Dehaze of Cataractous Retinal Images Using an Unpaired Generative Adversarial Network

Yuhao Luo [ID], Kun Chen [ID], Lei Liu [ID], Jicheng Liu [ID], Jianbo Mao [ID], Genjie Ke [ID], and Mingzhai Sun [ID]

Abstract—Cataracts are the leading cause of visual impairment worldwide. Examination of the retina through cataracts using a fundus camera is challenging and error-prone due to degraded image quality. We sought to develop an algorithm to dehaze such images to support diagnosis by either ophthalmologists or computer-aided diagnosis systems. Based on the generative adversarial network (GAN) concept, we designed two neural networks: Cataract-SimGAN and CataractDehazeNet. CataractSimGAN was intended for the synthesis of cataract-like images through unpaired clear retinal images and cataract images. Cataract-DehazeNet was trained using pairs of synthesized cataract-like images and the corresponding clear images through supervised learning. With two networks trained independently, the number of hyper-parameters was reduced, leading to better performance. We collected 400 retinal images without cataracts and 400 hazy images from cataract patients as the training dataset. Fifty cataract images and the corresponding clear images from the same patients after surgery comprised the test dataset. The clear images after surgery were used for reference to evaluate the performance of our method. CataractDehazeNet was able to enhance the degraded image from cataract patients substantially and to visualize blood vessels and the optic disc, while actively suppressing the artifacts common in application of similar methods. Thus, we developed an algorithm to improve the quality of the retinal images acquired from cataract patients. We achieved high structure similarity and fidelity between processed images and images from the same patients after cataract surgery.

Index Terms—Cataract, dehaze, retinal image, unpaired generative adversarial network.

I. INTRODUCTION

CATARACT is one of the most prevalent causes of vision loss, with an estimated 16 million people affected worldwide [1]. Since cataract is primarily an age-related disorder, this prevalence could double by 2020 due to aging of the world's population [1]–[3]. Retinal imaging of cataract patients using fundus cameras is challenging because the image quality is severely degraded by light scattering due to the turbidity of the ocular media [4]. Diagnosis based on hazy cataract images, either by human experts or computer-aided diagnosis systems (CAD) [5], [6], is error-prone. Scanning laser confocal retinal imaging (SLO) [7]–[10] is less affected by the features of cataract and offers better retinal images in cataract patients. However, SLO is much more expensive and less commonly available, especially in rural areas. Thus, developing an algorithm to enhance the contrast and improve the readability of cataract retinal images is of significant practical and clinical importance, especially for retinal health screening programs, such as those for diabetic retinopathy [11]. Conventional methods based on contrast limited adaptive histogram equalization [12] and luminance gain matrix in the HSV color space [13] were developed to enhance the color retina images. These methods were independent of the mechanism of the retinal image deterioration and were not specifically for cataractous retina images.

An optical model for imaging the cataractous retina was proposed by Peli [4] based on the imaging model of foggy natural scenes [14]–[16]. Algorithms used for natural scene haze removal can be divided into two categories: dark channel-based [17]–[19] and deep convolutional neural network (DCNN)-based. The dark channel-based algorithm was mainly developed from the dark channel prior first proposed by He [17]. Dark channel prior assumes that at least one channel of the hazy natural scene image is dark, with the majority of the pixel values close to zero. This assumption is needed to estimate the transmission function for the dehaze process. However, due to the significant differences between the retinal and natural scene images [20], this assumption may not hold, and can result in poor performance when the method is applied to retinal images. Although DCNNs [21] have previously been applied for haze removal for natural scenes [22]–[25] and medical images [26]–[31], these algorithms are not easily transferrable to cataractous retinal image. One reason is the lack of image pairs of an initially hazy image and the corresponding clear image following cataract surgery that are required for supervised deep learning [32] [33]. Even when such image pairs exist, the images taken before and after surgery are not usually aligned. Because of the low image quality of the cataract image, the alignment of the such image pairs is challenging, if not impossible. However, obtaining such image pairs is a critical step for applying supervised learning.

Recently, Zhu *et al.* [34] proposed CycleGAN, a generative adversarial network (GAN) [35]–[37] that utilizes unpaired images to train a dehaze network. It has been applied for haze removal in natural images without paired training images. However, unlike natural images, retinal images have many small features, like thin blood vessels. The consistency loss in the CycleGAN has greater effects on the dominant and global structures. CycleGAN produces small vessel-like artifacts when applied to dehaze cataractous images. During the revision of the current study, we noticed that Shen *et al.* [38] developed a clinical-oriented fundus enhancement network (cofeNet), which took into consideration of the major inferior-quality factors and achieved excellent performance.

In this study, we developed a two-stage process for haze removal in cataractous images. Inspired by the SimGAN [39], we first designed an unpaired GAN, CataractSimGAN, to generate synthesized cataract-like images. Based on the synthesized cataract-like image and the corresponding clear image, we trained a paired DCNN (CataractDehazeNet) for haze removal of cataractous images. The similarity of the synthesized cataract-like images to the real cataract images was critical for training CataractDehazeNet. To improve this similarity, we applied a multi-scale D net with different receptive fields to preserve the consistency of both the fine and global structures. Furthermore, we developed two specific loss functions to suppress artifacts: one to facilitate the fidelity of the vessel and optic disc structures, and the other was a multi-scale feature loss, which suppressed the background noise significantly. Our proposed method outperformed previous studies and achieved state-of-the-art performance in dehazing cataractous images.

II. RELATED WORK

We describe two categories of algorithms for haze removal in details: dark channel-based, and DCNN-based.

A. Dark Channel-Based Haze Removal

Peli [4] developed an optical model for imaging the retina through cataract, as shown in the following equation:

$$I(x) = \alpha \cdot L \cdot r(x) \cdot t(x) + L[1 - t(x)] \qquad (1)$$

where x denotes pixels of the image, $I(x)$ is the cataract image, L is the constant illumination of the camera, $r(x)$ is the reflectance function with all the distortions from either the eye or the optical system of the camera, $t(x)$ is the transmission function of the system, and α is the attenuation factor of the retinal illumination due to cataract [40]. This cataract imaging model is the same as the dehaze model [14]–[16] used in computer vision for images attenuated by haze or fog.

