An Annotation-free Restoration Network for Cataractous Fundus Images

Heng Li, Haofeng Liu, Yan Hu, Huazhu Fu, Yitian Zhao, Hanpei Miao, Jiang Liu

Abstract— **Cataracts are the leading cause of vision loss worldwide. Restoration algorithms are developed to improve the readability of cataract fundus images in order to increase the certainty in diagnosis and treatment for cataract patients. Unfortunately, the requirement of annotation limits the application of these algorithms in clinics. This paper proposes a network to annotation-freely restore cataractous fundus images (ArcNet) so as to boost the clinical practicability of restoration. Annotations are unnecessary in ArcNet, where the high-frequency component is extracted from fundus images to replace segmentation in the preservation of retinal structures. The restoration model is learned from the synthesized images and adapted to real cataract images. Extensive experiments are implemented to verify the performance and effectiveness of ArcNet. Favorable performance is achieved using ArcNet against state-of-the-art algorithms, and the diagnosis of ocular fundus diseases in cataract patients is promoted by ArcNet. The capability of properly restoring cataractous images in the absence of annotated data promises the proposed algorithm outstanding clinical practicability.**

Index Terms— **Cataracts, fundus image restoration, high-frequency component, domain adaptation.**

I. INTRODUCTION

CATARACTS, the leading cause of vision loss in the world, account for 33.4% blindness and 18.4% moderate-to-severe visual impairment [1], [2]. Surgery is the most cost-effective and commonly used treatment of cataracts and has been used to cure more than 30 million people worldwide

H. Li, Y. Hu, H. Liu, H. Miao are with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: lih3, huy3@sustech.edu.cn; 12032880, miaohp@mail.sustech.edu.cn). H. Li and H. Liu contributed equally to this manuscript.

Y. Zhao is with the Cixi Institute of Biomedical Engineering, Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, Ningbo 315201, China (e-mail: yitian.zhao@nimte.ac.cn).

H. Fu is with the Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore. (e-mail: hzfu@ieee.org).

J. Liu is with Research Institute of Trustworthy Autonomous Systems, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: liuj@sustech.edu.cn).

Corresponding authors: Y. Hu, J. Liu.

Code of this study is available at https://github.com/liamheng/Restoration-of-Cataract-Images-via-Domain-Adaptation.
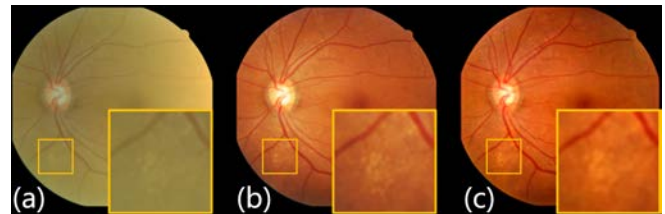
Fig. 1. Illustration of restoration for the cataractous image. (a) a cataractous fundus image before surgery. (b) the corresponding image after surgery. (c) the image restored from the cataractous one by ArcNet. The restoration enhances the abnormality in yellow boxes.

in 2020 [3]. Although the technology of cataract surgery has gradually advanced recently, treating cataract patients with ocular comorbidities is still cumbersome for the clinician. For instance, due to the sensitivity reduction and mesopic visual function resulting from glaucoma, intraocular lenses transplanted in cataract surgery might cause significant vision disturbances in patients with glaucoma [4]. For patients with diabetic retinopathy, cataract surgery often increases the risk of macular edema [5], which is a major cause of vision loss after cataract surgery [6]. Therefore, a complete and comprehensive preoperative examination is crucial in cataract surgery to discriminate ocular pathologies that could negatively influence the postoperative outcomes [7]. However, fundus assessment of cataract patients is a challenging task for clinicians. Since fundus images are captured through the lens, cataracts will attenuate and scatter the light passing through the lens degrading a photograph. To make matters worse, unlike the degradation caused by environmental noise, the degradation from cataracts is a pathological characteristic, which is unavoidable in fundus images. Hence, the restoration of fundus photographs taken through cataracts is clinically valuable [8]. As the example shown in Fig. 1, cataracts haze the fundus abnormalities before surgery, compared to the fundus image taken after surgery. And the proposed restoration algorithm enhances the observation of the abnormality in the cataractous fundus image.

To clearly visualize the fundus for reliable assessment and treatment, extensive work has long been done to enhance the cataractous fundus image [9]. According to the imaging principle of fundus photography, Peli et al. [10] developed an optical model for imaging the fundus through cataracts and proposed a restoration scheme for cataractous images using a homomorphic Wiener filter. Contrast enhancement is an efficient paradigm to promote the readability of images, by limiting contrast amplification, contrast limited adaptive

histogram equalization (CLAHE) has been presented [11] and introduced to restore degraded fundus images [12], [13]. For boosting the observations of the fundus, promising denoizing algorithms, such as dark channel prior [14] and guided image filtering (GIF) [15], have been modified to restore the fundus images obtained through cataracts [16]. However, due to the complexity of the imaging process of fundus photography and the mechanism of deterioration, it is challenging for these hand-crafted algorithms to preserve pathological characteristics and restore fundus images, simultaneously. The recent advancement in deep learning techniques has facilitated the application of convolutional neural networks (CNN) in image restoration [17]–[19], and specific approaches have been proposed to improve the quality of degraded fundus images [20]–[22].

Although these restoration approaches have achieved remarkable performance, practical difficulties appear when introducing them to the cataractous images. a) Collecting restoration annotations for cataractous fundus images, i.e. cataract-free images from identical patients, is costly and impractical in clinics. Therefore, the restoration algorithms demanding supervised data are unsuitable for cataractous images. b) Structure preservation by segmentation provides sub-optimal results. Preserving structures by segmentation [21], [22] not only aggravates annotation costs, but even worse, leads to missing unannotated pathological characteristics in restorations. c) These learned restoration models suffer significant performance drop encountering domain shifts. Domain shifts, referring to the changes in the data distribution between the training dataset and the dataset to deploy, are common and troublesome in clinics.

To develop a practical restoration scheme for cataractous images, an end-to-end network is proposed in this paper to learn the restoration model of clinical cataractous images without any annotations. Primarily, a source domain of high-low quality paired data is synthesized by simulating cataract-like images from publicly available data for the training of the restoration network. Subsequently, the segmentation annotation is replaced by the structure guidance of the high-frequency component (HFC), which is acquired from the fundus images, to preserve retinal structures during the restoration. Then a restoration network is designed for the fundus images obtained through cataracts using image-to-image translation and unsupervised domain adaptation. Finally, the comparisons with state-of-the-art algorithms and the applications for promoting fundus assessment were presented, while cataractous fundus images from four datasets were employed to demonstrate the performance of the proposed network. The proposed approach is significantly modified from our previous work published in ISBI-2021 [23], where the simulation model for cataracts has been improved and the HFC has replaced the Sobel operator as the structure guidance. The main contributions of this paper are summarized as follows:

1) A network for annotation-free restoration of cataractous fundus images (ArcNet) is proposed to promote the retina examination for cataract patients. Considering the acquirement of any annotation is impractical in clinical scenarios, ArcNet presents a practical solution for fundus restoration.

2) Instead of segmentation, the HFC is extracted from fundus images to guide the restoration, where fundus characteristics are automatically and comprehensively preserved.

3) The restoration model learned from the synthesized data is properly generalized to the cataractous images by unsupervised domain adaptation.

4) Extensive experiments validated the promising performance of the proposed ArcNet. Compared to the state-of-the-art restoration algorithms, ArcNet achieved favorable performance, and its effectiveness was demonstrated in quality restoration and diagnosis boosting.

II. RELATED WORK

Fundus photography, a routine examination for fundus assessment, tends to experience quality degradation in the attendance of cataracts, bias illumination, or unskilled operation. Since Peli et al. verified the feasibility and clinical value for enhancing the fundus images [8], and proposed an optical model for imaging the fundus photography through cataracts [10], restoration algorithms have been developed to analyze fundus images better.

A. Fundus image enhancement

CLAHE is an efficient technique for contrast enhancement and has been extensively applied to retinal fundus images [13], [24], [25]. Mitra et al. [12] combined CLAHE with Fourier transform to enhance the contrast of cataractous images. Alternatively, to enhance retinal fundus images, efforts have been made to estimate and compensate for the noise in imaging models. He [15] proposed GIF to filter images as an edge-preserving smoothing operator and efficiently compute haze removed outputs. Cheng et al. [16] developed a structure-preserving guided retinal image filtering (SGRIF), which modified GIF to preserve the structure in retinal images. Frequency domain approaches are another paradigm of image enhancement. Low-pass frequency filtering and α-rooting were conducted to improve the contrast of the retinal structures [26]. Although these methods perform efficiently, their heavy dependence on global image statistics leads to severe constraints in implementation.

