical analysis on vessel segmentation.



**Fig. 11. Visual comparison of optic disc/cup detection.** The proposed method can improve the clinical analysis on disc/cup detection.

quantitative results in Fig. 11, where our method recovers a high dynamic range in the low-quality images, boosting the discriminative representations of the disc/cup.

**Real clinical fundus image:** To demonstrate the effectiveness of clinical value, we also validate the proposed correction method for vessel segmentation and optic disc/cup detection tasks on the real fundus images. As shown in Fig. 12, our method produces legitimate corrected real images with clear



**Fig. 12. Vessel segmentation and optic disc/cup detection on real clinical images.**

clinical structure that can be successfully processed by the referenced vessel segmentation model, i.e. CE-Net [7]. As same as the vessel, the presences of optic disc/cup regions are also enhanced to be more cognizable.

## VI. CONCLUSION

In this paper, a clinical-oriented fundus enhancement network, named cofe-Net, has been proposed to correct the low-quality fundus image while preserving accurate lesion areas and retinal structures. Furthermore, a completed degradation model was also introduced to generate adequate training image pairs. Experiments supported our insights into the problems of fundus image correction and degradation factor modeling. Our cofe-Net boosted the performances of several clinical tasks such as vessel segmentation and disc/cup detection. Our method could assist ophthalmologists in ocular disease diagnosis through retinal fundus image observation and analysis, and simultaneously, also be beneficial to automated image analyzing system.