One of the simplest but most effective algorithms for haze removal is based on the dark channel prior proposed by He [17]. The dark channel prior is based on the critical observation that most outdoor images of natural scenes have very low intensity in at least one channel. For the dark channel prior, the model is as follows:

$$\frac{I^c(x)}{L^c} = t(x)\frac{D^c(x)}{L^c} + (1 - t(x)) \qquad (2)$$

where c is color channel index, and $D^c(x)$ is the clear image captured under ideal human-lens. To calculate the value of $I^c(x)$ from $D^c(x)$, we must obtain the values of $t(x)$ and L^c. With the help of the dark channel prior $min(\min_{y\in\Omega(x)}(\frac{D^c(x)}{L^c})) = 0$, $t(x)$ is:

$$t(x) = 1 - min(\min_{y\in\Omega(x)}(\frac{I^c(y)}{L^c})) \qquad (3)$$

where y is an element of $\Omega(x)$, a local patch centered at x of the image. For L^c, the top 0.1% brightest pixels are picked from the dark channel. Among these pixels, those with the highest intensity in $I^c(x)$ are selected as L^c. Then, with $t(x)$ and L^c, the clear image can be calculated from the hazy image. Many improved algorithms have been developed based on the original dark channel prior. For example, the weak dark channel prior (WDCP), a pixel-based dark channel, was designed to improve the speed of the original algorithm [41]. It achieved a similar effect more than 20 times faster. Later, a fast method for single image haze removal based on multiscale dark channel prior with significantly reduced complexity and increased speed was proposed [42].

Guided image filtering (GIF) [19] was proposed by He to optimize the transmission function $t(x)$. It utilizes a guidance image to provide filtering output that is more structured and less smoothed than the input.

Inspired by GIF, a novel structure-preserving guided retinal image filtering (SGRIF) has been used to dehaze cataract images [18]. This is a novel filtering method for retinal image, particularly in the optic disc region. It consisted of a global structure transfer filter and a global edge-preserving smoothing filter to preserve fine structures in regions that were flat.

B. DCNN for Haze Removal

With the rapid development of DCNN, several dehazing algorithms based on DCNN have been proposed. One category seeks to improve $t(x)$ with DCNN, and another uses an end-to-end design, with the cataract image as input and dehaze image as the output of the network. The third category, based on the CycleGAN, uses unpaired cataract images during the training process.

As discussed above, the transmission function $t(x)$ is critical for the dehazing result. DCNN-based methods have been proposed to calculate the relationship between the hazy image and $t(x)$. Such methods use the hazy image as input, and provide its medium transmission map as output. This $t(x)$ is subsequently applied to obtain the clear image [22], [43]–[45]. In [43], the authors designed a special deep architecture called DehazeNet to embody established image dehazing principles. They also proposed a novel nonlinear activation function to improve the quality of the recovered haze-free image. A cardinal (red, green and blue) color fusion network was proposed to solve the problem of color distortion in the "gloomy" environments with poor illumination [44].

Another type of DCNN-based algorithms bypass the transmission function $t(x)$. Dudhane *et al.* [23] designed a RI-GAN, which directly recovers the haze-free scene. Such a design may
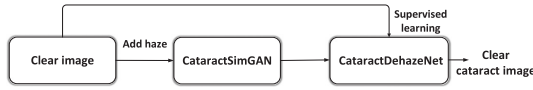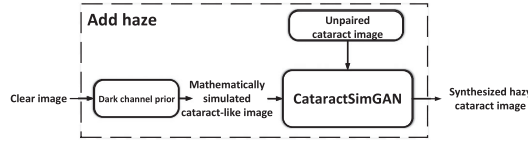
Fig. 1. Overview of the proposed algorithm.

Fig. 2. The flowchart of the process of generating synthesized cataract-like images.

avoid cascade error in the intermediate stages. The authors designed a novel residual inception (RI) module, which consists of dense connections within the multi-scale convolution layers to learn the integrated flavors of the haze-related features.

Unpaired CycleGAN has recently gained attention for medical image applications [46], [47]. The dehaze algorithm based on CycleGAN combines cycle-consistency loss and perceptual losses to improve the quality of textural information recovery and generated visually superior haze-free images [47]. A recent study by Zhao *et al.* demonstrated significant enhancement of retinal images using a CycleGAN with modified consistency loss functions [46].

III. METHOD

Due to the lack of clear and cataract image pairs for supervised training, we first synthesized hazy cataract-like images from clear images. We used the synthesized image pairs to train a neural network (dehaze network), which was later used to dehaze an unseen real cataract image. Fig. 1 provides an overview of the proposed algorithm.

A. Synthesized Cataract-Like Images

The similarity between the synthesized cataract-like images and the real cataract images is critical for training the dehaze network. We applied a two-stage process to synthesize cataract-like images, as shown in Fig. 2. First, we applied the mathematical haze model to generate a simulated cataract-like image. Then, we designed and applied an unpaired GAN (CataractSimGAN) to further refine the mathematically simulated cataract-like images to obtain the final, synthesized cataract-like images.

1) Simulated Cataract Images Based on the Dark Channel Prior: We applied the algorithm based on the dark channel prior proposed by He [17] to generate a simulated cataract image $(I^c(x))$ from a clear image $(D^c(x))$ using Eq. (2) and (3).

2) Simulated Cataract Images Based on CataractSimGAN: The mathematically simulated cataract-like images required further refinement because the dark channel prior assumption is not completely valid for retinal images. Thus, we designed CataractSimGAN to further improve the images using unpaired real cataract images.

a) Network structure: The architecture of the CataractSim-GAN is shown in Fig. 3. It consists of a G net and two identical D nets with different input image sizes.

The G net has four down-sampling and four up-sampling layers. Between the down-sampling and up-sampling, there are 9 Res-blocks. Both the down-sampling and up-sampling layers use modules of the form Convolution/Deconvolutional-BatchNorm-ReLu. For the res-block, it applies the form Convolution-Convolution-BatchNorm-ReLu and its skip connection is from the first convolutional layer to the last ReLu layer. The input size of the G net is fixed as 512×512.

The D net is an unconditional net [48], [49] consisting of three down-sampling layers and one convolutional layer. Each down-sampling layer has the same form of the down-sampling layer in G net. The form Convolution-Convolution-BatchNorm-ReLu is applied for the convolutional layer.

To discriminate the real and synthesized images both in the global and finer features, the D net should have a different receptive field. We proposed using a multi-scale (512×512 and 256×256) D net. We created an image pyramid of two scales by down-sampling the real and synthesized images by a factor of 2. The multi-scale D net could be divided into two parts, named D_1 net and D_2 net. D_1 net and D_2 net had identical network architectures; however, the input of D_1 net is an image of size 512×512, and the input of D_2 net is an image of 256×256.

b) The loss function: The loss function of the D net is the same as that of the patchGAN [50]. The patch losses of each scale are summed to form the final loss of the D net.

$$\mathcal{L}_D = \mathbb{E}\left(\log \sum_{i=1}^{2} D_i\left(I_r\right)\right) + \mathbb{E}\left(\log\left(1 - \sum_{i=1}^{2} D_i(G(I_i))\right)\right) \tag{4}$$

where I_i is the input image, I_r is the real image, G is the G net, and D is the D net.

In traditional image-to-image GAN, the fake image generated by the G net would be similar to the real image. However, for CataractSimGAN, we only need the generated image to be similar to the real image in terms of its hazy features. The dominant structures, the blood vessels (BV), and optic disc (OD) of the fake image should be the same as in the clear image. Inspired by SimGAN, we introduce the loss functions of "Multi-scale feature loss" and "BV-OD loss" between the input image and fake image to maintain the position of the BV and OD.