B. CNN-based restoration algorithms

Recently, owing to the advantage in image representation, CNN has become the most fashionable algorithm in the community of computer vision. Methods for promoting image quality have rapidly advanced based on CNN through image enhancement [17], [27], dehazing [18], [28], and denoizing [19], [29]. Unfortunately, due to the differences between medical and natural images, the fundamental assumption of these schemes may not hold in fundus images. As preserving retinal structures is significant in retinal image analysis, pixel-wise translation has thus been introduced to retinal fundus image enhancement. The networks applied to retinal image enhancement are composed of two categories: paired [30]

and unpaired [31] translation. As the paired ones require image pairs to learn a mapping from one representation to another, degradation was simulated to build retinal image pairs of high-low quality [21], [32]. CycleGAN utilizing unpaired data to model the translation, is one of the most frequently-used translation networks [33] and has been implemented in the restoration of fundus images [20], [34]. Further, for preserving retinal structures, segmentation was introduced to enhance retinal structures. A clinical-oriented fundus enhancement network (CofeNet) was proposed [21] to preserve the retinal structures in restoration according to segmentation outputs. Luo et al. [22] reported a two-stage dehazing algorithm, which restores cataractous fundus images under the supervision of segmentation. However, annotations of restoration and segmentation are necessary to implement these algorithms.

C. Domain adaptation

Most statistical learning algorithms strongly depend on the assumption that the source (train) and target (test) data are independent and identically distributed [35], while ignoring the out-of-distribution (OOD) scenarios commonly in practice. Due to the domain shift between the source and target domain, a model learned from source data will suffer significant performance drops on an OOD target domain [36]. Domain adaptation aims to bridge the shift between the source and target domain [37], [38], such that the model learned from the source domain is able to perform decently on the target data.

Domain adaptation algorithms are divided into three categories. The first one attempts to learn a mapping between source and target domains [39], [40]. The second one tries to find a shared latent space for source and target distributions [37], [41]. And the third one regularizes a classifier trained on a source distribution to work well on a target distribution [36], [42]. Supervision is necessary for the third one, while unsupervised domain adaptation requiring no annotated target data always follows the first two categories. Recently, with the advance of deep learning techniques, adversarial learning has been extensively employed to align source and target distributions in the image [39], output [43], or latent space [37].

Unsupervised domain adaptation is imported to achieve an annotation-free restoration network in this study. Cataract-like images are first simulated to build a source domain close to the target one, and then a shared network is implemented with unsupervised domain adaptation to generalize the restoration model from the source to the target domain.

III. METHOD

To mitigate the data requirement of annotations, in this study automatically captured structure guidance and unsupervised domain adaptation are introduced to restore unannotated cataractous fundus images.

An overview of the proposed ArcNet is exhibited in Fig. 2 (a). High-low quality paired data are synthesized for training the restoration model by simulating cataract-like images from any available clear ones. HFC in fundus images is extracted as

structure guidance to preserve retinal structures during restoration. Image-to-image translation and unsupervised domain adaptation are cooperated to learn the restoration model from the source domain of simulated data and adapt to the target domain of real images. Thus, the unannotated cataractous images can be unsupervisedly restored.

A. Cataract-like image simulation

According to the imaging model proposed in [10], the fundus image I degraded by cataracts is composed of retinal reflectance information and light reflected back from the cataract:

$$I(i,j) = \alpha \cdot L \cdot \gamma(i,j) \cdot \tau(i,j) + L(1 - \tau(i,j)), \quad (1)$$

where (i,j) indicates pixels in the image, α is the attenuation constant of retinal illumination. L is the illumination of the fundus camera, $\gamma(i,j)$ and $\tau(i,j)$ respectively denote the retinal reflectance function and the transmission function of the lens. Approaches to estimate $\tau(i,j)$ were proposed in previous studies [15], [16], but their application scenarios are quite limited in restoration.

The source domain of high-low quality image pairs is synthesized by simulating cataract-like images, to learn the restoration model for cataractous images. Reformulate Eq. 1 as

$$I(i,j) = \alpha \cdot L \cdot \gamma(i,j) + (1 - \tau(i,j))(L - \alpha \cdot L \cdot \gamma(i,j)). \quad (2)$$

Denote the ideal clear fundus image $L \cdot \gamma(i,j)$ as $s(i,j)$,

$$I(i,j) = \alpha \cdot s(i,j) + (1 - \tau(i,j))(L - \alpha \cdot s(i,j)). \quad (3)$$

As the uneven transmission function of cataract lens, a cataractous panel J is employed to model $(1 - \tau(i,j))$, where with the center of (a,b) by $J_{ij} = \sqrt{(i-a)^2 + (j-b)^2}$. Considering the disparity illumination intensity of the fundus under different spectral illuminations [44], decompose Eq. 2 into RGB channels. The cataract-like image simulation is given by:

$$C(s_c) = \alpha \cdot s_c * g_B(r_B, \sigma_B) + \beta \cdot J * g_L(r_L, \sigma_L) \cdot (L_c - s_c), \quad (4)$$

where s_c is the clear image, $c \in \{r, g, b\}$ refers to the red, green or blue channel of the image, and $*$ denotes the convolution operator. α and β denote the weight for the clear fundus image and the noise from cataracts. g_B and g_L are the Gaussian filters respectively for clear image smooth and cataractous panel, where $g(r, \sigma)$ represents a Gaussian filter with a radius of r and spatial constant σ. The highest intensity in s_c is selected as L_c. The simulated cataract-like image is defined as $s' = [C(s_r), C(s_g), C(s_b)]$.

Fig. 2 (b) demonstrates the simulation of cataracts from clear fundus images using Eq. 4. The source domain of high-low quality paired data is thus formed by clear fundus images and synthesized cataract-like ones. Additionally, owing to the domain adaptation algorithm in the following part, the rigorous requirements are unnecessary for the similarity between the simulated and real cataractous ones.
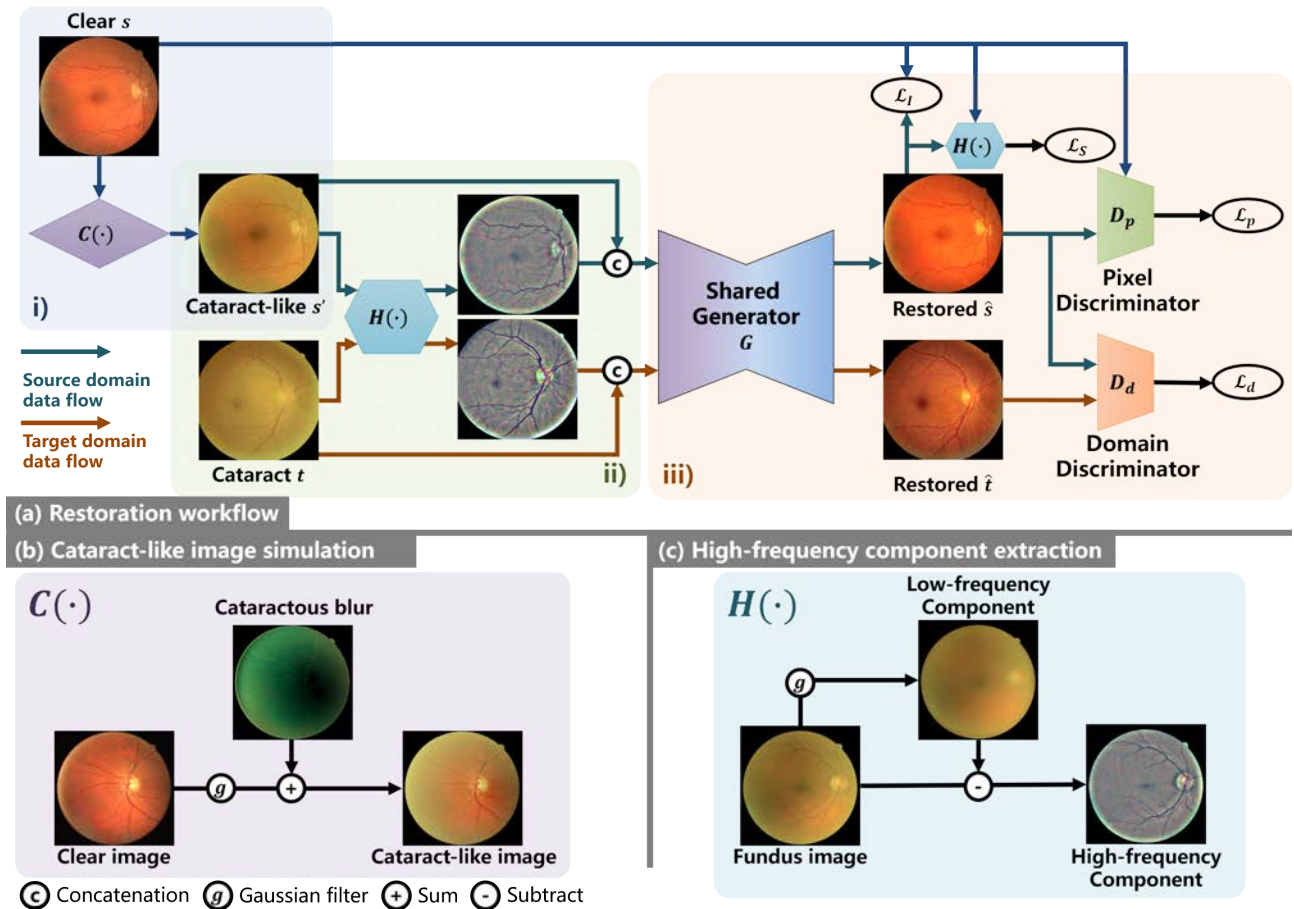
Fig. 2. (a) The overall workflow of the proposed restoration algorithm. $C(\cdot)$ and $H(\cdot)$ respectively refer to the simulation of cataract-like images in (b) and the extraction of the high-frequency component in (c). i) The cataract-like image s' is simulated from clear image s to synthesize the source domain of high-low quality image pairs. The target domain consists of the images to be restored. ii) Images from the source and target domain are simultaneously forwarded to the restoration network with the structure guidance of their high-frequency components. iii) $\mathcal{L}_p, \mathcal{L}_I, \mathcal{L}_S$ train the model to restore cataractous images, and $\mathcal{L}_d$ adapts the model from the source domain to the target domain. (b) Simulating cataract-like images from clear ones following Eq. 4. (c) According to Eq. 6, the HFC is extracted from the fundus image as the structure guidance in the restoration.

