

Individualized Statistical Modeling of Lesions in Fundus Images for Anomaly Detection

Yuchen Du¹, Lisheng Wang¹, Deyu Meng¹, Benzhi Chen, Chengyang An, Hao Liu, Weiping Liu, Yupeng Xu, Ying Fan, Dagan Feng², *Life Fellow, IEEE*, Xiuying Wang¹, and Xun Xu

Abstract— Anomaly detection in fundus images remains challenging due to the fact that fundus images often contain diverse types of lesions with various properties in locations, sizes, shapes, and colors. Current methods achieve anomaly detection mainly through reconstructing or separating the fundus image background from a fundus image under the guidance of a set of normal fundus images. The reconstruction methods, however, ignore the constraint from lesions. The separation methods primarily model the diverse lesions with pixel-based independent and identical distributed (i.i.d.) properties, neglecting the individualized variations of different types of lesions and their structural properties. And hence, these methods may have difficulty to well distinguish lesions from fundus image backgrounds especially with the normal personalized variations (NPV). To address these challenges, we propose a patch-based non-i.i.d. mixture of Gaussian (MoG) to model diverse lesions for adapting to their statistical distribution variations in different fundus images and their patch-like structural properties. Further, we particularly introduce the weighted Schatten p-norm as the metric of low-rank decomposition for enhancing the accuracy of the learned fundus image backgrounds and reducing false-positives caused by NPV. With the individualized modeling of the diverse lesions and the background learning, fundus image backgrounds and NPV are finely learned and subsequently

Manuscript received 22 August 2022; revised 25 October 2022; accepted 20 November 2022. Date of publication 29 November 2022; date of current version 3 April 2023. This work was supported in part by the China NSFC Projects under Grant 12090024, in part by the National Natural Science Foundation of China under Grant 82171100, in part by the Shanghai Clinical Science and Technology Innovation Project under Grant SHDC12019111, in part by the Shanghai Jiao Tong University Translational Medicine Cross Research Funds under Grant YG2019ZDA26, and in part by the Macao Science and Technology Development Fund under Grant 061/2020/A2. (*Corresponding authors:* Lisheng Wang; Xun Xu)

Yuchen Du, Lisheng Wang, Benzhi Chen, Chengyang An, and Hao Liu are with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: yuchendu@sjtu.edu.cn; lswang@sjtu.edu.cn; chenbenzhi1987@126.com; acy_98@sjtu.edu.cn; sine9489@sjtu.edu.cn).

Deyu Meng is with the School of Mathematics and Statistics and Ministry of Education Key Lab of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China, and also with the Macao Institute of Systems Engineering, Macau University of Science and Technology, Taipa, Macao (e-mail: dymeng@mail.xjtu.edu.cn).

Weiping Liu is with Shanghai MicroPort MedBot (Group) Company Ltd., Shanghai 201210, China (e-mail: Weiping.Liu@microport.com).

Yupeng Xu, Ying Fan, and Xun Xu are with the Shanghai Key Laboratory of Ocular Fundus Diseases, Shanghai Engineering Center for Visual Science and Photo Medicine, Department of Ophthalmology, SJTU School of Medicine, Shanghai General Hospital, Shanghai 200080, China (e-mail: kevinxy@live.com; mdafany@sjtu.edu.cn; drxuxun@sjtu.edu.cn).

Dagan Feng and Xiuying Wang are with the School of Computer Science, The University of Sydney, Sydney, NSW 2006, Australia (e-mail: dagan.feng@sydney.edu.au; xiu.wang@sydney.edu.au).

Digital Object Identifier 10.1109/TMI.2022.3225422

distinguished from diverse lesions, to ultimately improve the anomaly detection. The proposed method is evaluated on two real-world databases and one artificial database, outperforming the state-of-the-art methods.

Index Terms— Anomaly detection, non-independent and identical distribution, mixture of Gaussian, weighted Schatten p-Norm, normal personalized variations.