The final loss function for the G net of CataractSimGAN is as follows:

$$\mathcal{L}_G = \mathcal{L}_{G\_GAN} + \lambda_1 \mathcal{L}_{G\_Multi-scalefeature} + \lambda_2 \mathcal{L}_{G\_BV\_OD} \tag{5}$$

where $\mathcal{L}_{G\_GAN}$ is the GAN loss, $\mathcal{L}_{G\_Multi-scalefeature}$ is the Multi-scale feature loss of U-net, $\mathcal{L}_{G\_BV\_OD}$ is the BV-OD loss. λ_1 and λ_2 are set to 100 and 10, respectively. The details of each loss function are as follows:

- GAN loss
  The GAN loss for the G net is:

$$\mathcal{L}_{G\_GAN} = \mathbb{E}\left(\log(1 - \sum_{i=1}^{2} D_i\left(G(I_i)\right))\right) \tag{6}$$
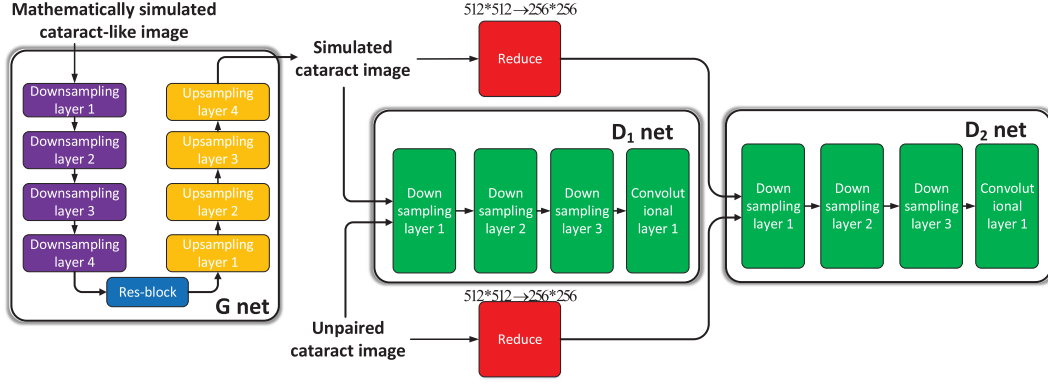
Fig. 3. Architecture of CataractSimGAN. It consists of a G net and two identical D nets with different input image sizes.
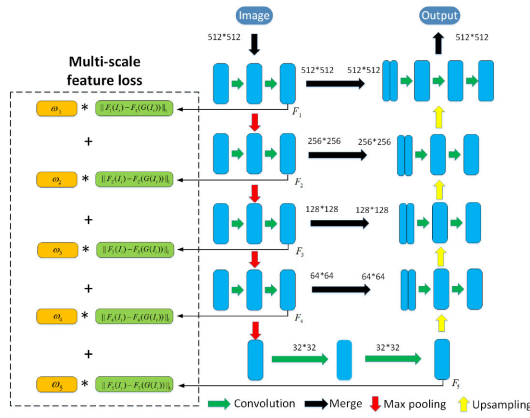

Fig. 4. Multi-scale feature loss function.

where I_i is the input image, G is the G net, and D_i is the ith D net.

- Multi-scale feature loss
  We trained two U-nets [51], one to segment the BV and the other to segment the OD. The two segmentation U-nets were pre-trained on datasets manually labeled by an ophthalmologist. In total 210 images were labeled for BV segmentation, with 140 images for training and 70 for testing. For OD segmentation, 500 images were manually labeled for training and 250 for testing. We applied the SGD optimizer with a learning rate of 0.005.
  As shown in Fig. 4, for each down-sampling we extracted a feature vector, for a total of five feature vectors extracted. The L1 distance between the vectors from the real input image (I_i) and the generated image ($G(I_i)$) was used as the loss. The weights of the five sub-losses were set as $[\frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1]$. The multi-scale feature loss is defined as:

$$\mathcal{L}_{G\_Multi-scalefeature} = (\sum_{k}^{5} \omega_k(||(F_k^{bv}(I_i)$$
$$- F_k^{bv}(G(I_i))||_1) + \sum_{k}^{5} \omega_k(||(F_k^{od}(I_i) - F_k^{od}(G(I_i))||_1)) \quad (7)$$

where $F_k^{bv}()$ and $F_k^{od}()$ represent the k-th feature map of the U-net for the BV and OD, respectively, ω_k is the k-th weight in ω, where $\omega = [\frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1]$.

- BV-OD loss
  The BV and OD are the dominant structures of retinal images. It is important to maintain the fidelity of these structures between the synthesized image and the real input image. Therefore, we introduced the BV-OD loss:

$$\mathcal{L}_{G\_BV\_OD} = ||(G(I_i) - I_i) * (I_{BV} + I_{OD})||_1 \quad (8)$$

where I_{BV} and I_{OD} are the binary BV and OD segmentation images, and $*$ represents the dot product for the matrix.

B. Supervised Learning Based on Synthesized Paired Cataract Images

We used pairs including the synthesized cataract images by CataractSimGAN to train the CataractDehazeNet, which was used later to dehaze unseen real cataract images.

The architecture of CataractDehazeNet is same as the pix2pix network [50], but with a customized loss function:

$$\mathcal{L}_{G'} = \mathcal{L}_{G'\_GAN} + \mathcal{L}_{G'\_L1} +$$
$$\lambda_3 \mathcal{L}_{G'\_Multi-scalefeature} + \lambda_4 \mathcal{L}_{G'\_BV\_OD} \quad (9)$$

where $\mathcal{L}_{G'\_GAN} = \mathbb{E}(\log(1 - D(I, G'(I_i))))$, $\mathcal{L}_{G'\_L1} = ||(G'(I_i) - I_r)||_1$. These two loss functions are the same as the pix2pix network. $\mathcal{L}_{G'\_Multi-scalefeature}$ and $\mathcal{L}_{G'\_BV\_OD}$ are similar to $\mathcal{L}_{G\_Multi-scalefeature}$ and $\mathcal{L}_{G\_BV\_OD}$ differing only in the replacement of I_i in the equation with I_r. λ_3 and λ_4 are 1000 and 1000, respectively.