B. Structure guidance with HFC

Segmentation masks have been used in previous algorithms [21], [22] to preserve retinal structures in fundus image restoration. A segmentation network was implemented in [21] to inject retinal structural information into the restoration network by feature map concatenation. While a structure segmentation loss was proposed in [22] to preserve retinal structures, which was calculated from two pretrained U-nets. However, this preservation requires segmentation annotation and ignores the structures beyond the segmentation masks, such as pathological characteristics, which are fundamental in the fundus assessment. To address the disadvantage of segmentation-based structure preservation, the HFC is automatically extracted from the cataractous image as the structure guidance.

Inspired by the frequency space analysis for cross-domain tasks [38], [45], the domain-invariant features in frequency space are extracted to preserve fundus structures. Based on the retinex theory [46], the blur is caused by illumination bias, which is considered to be a low-frequency noise in the image. Further, as presented in Eq. 1, the light reflected from the cataract belongs to the low-frequency component (LFC) in

the fundus image I, and the reflection of the retina is intact in HFC. Thus it indicates that the HFC of fundus images is more robust to cataracts and contains most domain-invariant structure features. Decompose the cataractous fundus image into two additive portions of LFC $L(I)$ and HFC $H(I)$:

$$I = L(I) + H(I). \tag{5}$$

Therefore, HFC of cataractous images is employed to preserve retinal structures during the restoration. As presented in Fig. 2 (c), LFC can be extracted efficiently using a low-pass Gaussian filter, and HFC is calculated from:

$$H(I) = I - I * g_P(r_P, \sigma_P), \tag{6}$$

where the low-pass Gaussian filter g_P with radius r_P and spatial constant σ_P is implemented to capture LFC from image I. Specifically, to guarantee that entire retinal structures are contained in HFC, the kernel size of the Gaussian filter should be sufficiently large. r_P was set as 26, and σ_P was calculated as 9 according to r_P [47].

With the structure guidance of HFC, the segmentation is replaced and retinal structures, especially pathological characteristics are comprehensively preserved.

C. Restoration with unsupervised domain adaptation

The synthesized high-low quality paired images are leveraged to train the restoration model, and HFC preserves retinal structures while removing cataracts from fundus images. Furthermore, unsupervised domain adaptation is implemented to generalize the restoration model from the synthesized source domain to the target domain of real cataracts.

1) Training restoration model: The paired image-to-image translation network with the generator of U-net is implemented to learn the restoration model from the synthesized source domain of paired images. To preserve retinal structures during the restoration, HFC is extracted from the cataract-like image s'. As expressed in Fig. 2(a), s' is forwarded to the generator G concatenating with the HFC as the structure guidance. The restored image $\hat{s}$ is given by

$$\hat{s} = G(s', H(s')). \tag{7}$$

For optimizing the generator G, threefold loss functions are defined to minimize. An adversarial loss $\mathcal{L}_p$ for G is quantified by D_p, a discriminator for the pixels from the restored image $\hat{s}$ and the clear one s, where

$$\mathcal{L}_p = \mathbb{E}\left[\log D_p(s)\right] + \mathbb{E}\left[\log\left(1 - D_p\left(G\left(s', H(s')\right)\right)\right)\right]. \tag{8}$$

L1 distance is also calculated between $\hat{s}$ and s, leading $\hat{s}$ to the ground truth.

$$\mathcal{L}_I = \mathbb{E}[\|s - G(s', H(s'))\|_1]. \tag{9}$$

To guarantee that the realistic and convincing contents are properly preserved, another L1 distance is measured between the HFCs of s and $\hat{s}$.

$$\mathcal{L}_S = \mathbb{E}[\|H(s) - H(\hat{s}))\|_1]. \tag{10}$$

Thus the restoration model is learned on the synthesized image pairs from the source domain by minimizing the loss function $\mathcal{L}_G$, where

$$\mathcal{L}_G = \mathcal{L}_p + \lambda_1 \mathcal{L}_I + \lambda_2 \mathcal{L}_S, \tag{11}$$

2) Unsupervised domain adaptation: The effectiveness of the restoration model strongly relies on an assumption, that the training and test data are independent and identically distributed. However, since the images for training are synthesized and the images to be restored are collected from clinics, the domain shift between the training and test data is inevitable in this study. Therefore, to effectively deploy the restoration model, unsupervised domain adaptation is introduced to transfer the restoration model from the synthesized to real images.

The unsupervised domain adaptation is implemented based on the outputs of the restoration model in the absence of annotation. As the aim of this study is to restore clear fundus images, our intuition is that properly restored images should belong to an identical domain no matter the cataract images are from the source or target domain. Consequently, as demonstrated in Fig. 2 (a) , another discriminator for domains, D_d, is adopted in the output space to conduct an unsupervised domain adaptation.

Specifically, the real cataractous images are employed as the target domain. The source domain and the target domain are simultaneously forwarded to the restoration model. The two domains share the generator G, and a discriminator D_d is used to import adversarial learning for domain adaptation. The loss of domain adaptation is subsequently calculated from the restoration result $\hat{s}$ of the source domain and $\hat{t}$ of the target one. The loss function is formulated as:

$$\mathcal{L}_d = \mathbb{E}\left[\log D_d\left(\hat{s}\right)\right] + \mathbb{E}\left[\log\left(1 - D_d\left(\hat{t}\right)\right)\right], \tag{12}$$

where $\hat{t} = G(t, H(t))$, and t represents real cataract fundus images. Append $\mathcal{L}_d$ into $\mathcal{L}_G$. During the training stage, to fool discriminator D_d, and the adversarial loss $\mathcal{L}_d$ leads G to generate $\hat{s}$ and $\hat{t}$, which shall share an identical distribution. Therefore, the restoration model adapts to the real cataractous images without supervision.