I. INTRODUCTION

RETINAL pathologies, such as diabetic retinopathy, macular degeneration, drusen, and pathologic myopia, are the major causes of visual impairment and blindness worldwide [1]. It is estimated that 55.7 million people will be affected by retinal pathologies by 2050 [2]. Since most retinal pathologies are irreversible, population screening becomes an important approach for reducing the visual impairment of the at-risk people at an early stage. The workload magnitude of screening, diagnosis, and treatment however is a significant challenge for ophthalmologists. Thus, an effective and efficient automatic anomaly screening method by utilizing the power of artificial intelligence is at an urgent need for screening the large population.

In the past decade, extensive efforts have been devoted to develop computer-aided detection (CAD) of retinal pathologies [3], [4]. These AI-based methods, either sample driven or feature driven, albeit being effective for detecting predefined types of lesions from fundus images, unfortunately, often lack the generalizability for detecting diverse types of lesions. While transfer learning and domain adaptation methods gain great popularity for detection or segmentation of the specific types of lesions [5], [6], it yet to further extend for detecting or segmenting previously undefined lesion types. The wide varieties of lesion types with various properties in shapes, locations, sizes, colors, and textures pose significant challenges on the general anomaly detection methods. The rare types and the previously undefined types of lesions make the automated detection even harder. Thus, though the general anomaly detection method is in a high demand, in the meanwhile remains a challenge [7], [8].

To address the challenge of detecting different types of lesions, current research methods are often based on two common hypotheses including (1) a fundus image with diverse lesions is regarded as a composite of the lesions being superimposed onto a fundus image background [7]; and (2) the fundus images from the normal populations are similar and also share similarities with the fundus image backgrounds from patient populations [9]. Accordingly, anomaly detection can be modeled as a computational task to compute the fundus

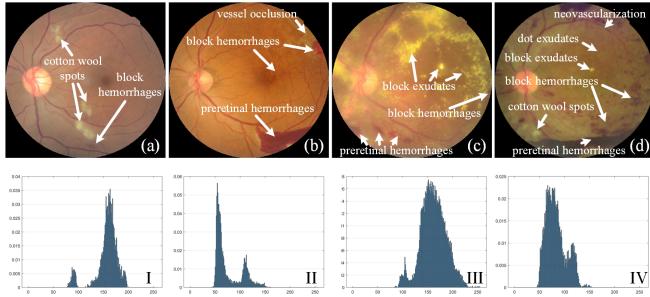


Fig. 1. The figures (I)-(IV) represent the intensity distributions of lesion regions in green channel of the corresponding top images (a)-(d). The complicated and different histogram distributions of lesions of an individual input image indicate that it would be more appropriate to model the diverse lesions with non-i.i.d. MoG.

image backgrounds for isolating diverse lesions from fundus images [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21]. Since the fundus image background is the learning objective, the anomaly detection methods are less lesion specific and dependent, therefore are able to achieve better generalizability for different lesion types.

The prevalent background learning based anomaly detection methods in fundus images can be divided into two groups: the generation/reconstruction methods [7], [8], [10], [11] and the separation methods [9]. The generation/reconstruction methods learned the basis vectors from normal fundus images to reconstruct the abnormal images for anomaly separation or isolation, which however faced a serious dilemma: both normal regions and lesions can be well reconstructed in more accurate fundus image backgrounds, which may lead to the false-negatives or false-positives. Alternatively, the separation methods were proposed to separate the various lesions (sparse structures) and fundus image backgrounds (low-rank structures) from fundus images by low-rank decomposition. However, different constraints on the anomaly detection modeling often lead to variations on the low-rank decomposition results. Thereby, the current challenge on more accurate anomaly detection is to determine the constraints, regularities, and metrics to be applied to low-rank decomposition of fundus images for optimal separation of the fundus image backgrounds from fundus images.