C. Evaluation Metrics

1) Assessment of Dehazing Results: We used images taken after surgery as the reference to evaluate the image quality of dehaze results. Because the images taken before and after surgery were not aligned, we first applied the SIFT algorithm [52] to align the images before assessing the dehaze image quality. Alignment was restricted to the OD region. Fig. 5 shows representative results of this alignment process.
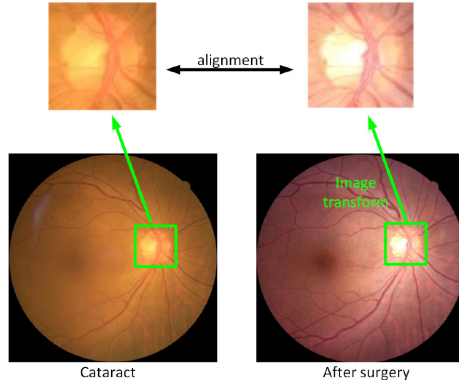
Fig. 5. Image alignment of cataract retinal images taken before (left) and after (right) surgery. Only the OD regions were aligned using the SIFT algorithm.

After alignment, we applied the peak signal-to-noise ratio (PSNR) and the structure similarity index (SSIM) [53] to evaluate the dehazed image relative to the corresponding image after surgery. The PSNR and SSIM are defined as follows:

$$PSNR = 10\log_{10}\left(\frac{MAX^2}{\frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}(I(i,j) - K(i,j))^2}\right) \quad (10)$$

where MAX is the max value of the image (255 for 8-bit grayscale images), M and N are the size of the image, $I(i,j)$ and $K(i,j)$ are the pixel values of the image and its reference image, respectively.

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (11)$$

where μ_x and μ_y are the mean values of the images, σ_x and σ_y are the standard deviations of the images, and σ_{xy} is the covariance of the images. Constants $C_1 = (K_1L)^2$ and $C_2 = (K_2L)^2$ are used to avoid zero in the denominator [53]. L is 255 for 8-bit gray images, and $K_1 = 0.01$ and $K_2 = 0.03$ are defined in reference [53].

2) Evaluation Using the Pix2pix Neural Network: It is difficult to evaluate the similarity between a synthesized cataract-like image and the corresponding real cataract image [54]. However, we assumed that if the synthesized cataract-like images are close enough to the real ones, a neural network trained using synthesized images should work well on real cataract images. Based on this rationale, we used the synthesized image pairs (clear image and corresponding cataract-like image) to train a publicly available pix2pix network. The trained pix2pix network was used to dehaze unseen images in the test dataset, which consisted of 50 real cataract images with corresponding images after surgery. Each dehaze image was compared to the corresponding image after surgery to calculate the SSIM and PSNR. The better the synthesizing algorithms are, the higher the SSIM and PSNR values are. We also compared the dehaze images visually, particularly checking the intensity line profile, to evaluate the synthesizing algorithm.

Cataract-like images	SSIM	PSNR
Mathematically simulated	0.702($\pm$0.105)	15.219($\pm$3.984)
CycleGAN generated	0.701($\pm$0.088)	19.053($\pm$3.851)
CataractSimGAN without prior	0.753($\pm$0.087)	20.277($\pm$4.076)
CataractSimGAN with prior	**0.763($\pm$0.084)**	**20.504($\pm$4.125)**

D. Dataset

For the CataractSimGAN, we selected 400 clear retinal images and 400 real cataract images to compose the training dataset. One clear image and one hazy image were randomly selected to form an image pair, which was used to train Cataract-SimGAN.

For the CataractDehazeNet, we utilized the 400 synthesized cataract-like images from the CataractSimGAN as the input and the corresponding 400 clear images as the label of the training dataset. Fifty cataract images and their corresponding clear images after surgery were used to evaluate the performance of the dehazing. These images were acquired from the Eye Hospital of Wenzhou Medical University with 20 male and 30 female patients. The age range of the patients is from 45 to 88 with the average of 72.9 ($\pm$8.9). We excluded unclear images with (i) presence of dust and dirt in the camera, (ii) dusty camera lenses, (iii) eye blink, and (iv) event of occlusion by eyelashes [20]. All images in the training and test datasets were 512$\times$512.

IV. RESULTS & DISCUSSION

We first evaluated the performance of the proposed Cataract-SimGAN, then compared our approach to the performance of recently published methods using SSIM, PSNR and vessel segmentation as the evaluation metrics. Finally, we demonstrated the performance of the CataractDehazeNet using images with pathological lesions and compared the results with the corresponding clear images acquired from the same patients after cataract surgery.

A. Evaluation of Synthesized Cataract-Like Images

1) Comparison of Mathematically Simulated, CycleGAN Generated, and CataractSimGAN Generated Cataract-Like Images: We compared the performance of pix2pix dehaze networks trained using the mathematically simulated, CycleGAN generated, and CataractSimGAN generated cataract-like images. Furthermore, we checked whether using the mathematically simulated images as the basis for synthesis would improve the performance of the CataractSimGAN. To that end, we evaluated the SSIM and PSNR of the pix2pix-dehaze net using images generated by CataractSimGAN with and without the dark channel prior.

Training the pix2pix-dehaze network with images generated using CataractSimGAN with prior resulted in the highest SSIM and PSNR values (Table I). This suggests that cataract-like images synthesized with this method closely mimicked real
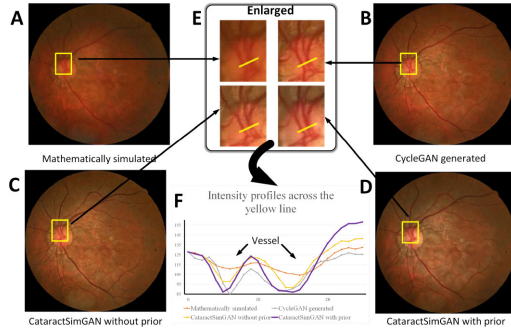
Fig. 6. Visualization of the dehaze results for cataract images obtained using the pix2pix network trained using synthesized cataract-like images.

TABLE II
SSIM AND PSNR OF PIX2PIX NET TRAINED USING SYNTHESIZED CATARACT-LIKE IMAGES GENERATED BY CATARACTSIMGAN WITH DIFFERENT SCALE D NETS IN THE OD REGION

Different scale D net	SSIM	PSNR
Single-scale D net	0.734($\pm$0.077)	15.743($\pm$3.678)
Multi-scale D net	**0.763($\pm$0.084)**	**20.504($\pm$4.125)**

cataract images. Furthermore, these values reveal it was constructive to train the CataractSimGAN on the basis of the mathematically simulated images, that is, with the dark channel prior.

To further evaluate the performance of the these methods, we visually examined the dehaze results of the pix2pix-dehaze networks that were trained using these methods (Fig. 6).

Visually, Fig. 6 A showed much less contrast than Fig. 6 B, C and D. This was further confirmed by examination of the zoomed-in regions enclosed in the yellow rectangles (Fig. 6 E). The differences between Fig. 6 B, C and D were much subtler. However, the line profiles (Fig. 6 F) clearly demonstrate that the dehaze results using the CataractSimGAN with the prior achieved the highest contrast indicated by the purple line in Fig. 6 D with the highest peak-valley difference.