Finally, based on the described modules, the annotation-free restoration network for cataractous images is trained without any annotations using the loss function

$$\mathcal{L}_{total} = \mathcal{L}_p + \lambda_1 \mathcal{L}_I + \lambda_2 \mathcal{L}_S + \lambda_3 \mathcal{L}_d, \tag{13}$$

where λ_1, λ_2 and λ_3 are the weights used to balance the losses and were respectively set to 100, 50 and 1 in this paper.
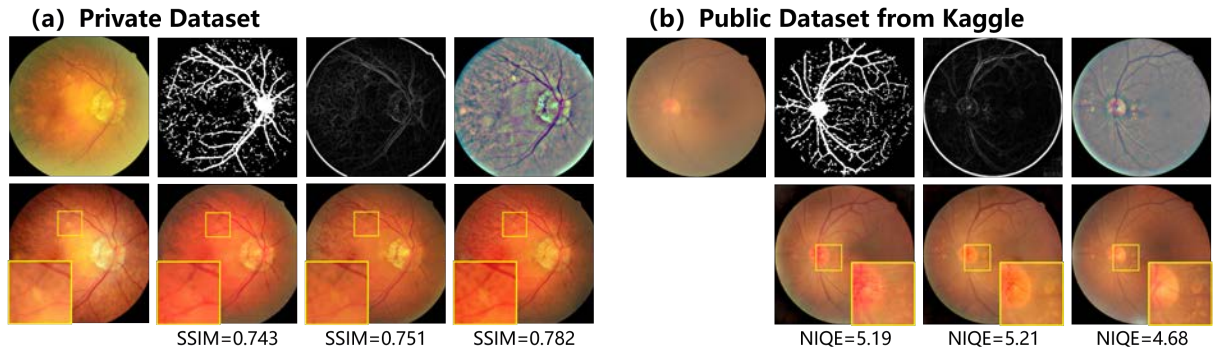
D. Network implementation

ArcNet adopts an Unet-like network as a generator and two patch-based fully convolutional networks as discriminators. The generator consists of 8 down-sampling layers and 8 up-sampling layers with skip connections between the corresponding layers. Each down-sampling layer contains a down convolution layer, a leaky ReLU layer, and a batch normalization layer. While each up-sampling layer contains a transposed convolution layer, a ReLU layer, and a batch normalization layer, and the output layer contains a ReLU layer, a convolution layer, and a Tanh function. The pixel discriminator and domain discriminator are with the same network architecture which adopts 5 convolutional layers with kernel size 4×4 and stride of $\{2, 2, 2, 1, 1\}$ for the channel number of $\{64, 128, 256, 512, 1\}$.

In the training phase, the input image size was set to 256×256 and the batch size was set to 8. To reduce the gap between the source domain and the target domain, the model loaded images randomly in a scale, including 286, 306, 326, and 346. After which it randomly cropped the images to the size of 256×256. All the networks were trained for 100 epochs using Adam optimizer with a learning rate equal to 2×10^{-4} for the 80 epochs and 5×10^{-5} for the next 20 epochs. During inference, the input image size was set to 256×256, and the batch size was set to 1.

The network was implemented in PyTorch, and was trained on an NVIDIA TITAN RTX with 24GB of memory. The number of model parameters is about 60 million.

IV. EXPERIMENTS

Five experiments were implemented to interpret the performance of ArcNet. The effectiveness of HFC and the proposed modules were first demonstrated. Then comparisons with state-of-the-art algorithms were implemented to verify ArcNet on the restoration of cataractous images. Subsequently,

(a) Private Dataset   **(b) Public Dataset from Kaggle**

SSIM=0.743 SSIM=0.751 SSIM=0.782   NIQE=5.19 NIQE=5.21 NIQE=4.68

Fig. 3. Restoration guided by segmentation, Sobel operator, and HFC. (a) Private dataset. (b) Public cataract subset of Kaggle. The images in the first row are successively the cataractous image, the structure guidance from segmentation, Sobel operator, and HFC. The second row exhibits the ground truth of clear fundus image (unavailable for Kaggle), the results of the restoration with the guidance of segmentation, Sobel operator, and HFC. The yellow boxes highlight pathological characteristics of drusen.

an investigation was provided to understand the impact on restoration from selecting source domains. An application of ArcNet was then presented to promote the diagnosis of ocular fundus diseases. Finally, the model setting and computational complexity of ArcNet were analyzed.

A. Experiment Setting

1) Datasets: Four fundus image datasets for two tasks were employed in the experiments to comprehensively verify the proposed ArcNet. The experimental protocol is in accordance with the Declaration of Helsinki and was approved by the local Ethics Committee.

As provided in Table I, the evaluation of restoration performances were conducted using three datasets: 1) a private dataset, RCF, consisting of 52 paired fundus images collected before and after cataract surgery from Peking University Third Hospital; two public fundus datasets were collected from 2) DRIVE[1] and 3) Kaggle[2]. To exhibit the effectiveness of ArcNet in diagnosis tasks, 4) the private ophthalmic dataset, Fundus-iSee, is used to diagnose ocular comorbidities in cataract patients.

TABLE I
DATASETS USED IN THE EXPERIMENTS

Tasks	Target domain datasets	Source domain datasets
Restoration	1) RCF: 26 images before surgery paired with 26 after surgery	2) DRIVE: 40 clear images with segmentation annotations
	3-1) Cataractous subset of Kaggle: 100 cataractous images	3-2) Normal subset of Kaggle: 300 clear images
Diagnosis	4) Fundus-iSee: 10000 fundus images from patients with (2669) and without (7331) cataracts are sorted into five categories according to fundus diseases	

2) Baseline: Seven state-of-the-art restoration algorithms for fundus images were conducted as the baseline. Three hand-crafted methods based on CLAHE [12], SGRIF [16], and image filtering [26], were leveraged to enhance the fundus images without training. In contrast, the image-to-image

[1] http://www.isi.uu.nl/Research/Databases/DRIVE/
[2] https://www.kaggle.com/jr2ngb/cataractdataset

translation networks of pix2pix [30] and CycleGAN [31] were trained on the source domain and tested on the target domain to restore cataract images. Moreover, the recent restoration approaches [21], [22] using segmentation annotations were also performed. Additionally, the source and target domain without annotations were forwarded to ArcNet during the training stage. Then the learned model was applied to restore the cataract images of the target domain.

3) Evaluation metrics: The annotation is only available in the private dataset because of the difficulties in collecting high-low quality paired data. Thus different evaluation metrics are selected to quantify the restoration performance according to the attendance and absence of the ground truth. For the annotated private dataset, the evaluation metrics of structural similarity (SSIM) and peak signal to noise ratio (PSNR) were used to quantify the restored image quality. While for the unannotated public dataset, the evaluation metrics of image quality, i.e., inception scores (IS) and natural image quality evaluator (NIQE) [20], [48], were adopted to assess the restoration performance.

Particularly, registration was applied between the paired images in the private dataset, and SSIM and PSNR were calculated from the overlap area of the paired images. The PSNR and SSIM values grow with an increase in the similarity between the restoration result and the ground truth. IS and NIQE quantify the image quality by evaluating the distortion, and the satisfying image quality is indicated by the high IS score and the low NIQE score.

B. Structure guidance for ArcNet

In previous restoration algorithms, structure segmentation and edge detection were imported to preserve retinal structures. In [21] and [22], segmentation networks were leveraged to provide retinal structural information and loss for restoration. Alternative to segmentation networks, the edge detected by the Sobel operator was utilized in our previous study [23] with an architecture same to ArcNet to preserve retinal structures during the restoration.

To demonstrate the advantage of the HFC and compare the proposed algorithm with the previous one [23], in this section, ArcNet is respectively implemented with structure guidances

of segmentation, Sobel operator, and HFC. Specifically, a U-net was trained on DRIVE with the segmentation annotations of vessels and optic disks. Thus the HFC is replaced by segmentation masks captured by the U-net as the structure guidance in ArcNet. And as our previous algorithm [23] shares an identical architecture with ArcNet, it is also presented and compared in the restoration with the Sobel operator.

The restoration was respectively conducted with the source-target domain groups of DRIVE and RCF, together with the normal and cataract subset of Kaggle. Fig. 3 exhibits the restoration results as well as the guidance of segmentation, Sobel operator, and HFC. Pathological characteristics of drusen are highlighted in the yellow boxes of the second row. As observed from the first row, the segmentation guidance only attends to the annotated structures, such that the unannotated pathological characteristics were lost in the second row. Moreover, due to the quality degradation caused by cataracts, the guidance suffers from segmentation errors. Though the Sobel operator is sensitive to the edge and forwards more pathological characteristics to the network, it leads our previous algorithm [23] to ignore region information. Thus artifacts appeared in the restored optic disk of Fig. 3 (b). HFC preserves most details of the retinal structures, such as the tiny particulars of the optic disc and retina. Consequently, the structures distorted in the restoration guided by segmentation and Sobel operator, were desirably preserved in the image restored with HFC.

TABLE II
COMPARISON ON RESTORATION RESULTS WITH DRIVE AND KAGGLE.