The majority of current studies either utilizes the low-rank property to reconstruct fundus image backgrounds of the test fundus images, or constrains lesions by i.i.d. and pixel-wise properties. However, the affluent characteristics of the lesions which could facilitate the performance of anomaly detection have seldom been explored yet that in different fundus images the lesions often manifest varieties in shapes, locations, sizes, and colors and exhibit individualized subject-wise statistical variations characteristics, as shown in Fig.1. And lesions usually have a structural property, where they generally exhibit patch-like appearances. Additionally, the fundus image background also encompasses inter-subject heterogeneity, namely, the normal personalized variations (NPV) that specifically refer to the intensity variations in the background of normal retinal fundus images from one subject to another. The NPV embraced in the normal regions or normal fundus images are difficult to discern from the diverse lesions, which is the often cause leading to false-positives.

To achieve more accurate anomaly detection in fundus images, it is important to model individualized statistical variation characteristics and the patch-like structural properties of diverse lesions in different fundus images with high accuracy. Such precise modeling will be used as the prior to adaptively constrain sparse structures decomposition from the fundus images, thereby will improve the discernment for diverse lesions. Accordingly, to achieve these objectives, a patch-based non-independent and identical distributed (non-i.i.d.) Mixture of Gaussian (PNMoG) is proposed to adaptively model the individualised variations and the patch-like characteristics of the lesions in different fundus images. More specifically, PNMoG models the lesions in different fundus images with different MoG distributions (i.e., parameters of MoG are different) in terms of their individualized statistical distributions; further, the MoG is encoded in a multivariate manner such that different lesion patches are encoded as different units into MoG to maintain lesion's patch-like structure.

Meanwhile, to better encompass the NPV, the metric with the weighted Schatten p -norm (WSPN) [22] is defined to constrain fundus image backgrounds. The WSPN is constructed through regularizing the low-rank property of fundus image backgrounds by a weighted and exponential trace-norm metric. Since WSPN can provide a relaxation constraint, normal regions with NPV are more likely to be decomposed into low-rank structures rather than sparse structures.

Consequently, to improve diverse lesions detection in fundus images, a new model is proposed by integrating PNMoG and WSPN into the low-rank decomposition of the fundus images, and used for individually and adaptively constraining diverse lesions and simultaneously encoding fundus image backgrounds, and diverse lesions detection in fundus images is well improved. The effectiveness and advantages of the proposed method are validated by experiments and comparisons with the state-of-the-art methods.

A. Related Works

The anomaly detection methods can be generally classified into background learning methods [7], [8], [9], [10], [11], [12], [13], [14] and deep learning methods [15], [16], [17], [18], [19], [20], [21]. Background learning methods, the state-of-the-art in diverse anomaly detection from fundus images, can be further classified into background reconstruction and background separation methods.

The background reconstruction methods firstly learn the basis vectors of the low-dimensional subspace of the normal fundus images, and then reconstruct a test fundus image by the basis vectors [7], [8], [10], [11]. The reconstructed image is regarded as the approximation of fundus image background of the test image. The pixels or local patches with large reconstruction errors will be identified as anomalies. There are different strategies for learning the basis and reconstructing the background. In [7], a set of orthogonal basis was learned by PCA decomposition from a fundus matrix formed by a set of stacked spatially-aligned and vectorized normal fundus images. In [8], the atoms in the dictionary that was trained from the fundus matrix by k-svd and OMP algorithms were used to sparsely reconstruct the input images. In [11],

a dictionary was trained with 3D patches of normal brain MRI images and then used to reconstruct the local patches for a test image. In these methods, since both the normal background and lesion regions of the test image were reconstructed indiscriminately, and consequently, these methods tended to reconstruct either high-precision fundus image backgrounds with lesions or low-precision fundus image backgrounds without lesions where NPV were also excluded, which hence may lead to either false-positives or false-negatives.