The similarity of the synthesized images to the real cataract images determined the dehaze results of the pix2pix network. The results described above showed that images generated using CataractSimGAN were more like the real cataract images than the mathematically simulated images based on the dark channel prior. However, the mathematically simulated images provided a good basis for CataractSimGAN, since training it on the mathematically simulated images provided better results than training on clear images. This could be because the dark channel prior serves to guide the convergence of the network.

2) Ablation Experiments on the CataractSimGAN:

a) Effectiveness of the multi-scale D nets: We proposed a multi-scale D net architecture for the CataractSimGAN (Fig. 3). Using the multi-scale D nets allowed the receptive field to be expanded, and not only the global features, but the local finer features could be effectively learned and generated as well.

To test the effectiveness of the multi-scale D net, we eliminated the D_2 (256$\times$256) net. The results in Table II show that the SSIM and PSNR of the multi-scale D net were higher than that of the single-scale D net (0.763 vs 0.734 and 20.504 vs 15.743), suggesting the effectiveness of the multi-scale D nets.
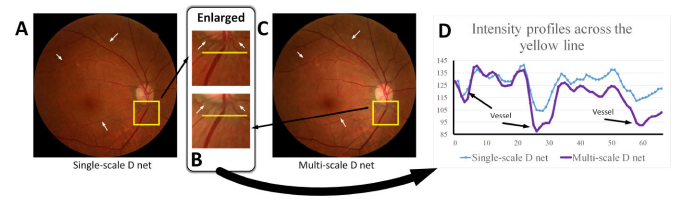

Fig. 7. Visual results of dehazing using pix2pix net trained using synthesized cataract-like images generated with different-scale D nets.

TABLE III
SSIM AND PSNR RESULTS OF THE PIX2PIX NET TRAINED USING THE SYNTHESIZED CATARACT-LIKE IMAGES BY CATARACTSIMGAN WITH DIFFERENT LOSS FUNCTIONS IN THE OD REGION

Loss function	SSIM	PSNR
Baseline	0.697($\pm$0.077)	15.460($\pm$3.906)
Multi-scale feature	0.722($\pm$0.085)	15.135($\pm$3.688)
BV-OD	0.708($\pm$0.077)	16.213($\pm$3.828)
Multi-scale feature & BV-OD	**0.763($\pm$0.084)**	**20.504($\pm$4.125)**

The visual results of dehazing using the pix2pix net trained with images from CataractSimGAN with D_1 only, and with both D_1 and D_2 are shown in Fig. 7. Compared with the single-scale D net (Fig. 7 A), images generated with the multi-scale D net (Fig. 7 C) had sharper contours for BV and were less noisy, especially in the regions indicated by white arrows. To further observe the fine details, we zoomed in the region enclosed in the yellow rectangles (Fig. 7 B) and evaluated their intensity profiles (Fig. 7 D).

The use of multi-scale D nets with different receptive fields helped to preserve the consistency of both the fine and global structures. This ability is particularly important for preserving small pathological features, such as exudates or hemorrhages, present in the retinal images. We also tried adding more D-nets, but did not observe improved performance. The reason could be that with more D-nets, the scale of the image was too small to retain enough structural information of the image.

b) Effectiveness of different loss functions: To improve the similarity of synthesized cataract-like images to the real cataract images, we developed two losses: multi-scale feature loss and BV-OD loss (Eq. (7) and (8)).

We used the synthesized images to train the pix2pix network and then applied the trained network on the test dataset to dehaze the images.

The highest SSIM and PSNR were achieved with the pix2pix network trained using images generated by the CataractSimGAN with both multi-scale loss and BV-OD loss (Table III). The pix2pix dehaze network trained with the L1 loss function (baseline) gave the lowest SSIM and PSNR, further suggesting that images synthesized using CataractSimGAN with specific loss functions more closely mimic real cataract images.

From Table III, we can see that multi-scale feature loss improved SSIM slightly, but with a slight decrease in PSNR. On the other hand, the use of BV-OD loss improved SSIM slightly but provided more improvement on the PSNR. However, the combination of these two loss functions provided significant improvement in both the SSIM and PSNR. This experiment demonstrated the effectiveness of both multi-scale feature and
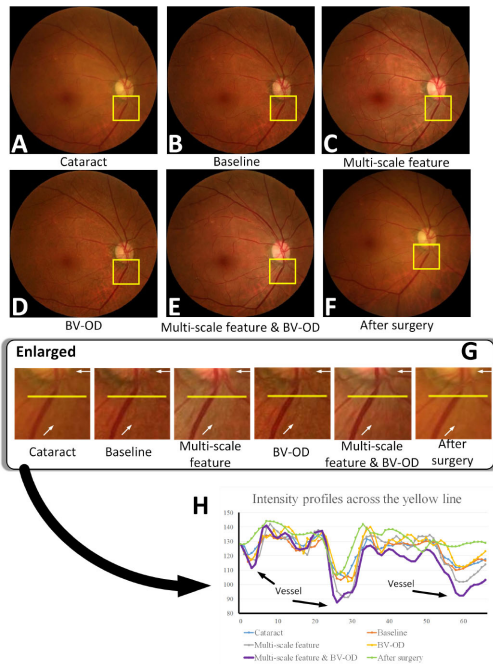
Fig. 8. Dehaze results following processing with different loss functions to generate cataract-like images, shown as both the whole image and enlarged regions.

TABLE IV
SSIM AND PSNR OF THE RESULTS OF CATARACTDEHAZENET APPLIED IN THE OD REGION WITH DIFFERENT LOSS FUNCTIONS

Loss function	SSIM	PSNR
Baseline	0.763($\pm$0.084)	20.504($\pm$4.125)
Dehaze multi-scale feature	0.778($\pm$0.076)	19.924($\pm$3.958)
Dehaze BV-OD	0.771($\pm$0.073)	20.335($\pm$3.978)
Dehaze multi-scale feature & BV-OD	**0.781($\pm$0.079)**	**20.683($\pm$3.878)**

BV-OD losses, and in particular, the complementary effects of these two. BV-OD loss focused more on the dominant BV and OD structures of the retina images, while the multi-scale feature loss improved the whole image quality and suppressed the artifacts.

We further checked the whole images (Fig. 8) with zoomed-in vessel regions (yellow rectangle). Visually, Fig. 8 E showed high contrast with a smooth background, which was confirmed by the intensity profile of the zoomed-in region (enlarged inset and Fig. 8 H). Comparing Fig. 8 B-D and the corresponding enlarged regions revealed that multi-scale feature loss significantly suppressed the noise and artifacts, while BV-OD loss was able to enhance the vessels.

B. Results of the Dehaze Network

The ultimate goal of our study was to develop a method to dehaze images acquired from cataract patients. The efforts described to this point were intended to synthesize cataract-like images appropriate to train a network (CataractDehazeNet) via supervised learning.