Guidance	DRIVE		Kaggle	
	SSIM ↑	PSNR ↑	IS ↑	NIQE ↓
Segmentation	0.738	18.05	1.46	5.12
Sobel [23]	0.755	18.07	1.50	5.32
HFC (ours)	**0.767**	**18.29**	**1.52**	**4.52**

The quantitative comparison is summarized in Table II. The evaluation metrics of SSIM and PSNR were calculated for the private dataset, and the metrics of IS and NIQE assessed the restoration of Kaggle. Due to the capacity of ArcNet, all structure guidances achieve decent scores in the evaluation metrics. Furthermore, superior performance was displayed using the proposed guidance of HFC. More meaningful details are forwarded to the network by HFC without annotation. Therefore, it indicates that HFC is more efficient in preserving the retinal structures during the restoration.

C. Ablation study of ArcNet

For illuminating the effectiveness of the modules in ArcNet, a comparison against the ablation of modules was conducted. DRIVE and RCF were still used as the source and the target domain here. The fundamental framework of ArcNet follows pix2pix [30], which in Fig. 2 includes the generator G, the adversarial loss $\mathcal{L}_p$ of the discriminator D_p, as well as the L1 distance $\mathcal{L}_I$. Subsequently, the proposed modules refer to: a) $H(\cdot)$ introduces the structure guidance of HFC to the restoration network; b) the structure loss $\mathcal{L}_S$ emphasizes

structure preservation in the restoration; c) the domain loss $\mathcal{L}_d$ of the discriminator D_d achieves unsupervised domain adaptation between the source and target domain in ArcNet.
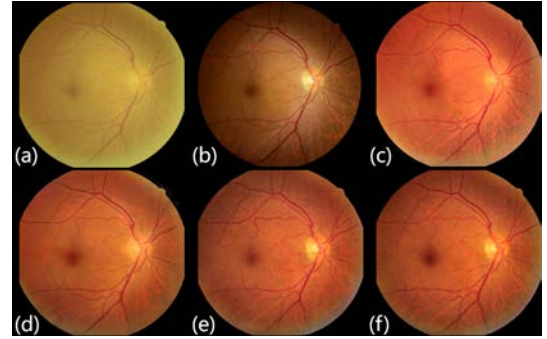

Fig. 4. The effect of the proposed modules. (a) cataract image. (b) the corresponding clear image (c) The fundamental network. (d) w/ $H(\cdot)$, HFC is imported to the restoration. (e) w/ $H(\cdot)$ and $\mathcal{L}_S$, structure preservation is emphasized. (f) w/ $H(\cdot)$, $\mathcal{L}_S$, and $\mathcal{L}_d$, unsupervised domain adaptation is further achieved in ArcNet. Zoom in for details.

The ablation study was conducted by individually deploying the modules to the fundamental network. $H(\cdot)$ was deployed first to import HFC, and then $\mathcal{L}_S$ promoted structure preservation, finally domain adaptation was presented by $\mathcal{L}_d$. Fig. 4 visualizes the results of the restoration, and quantitative evaluation is summarized in Table III.

Due to unsupervised restoration for the target domain, the gap between the intensity distributions of the ground truth (Fig. 4 (b)) and the restored image (Fig. 4 (c-f)) was inevitable. Therefore, our observation and analysis concentrate on content consistency. Compared with the result in Fig. 4 (c) restored by the fundamental network, HFC improves the preservation of retinal structures in Fig. 4 (d). In Fig. 4 (e), $\mathcal{L}_S$ further enhances the structures, especially in the optic disk area. Finally, in the comparison between Fig. 4 (e) and (f), domain adaptation by $\mathcal{L}_d$ promises ArcNet a restoration result with realistic textures and appearance. Moreover, the quality of the restored images gradually increased in Table III further demonstrating the proposed modules' effectiveness.

TABLE III
ABLATION STUDY. OUR FULL MODEL ACHIEVES REASONABLE RESULTS IN BOTH METRICS.

$H(\cdot)$	$\mathcal{L}_S$	$\mathcal{L}_d$	SSIM ↑	PSNR ↑
			0.733	17.69
√			0.755	17.83
√	√		0.763	17.96
√	√	√	**0.767**	**18.29**

D. Comparison with existing methods

To verify the advantages of ArcNet, restoration comparisons on cataractous images were conducted in this section. The comparisons were implemented on the private dataset with ground truth, as well as the unannotated public datasets. The state-of-the-art methods described in Section IV-A.2 were introduced as the baselines.
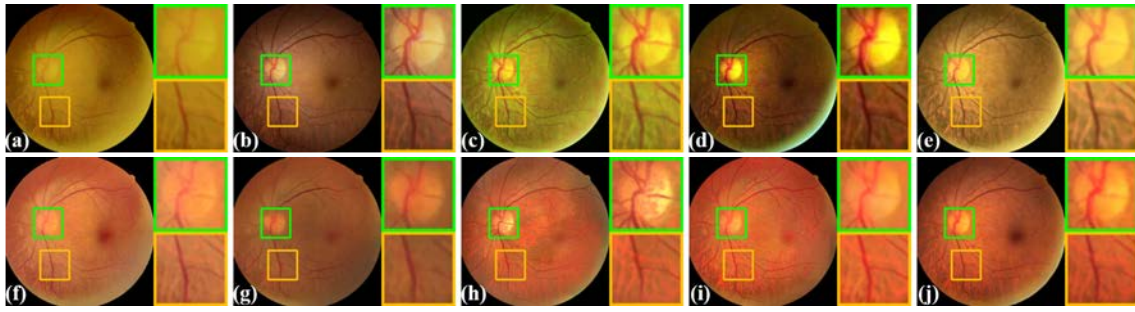
Fig. 5. Visual comparison of images restored from cataract ones. (a) cataract image. (b) clear fundus image after surgery. (c) Mitra et al. [12]. (d) SGRIF [16]. (e) Cao et al. [26]. (f) pix2pix [30]. (g) CycleGAN [31]. (h) Luo et al. [22]. (i) CofeNet [21]. (j) ArcNet (ours).
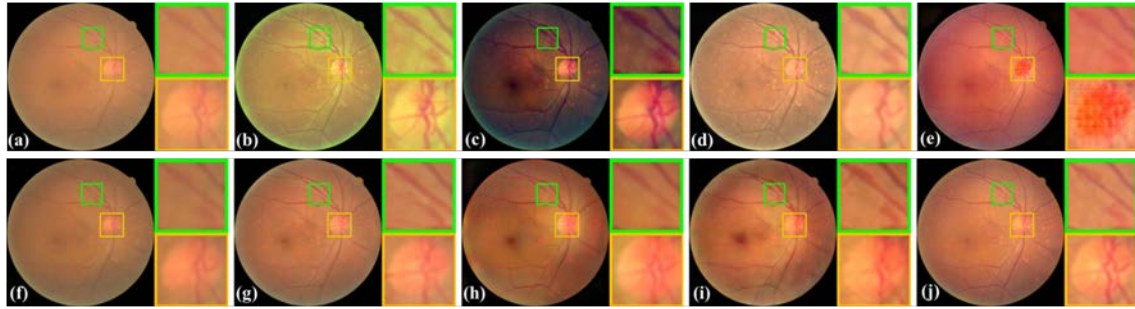

Fig. 6. Visual comparison of results for the public Kaggle dataset. (a) cataract image. (b) Mitra et al. [12]. (c) SGRIF [16]. (d) Cao et al. [26]. (e) pix2pix [30]. (f) CycleGAN [31]. (g) CycleGAN-real. (h) Luo et al. [22]. (i) CofeNet [21]. (j) ArcNet (ours).

1) Restoration of private dataset: The source domain of DRIVE and the target domain of RCF were employed to conduct this experiment. The visualized and quantitative comparisons against existing methods were summarized in Fig. 5 and Table IV. The proposed ArcNet achieves outstanding performance, and as shown in Fig. 5 the fundus images restored by ArcNet appear relatively realistic and clean details. Additionally, consistent with Table III, in Table IV the guidance of HFC and domain adaptation endows ArcNet with particularly favorable performance against the existing methods.

TABLE IV
QUANTITATIVE COMPARISON WITH STATE-OF-THE-ART METHODS ON THE PRIVATE DATASET. OUR METHOD PRODUCES SUPERIOR RESULTS IN THE METRICS WITH REFERENCE.