## REFERENCES

- [1] M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal Imaging and Image Analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, 2010.
- [2] U. Schmidt-Erfurth, A. Sadeghipour, B. S. Gerendas, S. M. Waldstein, and H. Bogunović, "Artificial intelligence in retina," *Progress in Retinal and Eye Research*, vol. 67, pp. 1–29, 2018.
- [3] A. W. Scott, S. Farsiu, L. B. Enyedi, D. K. Wallace, and C. A. Toth, "Imaging the infant retina with a hand-held spectral-domain optical coherence tomography device," *American journal of ophthalmology*, vol. 147, no. 2, pp. 364–373, 2009.
- [4] R. Qian, O. M. Carrasco-Zevallos, S. Mangalesh, N. Sarin, L. Vajzovic, S. Farsiu, J. A. Izatt, and C. A. Toth, "Characterization of long working distance optical coherence tomography for imaging of pediatric retinal pathology," *Translational vision science & technology*, vol. 6, no. 5, pp. 12–12, 2017.
- [5] S. Philip, L. M. Cowie, and J. A. Olson, "The impact of the health technology board for scotland's grading model on referrals to ophthalmology services," *British Journal of Ophthalmology*, vol. 89, no. 7, pp. 891–896, 2005.
- [6] H. Fu, B. Wang, J. Shen, S. Cui, Y. Xu, J. Liu, and L. Shao, "Evaluation of Retinal Image Quality Assessment Networks in Different Color-Spaces," in *MICCAI*, 2019, pp. 48–56.
- [7] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for medical image segmentation," *IEEE Transactions on Medical Imaging*, 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 Transactions on Graphics (TOG)*, 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 *CVPR*, 2018.
- [10] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *CVPR*, 2019.
- [11] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M. Yang, "Low-light image enhancement via a deep hybrid network," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4364–4375, 2019.
- [12] M. Foracchia, E. Grisan, and A. Ruggeri, "Luminosity and contrast normalization in retinal images," *Medical Image Analysis*, vol. 9, no. 3, pp. 179–190, 2005.
- [13] F. Marco, G. Enrico, and R. Alfredo, "Luminosity and contrast normalization in retinal images," *Medical Image Analysis*, vol. 9, no. 3, pp. 179–190, 2005.
- [14] S. Jenifer, S. Parasuraman, and A. Kadirvelu, "Contrast enhancement and brightness preserving of digital mammograms using fuzzy clipped contrast-limited adaptive histogram equalization algorithm," *Applied Soft Computing*, vol. 42, pp. 167–177, 2016.
- [15] S. J. Hwang, A. Kapoor, and S. B. Kang, "Context-based automatic local image enhancement," in *ECCV*, 2012, pp. 569–582.
- [16] A. W. Setiawan, T. R. Mengko, O. S. Santoso, and A. B. Suksmono, "Color retinal image enhancement using clahe," in *International Conference on ICT for Smart Society*, 2013, pp. 1–3.
- [17] M. K. Ng and W. Wang, "A total variation model for retinex," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 345–365, 2011.
- [18] W. Wang and M. K. Ng, "A nonlocal total variation model for image decomposition: illumination and reflectance," *Numerical Mathematics: Theory, Methods and Applications*, vol. 7, no. 3, pp. 334–355, 2014.
- [19] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *CVPR*, 2016, pp. 2782–2790.
- [20] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2017.
- [21] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 1032–1041, 2012.
- [22] M. Zhou, K. Jin, S. Wang, J. Ye, and D. Qian, "Color retinal image enhancement based on luminosity and contrast adjustment," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 3, pp. 521–527, 2017.
- [23] Q.-C. Tian and L. D. Cohen, "Global and local contrast adaptive enhancement for non-uniform illumination color images," in *ICCV Workshops*, Oct 2017.
- [24] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 460–489, 2005.
- [25] X. Lu, Y. Yuan, and P. Yan, "Image super-resolution via double sparsity regularized manifold learning," *IEEE transactions on circuits and systems for video technology*, vol. 23, no. 12, pp. 2022–2033, 2013.
- [26] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *ICCV*, 2011, pp. 479–486.
- [27] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *CVPR*, 2009, pp. 1964–1971.
- [28] W. Ren, X. Cao, J. Pan, X. Guo, W. Zuo, and M. Yang, "Image deblurring via enhanced low-rank prior," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3426–3437, 2016.
- [29] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.
- [30] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, 2013.
- [31] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *MICCAI*, 1998.
- [32] 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," *Optics & Laser Technology*, vol. 58, pp. 56–62, 2014.
- [33] 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 *ISBI*, 2015, pp. 226–230.
- [34] 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 Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2536–2546, 2018.
- [35] K. G. Lore, A. Akintayo, and S. Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.
- [36] K. Wei, J. Yang, Y. Fu, D. Wipf, and H. Huang, "Single image reflection removal exploiting misaligned training data and network enhancements," in *CVPR*, 2019, pp. 8178–8187.
- [37] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang, "Gated fusion network for single image dehazing," 2018.
- [38] 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 *CVPR*, 2018, pp. 6286–6295.
- [39] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, "Deep semantic face deblurring," in *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 Transactions on Image Processing*, vol. 27, no. 1, pp. 194–205, 2017.
- [41] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *ECCV*, 2016, pp. 154–169.
- [42] F. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: low-light image/video enhancement using cnns," in *BMVC*, 2018, p. 220.
- [43] H. Talebi and P. Milanfar, "Learned perceptual image enhancement," in *ICCP*, 2018, pp. 1–13.
- [44] R. Liu, L. Ma, Y. Wang, and L. Zhang, "Learning converged propagations with deep prior ensemble for image enhancement," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1528–1543, 2019.
- [45] H. Zhao, B. Yang, L. Cao, and H. Li, "Data-driven enhancement of blurry retinal images via generative adversarial networks," in *MICCAI*, 2019, pp. 75–83.
- [46] U. Upadhyay and S. P. Awate, "A mixed-supervision multilevel GAN framework for image quality enhancement," in *MICCAI*, 2019, pp. 556–564.
- [47] Kaggle diabetic retinopathy detection. [Online]. Available: <https://www.kaggle.com/c/diabetic-retinopathy-detection/data>
- [48] 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 Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.
- [49] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV*, 1994, pp. 474–485.
- [50] J. I. Orlando, H. Fu, J. B. Breda *et al.*, "REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *CoRR*, vol. abs/1910.03667, 2019.
- [51] 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 Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1597–1605, 2018.