CataractDehazeNet was based on the pix2pix network, but with customized loss functions: multi-scale feature loss and
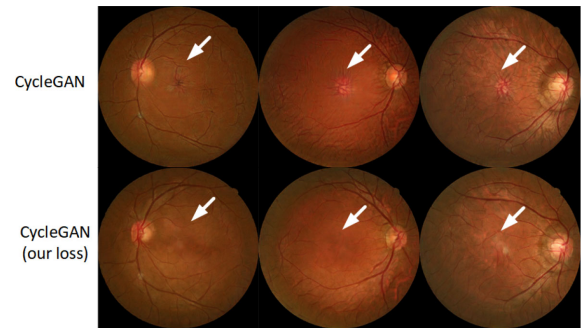

Fig. 9. The designed loss functions could significantly suppress the artifacts. The typical dehazed results using CycleGAN (first row) and the CycleGAN with our loss functions (second row), with arrows pointing to the artifacts (CycleGAN) and the corresponding regions (CycleGAN with the designed losses).
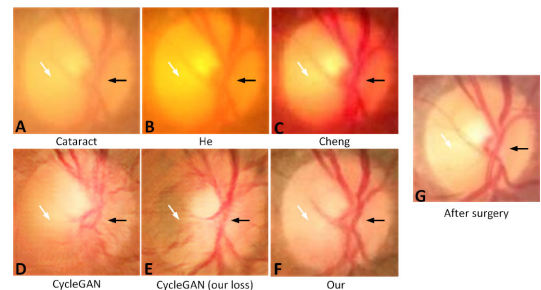

Fig. 10. The visual dehaze results for cataract images by the different approaches in the OD region.

BV-OD loss. Table IV demonstrated the necessity of the customized loss functions. From the table, we could see that either multi-scale feature loss or BV-OD loss improved the network performance slightly, as shown by the SSIM, but with a slight decrease with PSNR. However, the combination of these two showed significant improvement as measured by both the SSIM and PSNR.

C. Comparison With Previous Methods

We compared our CataractDehazeNet with previously published algorithms. We also used a U-net to segment both the BV and the OD, and utilized sensitivity, specificity, and dice as the metrics. In total, four algorithms were compared with our proposed method. The guided filter method for cataract image dehaze proposed by Cheng *et al.* [18] was referred to as "Cheng," and the dark channel dehaze method proposed by He [17] was referred to as "He". We also compared the unpaired CycleGAN method proposed by Zhu [34]. In addition, we applied our multi-feature loss and BV-OD loss in the CycleGAN to see whether it could improve upon its performance.

From Table V, it was clear that the proposed method outperformed the existing methods, as shown by both SSIM and PSNR. However, it was worth to note that adding the designed loss functions to the CycleGAN seemed to worsen the performance of the CycleGAN, as suggested by the decreased SSIM and PSNR. We visually investigated the images (Fig. 9) and found that the designed loss functions could significantly suppress
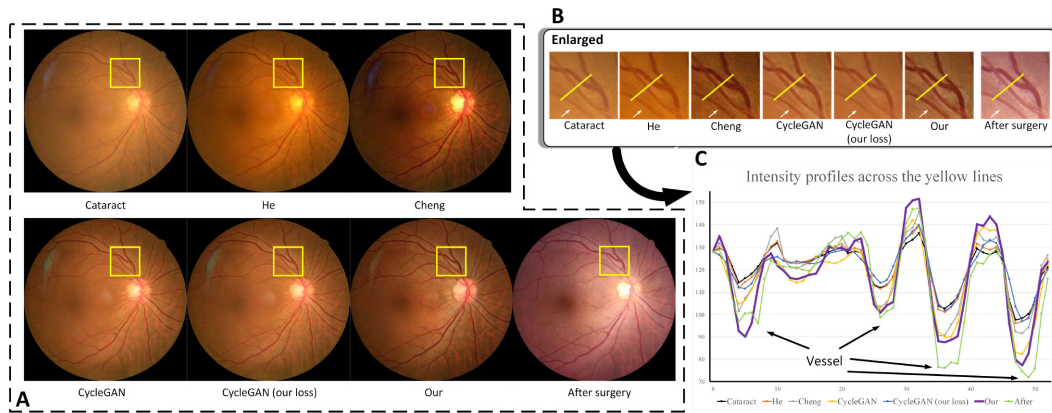
Fig. 11. Visual dehaze results of whole and enlarged portions of cataract images processed using different algorithms. A is dehaze images corresponding to each cataract image. B is partial enlarged images which located in the yellow rectangle marked in the whole image. C is corresponding intensity profiles across the yellow line in the partial enlarged images.

TABLE V
SSIM AND PSNR OF THE DEHAZE RESULTS IN THE OD REGION COMPARED WITH THOSE OF OTHER ALGORITHMS

Algorithms	SSIM	PSNR
He	0.771($\pm$0.092)	18.798($\pm$4.719)
Cheng	0.717($\pm$0.103)	16.337($\pm$4.304)
CycleGAN	0.667($\pm$0.105)	18.562($\pm$3.877)
CycleGAN (our loss)	0.633($\pm$0.119)	17.817($\pm$3.924)
Our	**0.781($\pm$0.079)**	**20.683($\pm$3.878)**

TABLE VI
SEGMENTATION OF BV AND OD ON IMAGES FROM THE TEST DATASET DEHAZED USING THE VARIOUS ALGORITHMS (SE:SENSITIVITY, SP:SPECIFICITY)

Algorithms		Cataract	He	Cheng	CycleGAN	CycleGAN (our loss)	Ours
	Se	0.528	0.518	0.511	0.679	0.636	**0.744**
BV	Sp	**0.994**	**0.994**	**0.994**	0.959	0.944	0.989
	Dice	0.646	0.638	0.628	0.554	0.472	**0.762**
	Se	0.947	0.940	0.939	0.730	0.718	**0.947**
OD	Sp	0.998	0.998	0.998	0.999	0.999	0.998
	Dice	0.928	0.923	0.927	0.809	0.809	**0.932**

Fig. 12. Dehaze results of cataract images with lesions using CataractDehazeNet.

the artifacts, particularly at the macular region (Fig. 9 white arrows). Since the artifacts mainly appeared in the macular region, evaluations of SSIM and PSNR at the OD regions could not pick up the differences.

Fig. 10 showed the typical dehazed results using different methods in the OD region, which has the most complicated structures to better demonstrate the subtle effects. The contrast of images processed using He's and Cheng's algorithms was lower than that in images processed using the CycleGAN, CycleGAN with our proposed loss functions, and our proposed CataractDehazeNet method. Compared to the CycleGAN, and CycleGAN with our losses, the proposed method rendered much less noise (white arrow), much sharper and more natural-looking contours (black arrow).