Algorithms	SSIM ↑	PSNR ↑
Mitra et al. [12]	0.698	15.82
SGRIF [16]	0.609	15.07
Cao et al. [26]	0.709	16.40
pix2pix [30]	0.731	17.68
CycleGAN [31]	0.732	17.25
Luo et al. [22]	0.704	17.29
CofeNet [21]	0.754	18.03
ArcNet (ours)	**0.767**	**18.29**

It can be observed from the results that the hand-crafted methods [12], [16], [26] fail to restore the fundus images in the appearance of ground truth. These methods are manually designed according to publicly distributed attributes of hazed fundus images, which often results in limited versatility to noise distributions. In contrast, the images restored using

the methods of image-to-image translation [30], [31], appear similar to the clear images. The paired training data allow image-to-image translation models to explore a particular mapping between the cataract-like and clear images. Segmentation masks were introduced to preserve retinal structures in [21], [22], and thus they are more sensitive to the vessels and optic disks.

However, the inherent difference between the cataract-like images from the source domain and real cataract images from the target domain leads the methods to alias the retinal structures, such that the details became blurry and of low contrast. ArcNet is capable of mapping the appearance of clear fundus images and enhancing the fine retinal structures.

2) Restoration of public data: The comparison is also performed on a dataset of a large volume to examine the proposed method. The dataset from Kaggle was employed to conduct this experiment, where the source domain was synthesized from the normal subset, and the cataractous subset was used as the target domain. Due to the absence of clear images paired to the cataractous ones, the results were assessed using the evaluation metrics of IS and NIQE without the reference of ground truth.

Fig. 6 exhibits the results of the restoration. And as plenty of unpaired normal and cataractous images were contained in the Kaggle dataset, CycleGAN was not only trained on the synthesized data, but also on the real unpaired data (denoted as CyclaGAN-real). The optic disk and a retinal hemorrhage are highlighted to demonstrate the capacity of retinal structure preservation. The structures are properly preserved by [12], [16], but the image is mapped to an inferior color and illumination space. In the results of [21], [26], the retinal hemorrhage is preserved, while the vessels in the optic disk

TABLE V

QUANTITATIVE COMPARISON WITH STATE-OF-THE-ART METHODS ON THE PUBLIC DATASET. SUPERIOR RESULTS ARE PRESENTED BY OUR METHOD IN THE METRICS WITHOUT REFERENCE.

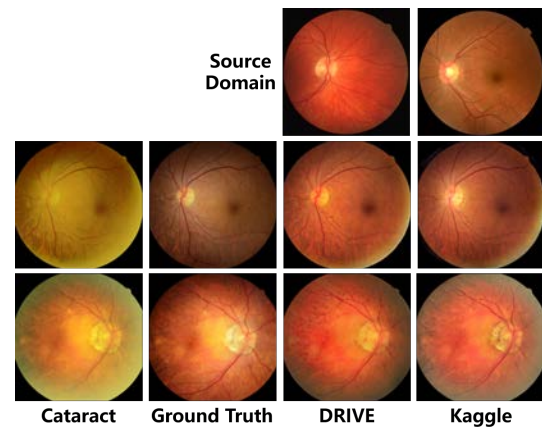| Algorithms	IS ↑	NIQE ↓
Mitra et al. [12]	1.31	6.44
SGRIF [16]	1.44	5.98
Cao et al. [26]	1.45	5.98
pix2pix [30]	1.49	6.30
CycleGAN [31]	1.40	5.25
CycleGAN-real [31]	1.47	5.30
Luo et al. [22]	1.45	5.56
CofeNet [21]	1.51	5.49
ArcNet (ours)	**1.52**	**4.52**

Fig. 7. Images were restored using different source domains. The first row exhibits the examples from DRIVE and Kaggle, which are used as the source domain to restore cataractous images. The following rows provide the restoration results using the corresponding source domain.

are fuzzy. Compared with the existing algorithms, ArcNet satisfactorily preserved the optic disk and retinal hemorrhage, as well as mapped the image to proper image space.

The metrics of IS and NIQE provides quantitative evaluations. The quality of restoration results is summarized in Table V. The quantitative evaluations in Table V further demonstrate that ArcNet achieves reasonable and obvious improvements of the restoration.

E. Selection of source domain

Since the restoration model was learned from the source domain, it can infer that the selection of the source domains impacts the restoration performance. To understand the impact of the selection of source domains, an investigation was conducted by comparing the images restored using different source domains. The dataset of DRIVE, and the normal subset from Kaggle were respectively employed to simulate cataract-like images and synthesize the source domain. And RCF was used as the target domain.

The images restored using different source domains are visualized in Fig. 7. The impact of the source domain was observed from the restored images. Because the images in DRIVE have a higher brightness in red than those in Kaggle, the higher brightness in red was observed from the images restored using DRIVE compared to Kaggle. Due to the training on the source domain and the alignment between the source domain's clear images and the target domain's outputs, it is no surprise that the learned model tries to improve the quality of cataractous images with the appearances of the source domain. Alternatively, due to the high image quality of DRIVE and Kaggle, the images restored using both source domains by the proposed ArcNet always achieve a remarkable quality.

The quantitative comparison of the restoration using the target domain of RCF with the source domain of DRIVE and Kaggle is presented in Table VI, and the performances of the existing CNN-based algorithms were summarized as well. The hand-crafted methods were omitted in this experiment, since their performances are independent of the source domain. Compared to the existing algorithms, the proposed method stably achieves the best image quality with both source domains.

This experiment confirms the impact on restoration from the selection of the source domain. The intensity distribution

TABLE VI

COMPARISON ON RESTORATION WITH THE SOURCE DOMAIN OF DRIVE AND KAGGLE.

Algorithms	DRIVE		Kaggle	
	SSIM	PSNR	SSIM	PSNR
pix2pix [30]	0.731	17.68	0.739	17.71
CycleGAN [31]	0.732	17.25	0.727	17.40
Luo et al. [22]	0.704	17.29	0.735	17.74
CofeNet [21]	0.754	18.03	0.759	18.54
Previous work [23]	0.755	18.07	0.748	18.24
ArcNet (ours)	**0.767**	**18.29**	**0.760**	**18.59**

of restored images follows the source domain. Fortunately, despite the discrepancy in color, the readability and quality of cataractous images could be improved once the source domain of high-quality is employed. And the outstanding performance of ArcNet is validated using different source domains.

F. Clinical Image Restoration and Diagnosis

Considering that the restoration of cataractous fundus images was conducted to promote the clinical observation and diagnosis, ArcNet was imported to a diagnosis task of fundus ocular diseases and compared with the state-of-the-art methods. The private ophthalmic dataset Fundus-iSee was used to conduct automatic diagnosis. Fundus-iSee consists of 720 (492 clear VS 228 cataractous) images of age-related macular degeneration (AMD), 270 (181 VS 89) of diabetic retinopathy (DR), 450 (312 VS 138) of glaucoma, 790 (478 VS 312) of high myopia, and 7770 (5868 VS 1902) normal fundus. Because cataracts distort the observation of the retinal fundus and increase the risk of misdiagnosis, restoration algorithms were applied to boost the diagnosis.

The diagnosis model was learned by ResNet-50 from 5000 randomly sampled training images. The rest 5000 images in Fundus-iSee were sorted into a clear and a cataractous subset for testing. Additionally, restoration was conducted on the cataractous subset by the state-of-the-art methods, where the restoration models were trained with the source and target domain composed of 2000 clear and 700 cataractous images
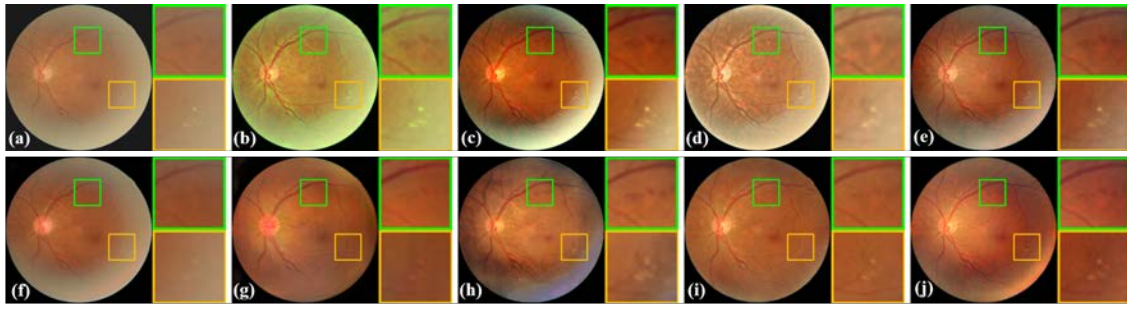
Fig. 8. Restoration example of the lesions in the fundus image with DR. (a) cataract image. (b) Mitra et al. [12]. (c) SGRIF [16]. (d) Cao et al. [26]. (e) pix2pix [30]. (f) CycleGAN [31]. (g) Luo et al. [22]. (h) CofeNet [21]. (i) Previous work [23]. (j) ArcNet (ours).
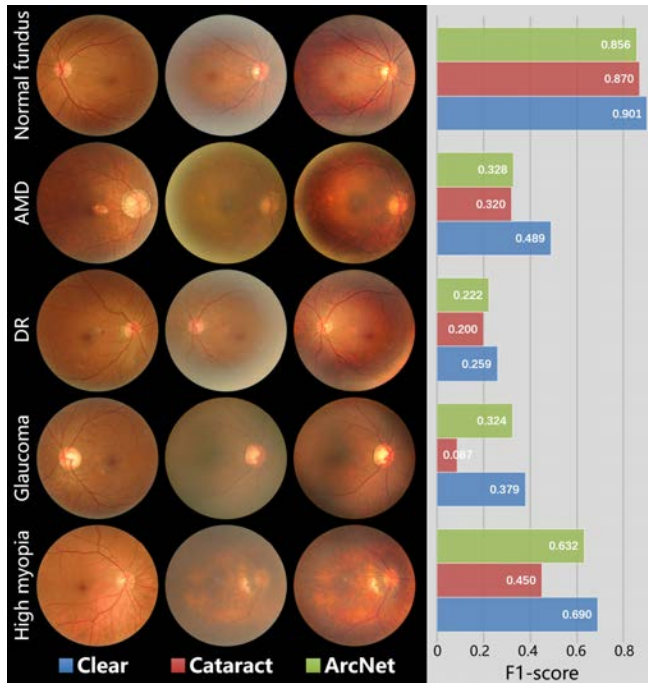

Fig. 9. Samples of Fundus-iSee and the diagnosis performance with ArcNet. The diagnosis F1-scores of each category are presented.

that fundus abnormalities are more clearly noticed in restored ones. Therefore, the diagnosis performance of ocular diseases has been remarkably boosted by ArcNet. Moreover, due to data imbalance, the diagnosis model tends to sort any uncertain images into the category of the normal fundus. And thus, diagnosis improvement was not observed in the F1-score of the restored normal fundus images, while the advance in Ckappa verified the improvement by the restoration.

TABLE VII
APPLICATION OF RESTORATION IN DIAGNOSIS.

Fundus Image	F1-score	Ckappa
Clear	0.838	0.448
Cataract	0.730	0.310
Mitra et al. [12]	0.722	0.364
SGRIF [16]	0.760	0.420
Cao et al. [26]	0.667	0.326
pix2pix [30]	0.713	0.381
CycleGAN [31]	0.724	0.286
Luo et al. [22]	0.712	0.370
CofeNet [21]	0.754	0.416
Previous work [23]	0.747	0.405
ArcNet (ours)	**0.761**	**0.428**

from Fundus-iSee. Fig. 8 shows an example of the restored images with lesions of DR. The diagnosis results on the clear, cataractous, and restored subset are exhibited in Fig. 9, while a comparison with the state-of-the-art methods is summarized in Table VII.

The restoration comparison for the lesions of DR is exhibited in Fig. 8. As a result of cataracts, the retinal hemorrhages (green box) and hard exudates (yellow box) are obscure in the original image. For the state-of-the-art methods, the lesions are diminished in the images restored by Cao et al. [26], CycleGAN [31], Luo et al. [22], and our previous work [23], while Mitra et al. [12], SGRIF [16], pix2pix [30], and CofeNet [21] preserve the lesions but result in limited image quality. ArcNet properly restores the image, and desirably enhances the lesions of retinal hemorrhages and hard exudates.

The image samples and F1-scores of each category are respectively presented in Fig. 9. Compared to the clear images, cataracts obstruct the observation of the retinal fundus. ArcNet improves the readability of the cataractous fundus images such

A diagnosis comparison with images restored by state-of-the-art methods is summarized in Table VII, which is evaluated by the metrics of F1-score and Cohen's kappa (Ckappa). It should be noted that Ckappa is more sensitive than F1-score in multi-categories classification, as it has a more realistic view of the model's performance when using imbalanced data [49]. According to the diagnostic results, ResNet-50 provides outstanding performance for diagnosing ocular fundus diseases on clear fundus images. However, cataracts impact the diagnosis that the risk of misclassification significantly increases in patients with cataracts. The effectiveness of the proposed algorithm is illustrated in the performance rebound of the images restored by ArcNet.

G. Model setting analysis

In this section, various model settings, including image size and HFC, are compared to analyze their impacts on the performance of ArcNet. And the corresponding computational cost and time are also presented.

Fig. 10 demonstrates the restoration performance with disparate image sizes and HFCs. The restoration by ArcNet was

implemented in three image sizes, i.e. 256×256, 512×512, and 768×768, while the HFC extracted in Eq. 6 was adjusted by changing the radius r_P and spatial constant σ_P of the low-pass Gaussian filter g_P. Under the identical model setting of HFC in Fig. 10, the image size of 768×768 provides the best performance, followed by 512×512 and 256×256. This illuminates that as a result of the increased resolution, the restoration performance grows with the image size. On the other hand, according to Eq. 6, the HFC is obtained through removing the LFC captured by g_P. Therefore, structure details will remain in the LFC and then be removed from the HFC if r_P is not large enough. As shown in Fig. 10 (a), the restoration performance improves with the increase of r_P, and the performance starts to converge once r_P reaches 26. Further, consistent with [47], the optimal performance is achieved in Fig. 10 (b) at σ_P of 9 cooperating with r_P of 26.
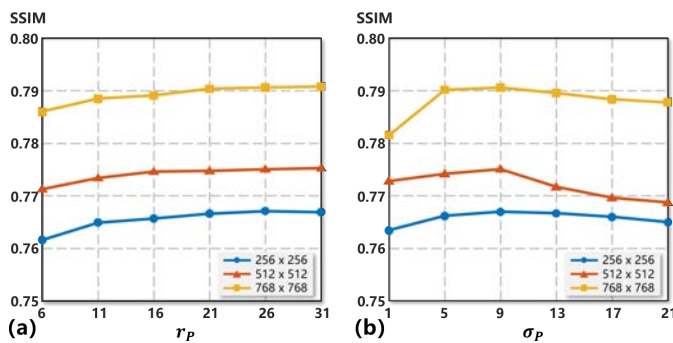

Fig. 10. Restoration performance with various image sizes and HFC. Three images sizes are provided and the performance of the HFC extracted with different r_p is presented in (a), diverse σ_P in (b).

The computational cost, training time, and inference time of ArcNet are summarized in Table VIII to demonstrate the overall complexity. The computational cost is quantified by multiply-accumulate operation (GMac), and training and inference time refer to the number of hours (h) for training ArcNet and the number of seconds (s) for restoring an image in the inference phase. As the image size increases, the computational cost increases exponentially, and the training and inference time grow correspondingly. Combining Fig. 10 and Table VIII, it indicates that a larger image size achieves better restoration performance, but the computational complexity accordingly increases. Thus considering the computational cost, the image size of 256×256 was employed to conduct the experiments.

TABLE VIII
COMPARISON OF COMPUTATIONAL COMPLEXITY WITH DIFFERENT IMAGE SIZES.

Image size	Computational cost (GMac)	Training time (h)	Inference time (s)
256×256	18.16	3.02	0.10
512×512	72.81	3.22	0.32
768×768	163.83	3.56	0.51

V. DISCUSSION

According to the experiments, the proposed ArcNet is an efficient annotation-free restoration algorithm for cataractous

fundus images. Compared to segmentation, HFC is more suitable to preserve retinal structures since it is free from segmentation annotation, and more importantly, the unannotated structure details are favorably preserved by HFC. The modules in ArcNet allow it to outperforms the state-of-the-art algorithms in the absence of annotations. Notably high-quality fundus images are still necessary to guarantee the performance of ArcNet. In the diagnosis of ocular fundus diseases, the clinical application value of ArcNet is presented that it promotes the accuracy of automatic diagnosis algorithms.

The annotation of clinical data is a bottleneck for the application of deep learning in many medical scenarios. Developing restoration algorithms for cataractous images requires the annotation from clear images of identical patients, which is extremely troublesome in clinics. Additionally, segmentation is employed in recent algorithms to preserve retinal structures in restoration. Under the prerequisite of segmentation annotations, these algorithms have practical difficulties in clinical deployment. The proposed ArcNet circumvents these rigorous requirements to restore cataractous fundus images without any annotation. Cataract-like images are simulated to train the restoration model, and structure guidance is automatically captured from the HFC of fundus images to preserve the fine retinal structures. The learned model is efficiently adapted to real cataract images by unsupervised domain adaptation. Thus ArcNet enjoys outstanding clinical practicability.