Whole-retina images with zoomed-in regions (areas within yellow squares) processed using the different algorithms are shown in Fig. 11. We also compared the results of the algorithms with the clear image from the same patient after surgery. Our
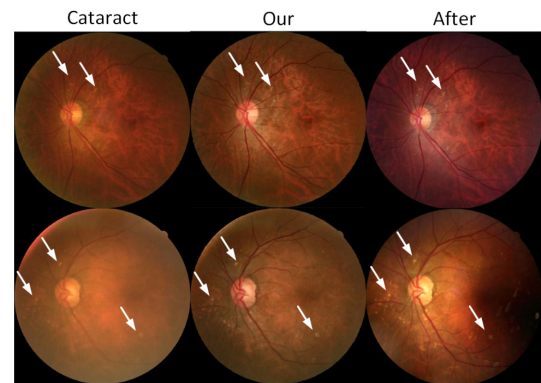
proposed method rendered an image with high contrast, less noise, and fewer artifacts, as shown by the purple line profile (Fig. 11 C).

Segmentation of BV is an important task. We examined how the proposed dehaze algorithm can improve BV and OD segmentations. We manually segmented all of the retinal images in the test dataset and used the U-net network to segment the OD and BV. In comparison with other methods, the proposed algorithm resulted in the highest dice coefficients for both OD and BV segmentations (Table VI), further suggesting the superior performance of the proposed algorithm.

D. Effect of the Dehaze Algorithm on Lesions in Cataract Images

Small lesions, such as exudates and hemorrhages, are important pathological features. Thus, it is crucial that the dehaze algorithm does not affect these lesions. Fig. 12 showed that our dehaze algorithm effectively removed haze and enhanced the contrast of the image. The pathological lesions became much clearer in the dehaze images and had good correspondence with those in images acquired from the same patients after surgery (white arrows).
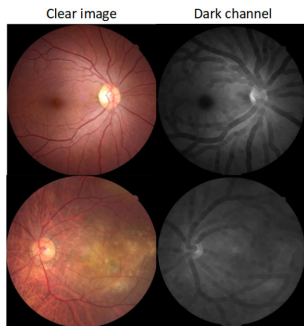
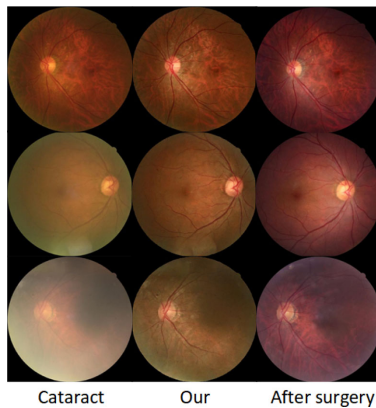Fig. 13. The dark channels of two typical clear retinal images.

Fig. 14. The dehaze images using the proposed method on images of different cataract severeness.

E. Discussion

In the field of natural image processing, the dehaze method based on the dark channel prior has achieved good results in removing haze to obtain clearer images. However, this method has limited performance when the dark channel assumption does not hold. Fig. 13 showed the corresponding "dark channel" of two typical clear retina images. It is clear that the intensity values are not close to zero, particularly at the OD regions. Therefore, we could not use Eq. (3) to calculate $t(x)$, the critical parameters for dehazing. Another impact came from L^c. Because the brightest region of the image is the OD region, $t(x)$ calculated by Eq. (3) would be or close to zero in the OD regions, making it fail to dehaze in the regions. This could also be the reason that Cheng's algorithm [18] is outperformed by the proposed method.

CycleGAN has been widely applied for unpaired image-image translation. It uses the cycle consistency loss function to train the network based on unpaired image datasets. However, when applied to retinal images, it introduces small, vessel-like artifacts and interrupts the internal structure of the OD region (as shown in Figs. 10 and 11). A possible reason could be that the loss functions in the CycleGAN are not able to provide enough constrains to suppress vascular artifacts. The artifacts are subtle and only count for a very small portion of the whole image. The loss functions in the CycleGAN constrain the input and output images to be similar on the whole-image level, but fail at the small structures. Applying different loss functions, such as the customized multi-scale feature and BV-OD loss functions, could

help to improve the performance. However, CycleGAN already consists of two GANs with total six loss functions (two GAN losses, two cycle-consistent losses and two identity mapping losses). Adding the customized loss functions would make it to ten. When the number of hyper-parameters is high, it is much more difficult to achieve the optimum. In addition, the two GANs in the CycleGAN were trained simultaneously, which make it more difficult to achieve the optimum. In contrast, we created two independent networks: one was used to generate the cataract-like image (CataractSimGAN) and one to dehaze cataract images (CataractDehazeNet). This approach promoted the convergence of the network and achieved satisfactory results (Fig. 14). With the generated paired images, the choice of paired DCNNs could also be extended, and the training of a paired DCNN is easier than that of an unpaired DCNN. We compared the results of our approach with those of previous state-of-the-art approaches and showed that our method offers the superior results.

V. Conclusion

Retinal images acquired from cataract patients are hazy due to opacification of the lens. To enhance the contrast of such images, we constructed an unpaired GAN, CataractSimGAN, to generate cataract-like retinal images. Using the synthesized paired cataract images generated using CataractSimGAN, we trained CataractDehazeNet, which was then used to dehaze unseen real cataract images. We also designed two specific loss functions for retinal images to improve the effectiveness of these two DCNNs. Compared with previous proposed methods, our algorithm obtained clearer blood vessel and optic disc structures, and substantially suppressed the artifacts. The proposed algorithm can not only provide visually better images but also can help physicians to diagnose retinal diseases in patients with cataracts. Most of the patients with cataracts are elders who may very likely have other retinal diseases, such as wet age-related macular degeneration (AMD). Diagnosis of the retinal diseases before cataract surgery enables the physician to provide a more comprehensive treatment plan, for example, intravitreal injection of anti-VEGF to treat AMD during the surgical process. Due to the prevalence of cataract patients, we firmly believe that the proposed dehaze algorithm can help the ophthalmologist to see the retinas of the cataract patients better, and therefore provide better care and treatments.

References

[1] P. A. Asbell, I. Dualan, J. Mindel, D. Brocks, M. Ahmad, and S. Epstein, "Age-related cataract," *Lancet*, vol. 365, no. 9459, pp. 599–609, 2005.
[2] D. Pascolini and S. P. Mariotti, "Global estimates of visual impairment: 2010," *Brit. J. Ophthalmol.*, vol. 96, no. 5, pp. 614–618, 2012.

[3] W. H. Organization, "Blindness and visual disability: Major causes worldwide," *WHO Fact Sheet*, vol. 143, 1999.

[4] E. Peli and T. Peli, "Restoration of retinal images obtained through cataracts," *IEEE Trans. Med. Imag.*, vol. 8, no. 4, pp. 401–406, Dec. 1989.

[5] L. Guo, J.-J. Yang, L. Peng, J. Li, and Q. Liang, "A computer-aided healthcare system for cataract classification and grading based on fundus image analysis," *Comput. Ind.*, vol. 69, pp. 72–80, 2015.

[6] X. Xu, L. Zhang, J. Li, Y. Guan, and L. Zhang, "A hybrid global-local representation CNN model for automatic cataract grading," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 556–567, Feb. 2020.

[7] R. H. Webb, G. W. Hughes, and F. C. Delori, "Confocal scanning laser ophthalmoscope," *Appl. Opt.*, vol. 26, no. 8, pp. 1492–1499, 1987.

[8] A. Roorda, F. Romero-Borja, W. J. Donnelly III, H. Queener, T. J. Hebert, and M. C. Campbell, "Adaptive optics scanning laser ophthalmoscopy," *Opt. Express*, vol. 10, no. 9, pp. 405–412, 2002.

[9] D. X. Hammer, R. D. Ferguson, C. E. Bigelow, N. V. Iftimia, T. E. Ustun, and S. A. Burns, "Adaptive optics scanning laser ophthalmoscope for stabilized retinal imaging," *Opt. Express*, vol. 14, no. 8, pp. 3354–3367, 2006.

[10] J. Fischer, T. Otto, F. Delori, L. Pace, and G. Staurenghi, "Scanning laser ophthalmoscopy (SLO)," in *Proc. High Resolution Imag. Microscopy Ophthalmol.*, 2019, pp. 35–57.

[11] H. Yu, C. Agurto, S. Barriga, S. C. Nemeth, P. Soliz, and G. Zamora, "Automated image quality evaluation of retinal fundus photographs in diabetic retinopathy screening," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, 2012, pp. 125–128.

[12] A. W. Setiawan, T. R. Mengko, O. S. Santoso, and A. B. Suksmono, "Color retinal image enhancement using clahe," in *Proc. IEEE Int. Conf. ICT Smart Soc.*, 2013, pp. 1–3.

[13] 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.

[14] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vision*, vol. 48, no. 3, pp. 233–254, 2002.

[15] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–9, 2008.

[16] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Conf. Comput. VisionPattern Recognit.*, 2008, pp. 1–8.

[17] 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.

[18] 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.

[19] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[20] A. Raj, A. K. Tiwari, and M. G. Martini, "Fundus image quality assessment: Survey, challenges, and future scope," *IET Image Process.*, vol. 13, no. 8, pp. 1211–1224, 2019.

[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[22] X. Yang, H. Li, Y.-L. Fan, and R. Chen, "Single image haze removal via region detection network," *IEEE Trans. Multimedia*, vol. 21, no. 10, pp. 2545–2560, Oct. 2019.

[23] A. Dudhane, H. S. Aulakh, and S. Murala, "RI-GAN: An end-to-end network for single image haze removal," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, pp. 1–10, 2019.

[24] Z. Shen *et al.*, "Human-aware motion deblurring," in *Proc. IEEE Int. Conf. Comput. Vision*, 2019, pp. 5572–5581.

[25] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, "Deep semantic face deblurring," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 8260–8269.

[26] J. Mao *et al.*, "Automated diagnosis and quantitative analysis of plus disease in retinopathy of prematurity based on deep convolutional neural networks," *Acta Ophthalmologica*, vol. 98, no. 3, pp. e339–e345, May 2020.

[27] R. Zheng *et al.*, "Detection of exudates in fundus photographs with imbalanced learning using conditional generative adversarial network," *Biomed. Opt. Express*, vol. 9, no. 10, pp. 4863–4878, 2018.

[28] H. Fu *et al.*, "A deep learning system for automated angle-closure detection in anterior segment optical coherence tomography images," *Amer. J. Ophthalmology*, vol. 203, pp. 37–45, 2019.

[29] H. Fu *et al.*, "Angle-closure detection in anterior segment OCT based on multilevel deep network," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2019.2897162.

[30] M. Pekala, N. Joshi, T. A. Liu, N. M. Bressler, D. C. DeBuc, and P. Burlina, "Deep learning based retinal OCT segmentation," *Comput. Biol. Med.*, vol. 114, 2019, Art. no. 103445.

[31] H. Li *et al.*, "Dense deconvolutional network for skin lesion segmentation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 527–537, Mar. 2019.

[32] S. Zhang *et al.*, "Simultaneous arteriole and venule segmentation of dual-modal fundus images using a multi-task cascade network," *IEEE Access*, vol. 7, pp. 57 561–57 573, 2019.

[33] S. Wang *et al.*, "Unsupervised deep learning features for lung cancer overall survival analysis," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2018, pp. 2583–2586.

[34] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 2223–2232.

[35] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[36] X. Yang, Y. Lin, Z. Wang, X. Li, and K.-T. Cheng, "Bi-modality medical image synthesis using semi-supervised sequential generative adversarial networks," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 3, pp. 855–865, Mar. 2020.

[37] T. Zhang *et al.*, "Noise adaptation generative adversarial network for medical image analysis," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1149–1159, Apr. 2020.

[38] Z. Shen, H. Fu, J. Shen, and L. Shao, "Understanding and correcting low-quality retinal fundus images for clinical analysis," 2020, *arXiv:2005.05594*.

[39] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 2107–2116.

[40] K. Leutwein and H. Littmann, "The fundus camera," *Refraction and Clinical Optics*. New York, NY, USA: Harper & Row, 1980, pp. 457–466.

[41] C.-H. Hsieh, Q. Zhao, and W.-C. Cheng, "Single image haze removal using weak dark channel prior," in *Proc. IEEE 9th Int. Conf. Awareness Sci. Technol.*, 2018, pp. 214–219.

[42] H. Hong, Z. Liu, Y. Shi, and X. Hua, "A fast method for single image haze removal based on multiscale dark channel prior," in *Proc. MIPPR Parallel Process. Images Optim. Techn.; Med. Imag.*, 2018, vol. 10610, Art. no. 1061002.

[43] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.

[44] A. Dudhane and S. Murala, "Cˆ2MSNet: A novel approach for single image haze removal," in *Proc. IEEE Winter Conf. Appl. Comput. Vision*, 2018, pp. 1397–1404.

[45] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vision*, 2016, pp. 154–169.

[46] H. Zhao, B. Yang, L. Cao, and H. Li, "Data-driven enhancement of blurry retinal images via generative adversarial networks," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2019, pp. 75–83.

[47] D. Engin, A. Genç, and H. K. Ekenel, "Cycle-dehaze: Enhanced cyclegan for single image dehazing," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2018, pp. 825–833.

[48] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, and S. Belongie, "Stacked generative adversarial networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 5077–5086.

[49] T. Zhang *et al.*, "SkrGAN: Sketching-rendering unconditional generative adversarial networks for medical image synthesis," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2019, pp. 777–785.

[50] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 1125–1134.

[51] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2015, pp. 234–241.

[52] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[53] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[54] H. Fu *et al.*, "Evaluation of retinal image quality assessment networks in different color-spaces," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2019, pp. 48–56.