Although ArcNet circumvents the impractical requirement of annotations for fundus images, collecting high-quality fundus images as the source domain is still significant to guarantee the performance of ArcNet. In clinical applications, restoring the cataractous images using the source domain synthesized by the clear fundus images obtained under the conditions same as the cataractous ones will help ArcNet achieve the best restoration performance.

VI. CONCLUSION

Image restoration is a solution to the challenge of retinal examination caused by cataracts in fundus images. However, the annotation requirement limits restoration algorithms' development and application for cataractous fundus images. To mitigate this limitation, in this paper an annotation-free restoration network, named ArcNet, was proposed to restore cataractous fundus images while preserving the fine retinal structures. Extensive experiments were conducted to validate the performance and effectiveness of the proposed ArcNet. ArcNet restored cataractous images without annotations and outperformed the state-of-the-art restoration algorithms. Also, in cataracts patients, ArcNet boosted the performance of automatic diagnostic algorithms for ocular fundus diseases.

REFERENCES

[1] S. R. Flaxman, R. R. Bourne, S. Resnikoff, P. Ackland, T. Braithwaite, M. V. Cicinelli, A. Das, J. B. Jonas, J. Keeffe, J. H. Kempen *et al.*, "Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis," *The Lancet Global Health*, vol. 5, no. 12, pp. e1221–e1234, 2017.
[2] C. M. Lee and N. A. Afshari, "The global state of cataract blindness," *Current opinion in ophthalmology*, vol. 28, no. 1, pp. 98–103, 2017.

[3] W. H. Organization *et al.*, "Blindness: vision 2020—control of major blinding diseases and disorders," 2015.

[4] B. V. Kumar, R. P. Phillips, and S. Prasad, "Multifocal intraocular lenses in the setting of glaucoma," *Current opinion in ophthalmology*, vol. 18, no. 1, pp. 62–66, 2007.

[5] P. Flesner, B. Sander, V. Henning, H.-H. Parving, M. D. De La Cour, and H. Lund-Andersen, "Cataract surgery on diabetic patients. a prospective evaluation of risk factors and complications," *Acta ophthalmologica Scandinavica*, vol. 80, no. 1, pp. 19–24, 2002.

[6] S. J. Kim, R. Equi, and N. M. Bressler, "Analysis of macular edema after cataract surgery in patients with diabetes using optical coherence tomography," *Ophthalmology*, vol. 114, no. 5, pp. 881–889, 2007.

[7] J. L. Alio, A. B. Plaza-Puche, R. Férnandez-Buenaga, J. Pikkel, and M. Maldonado, "Multifocal intraocular lenses: an overview," *Survey of Ophthalmology*, vol. 62, no. 5, pp. 611–634, 2017.

[8] E. Peli and B. Schwartz, "Enhancement of fundus photographs taken through cataracts," *Ophthalmology*, vol. 94, pp. 10–13, 1987.

[9] T. Li, W. Bo, C. Hu, H. Kang, H. Liu, K. Wang, and H. Fu, "Applications of deep learning in fundus images: A review," *Medical Image Analysis*, p. 101971, 2021.

[10] E. Peli and T. Peli, "Restoration of retinal images obtained through cataracts," *IEEE transactions on medical imaging*, vol. 8, no. 4, pp. 401–406, 1989.

[11] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics gems*, pp. 474–485, 1994.

[12] A. Mitra, S. Roy, S. Roy, and S. K. Setua, "Enhancement and restoration of non-uniform illuminated fundus image of retina obtained through thin layer of cataract," *Computer methods and programs in biomedicine*, vol. 156, pp. 169–178, 2018.

[13] 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*. IEEE, 2013, pp. 1–3.

[14] 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, 2010.

[15] ——, "Guided image filtering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1397–1409, 2012.

[16] 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.

[17] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M.-H. Yang, "Low-light image enhancement via a deep hybrid network," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4364–4375, 2019.

[18] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[19] T. Huang, S. Li, X. Jia, H. Lu, and J. Liu, "Neighbor2neighbor: Self-supervised denoising from single noisy images," *arXiv preprint arXiv:2101.02824*, 2021.

[20] H. Zhao, B. Yang, L. Cao, and H. Li, "Data-driven enhancement of blurry retinal images via generative adversarial networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 75–83.

[21] Z. Shen, H. Fu, J. Shen, and L. Shao, "Modeling and enhancing low-quality retinal fundus images," *IEEE transactions on medical imaging*, vol. 40, no. 3, pp. 996–1006, 2020.

[22] Y. Luo, K. Chen, L. Liu, J. Liu, J. Mao, G. Ke, and M. Sun, "Dehaze of cataractous retinal images using an unpaired generative adversarial network," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3374–3383, 2020.

[23] H. Li, H. Liu, Y. Hu, R. Higashita, Y. Zhao, H. Qi, and J. Liu, "Restoration of cataract fundus images via unsupervised domain adaptation," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2021, pp. 516–520.

[24] R. GeethaRamani and L. Balasubramanian, "Retinal blood vessel segmentation employing image processing and data mining techniques for computerized retinal image analysis," *Biocybernetics and Biomedical Engineering*, vol. 36, no. 1, pp. 102–118, 2016.

[25] 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.

[26] L. Cao, H. Li, and Y. Zhang, "Retinal image enhancement using low-pass filtering and α-rooting," *Signal Processing*, vol. 170, p. 107445, 2020.

[27] 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.

[28] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6286–6295.

[29] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.

[30] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[31] J.-Y. Zhu *et al.*, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, 2017, pp. 2223–2232.

[32] A. D. Pérez, O. Perdomo, H. Rios, F. Rodríguez, and F. A. González, "A conditional generative adversarial network-based method for eye fundus image quality enhancement," in *International Workshop on Ophthalmic Medical Image Analysis*. Springer, 2020, pp. 185–194.

[33] D. Engin, A. Genç, and H. Kemal Ekenel, "Cycle-dehaze: Enhanced cyclegan for single image dehazing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 825–833.

[34] T. K. Yoo, J. Y. Choi, and H. K. Kim, "Cyclegan-based deep learning technique for artifact reduction in fundus photography," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 258, no. 8, pp. 1631–1637, 2020.

[35] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, "Domain generalization: A survey," *arXiv preprint arXiv:2103.02503*, 2021.

[36] S. Motiian, M. Piccirilli, D. A. Adjeroh, and G. Doretto, "Unified deep supervised domain adaptation and generalization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 5715–5725.

[37] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7167–7176.

[38] Y. Yang and S. Soatto, "Fda: Fourier domain adaptation for semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4085–4095.

[39] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. Efros, and T. Darrell, "Cycada: Cycle-consistent adversarial domain adaptation," in *International conference on machine learning*. PMLR, 2018, pp. 1989–1998.

[40] S. Zhao, B. Li, X. Yue, Y. Gu, P. Xu, R. Hu, H. Chai, and K. Keutzer, "Multi-source domain adaptation for semantic segmentation," *arXiv preprint arXiv:1910.12181*, 2019.

[41] Y. Luo, L. Zheng, T. Guan, J. Yu, and Y. Yang, "Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2507–2516.

[42] Q. Yu, A. Hashimoto, and Y. Ushiku, "Divergence optimization for noisy universal domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2515–2524.

[43] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7472–7481.

[44] F. C. Delori, E. S. Gragoudas, R. Francisco, and R. C. Pruett, "Monochromatic ophthalmoscopy and fundus photography: the normal fundus," *Archives of ophthalmology*, vol. 95, no. 5, pp. 861–868, 1977.

[45] J. Huang, D. Guan, A. Xiao, and S. Lu, "Fsdr: Frequency space domain randomization for domain generalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6891–6902.

[46] E. H. Land, "The retinex theory of color vision," *Scientific american*, vol. 237, no. 6, pp. 108–129, 1977.

[47] P. Getreuer, "A survey of gaussian convolution algorithms," *Image Processing On Line*, vol. 2013, pp. 286–310, 2013.

[48] J. Guo, Z. Pang, F. Yang, J. Shen, and J. Zhang, "Study on the method of fundus image generation based on improved gan," *Mathematical Problems in Engineering*, vol. 2020, 2020.

[49] L. A. Jeni, J. F. Cohn, and F. De La Torre, "Facing imbalanced data–recommendations for the use of performance metrics," in *Humaine association conference on affective computing and intelligent interaction*, 2013, pp. 245–251.