As an alternative, in the background separation methods, the fundus image background (or diverse lesions) can be separated from a test fundus image mainly utilizing the low-rank (or sparse) characteristics between the fundus image background (or diverse lesions) and the normal fundus images [9]. However, these methods ignored the important lesion properties [7], [8], [11]. Methods were proposed to model lesion's statistical properties by pixel-based i.i.d. Mixture of Gaussian (MoG) [9], [12], [14] or Mixture of Exponential Power (MoEP) [13], in the meanwhile to constrain the lesions from the UV^T low-rank decomposition to better fit to MoG or MoEP. However, lesions in different fundus images manifest individualized statistical variations, as shown in Fig.1, and often present as patch-like regions, and hence neither a pixel-based i.i.d MoG/MoEP nor pixel-based modeling is sufficient to depict these characteristics. Further, the simple UV^T constraint on fundus image backgrounds also would not be able to encompass NPV in fundus image backgrounds, thus resulting in inaccurate background reconstruction and false-positives.

More recently, deep learning methods have been adapted for anomaly detection from various images. For instance, the AnoGAN, f-AnoGAN and DCGAN [16], [17], [18] were proposed for anomaly detection. However, these models might face with challenges to reconstruct the image details, which hence may generate false-positives. The ProxyAno model [19] was proposed to detect anomalies in fundus images and OCT images through measuring the feature correspondence of test images with their superpixel-images (SI) counterparts. A DGM model [20] was designed to detect and classify lesions by synthesizing normal images and disentangling anomalies. The DRAEM model [21] was proposed by combining a discriminative module with reconstruction module to discern anomalies. However, these methods may also confront a dilemma to simultaneously satisfy both the precise background reconstruction and fine lesions suppression in the whole or partial of their algorithms, and thereby may generate either false-negatives or false-positives.

B. Contributions

The main contributions of this paper can be summed up as follows:

- 1) A patch-based non-i.i.d. MoG is proposed to more accurately model the individualized variations of diverse lesions in different fundus images as well as their patch-like structures.
- 2) The weighted Schatten p -norm is utilized as the metric of the low-rank decomposition to enhance the accuracy of decomposed fundus backgrounds, cooperating with

PNMoG, false-positives caused by NPV are well alleviated.

- 3) The individualized lesions modeling and new metric for low-rank decomposition problem is assembled as an optimization problem with a concise form for anomaly detection, and an efficient algorithm is provided for solving this problem.

II. METHODOLOGY

The method is proposed for anomaly detection from fundus images. Firstly, all fundus images will be preprocessed for reducing bias from different imaging conditions and for excluding personalized differences of blood vessels in different images. Secondly, anomaly detection in fundus images is modeled as a joint model consisted of two components including (1) the PNMoG to depict the lesion characteristic diversities and varieties that are specific to the individualized different types of lesions and subject-wise images, and (2) the WSPN to learn the fundus image backgrounds to better encompass the NPV, as illustrated in Fig.2.

A. Preprocessing

Similar as [7] and [8], all fundus images are preprocessed for normalizing them in size and color [23] and spatially-aligning them [24] for low-rank decomposition computation. Meanwhile, blood vessels are detected by U-net [25] and removed from fundus images and their regions are filled-up by surrounding colors [26]. Only the green channels of all color fundus images are used in the computation [7], [8].

B. The PNMoG-WSPN Model for Anomaly Detection

After preprocessing, the normal fundus images and the fundus image backgrounds of abnormal fundus images are much similar with each other, which can be formulated by low-rank representation. Accordingly, the anomaly detection problem can be conducted by low-rank decomposition of the fundus images under certain constraints so as to separate diverse lesions from the fundus images. To achieve the low-rank decomposition, the arrangement of the fundus images is discussed below.

The input abnormal fundus images are preprocessed and stacked as a tensor $\mathcal{X}_1 \in \mathbb{R}^{h \times w \times m_1}$, where h , w , and m_1 denote the height, width, and number of abnormal fundus images. Then \mathcal{X}_1 is stacked with a prepared and preprocessed normal fundus image set $\mathcal{X}_2 \in \mathbb{R}^{h \times w \times m_2}$ which helps learn the fundus image backgrounds, where m_2 is the number of normal fundus images. Finally, a new tensor $\mathcal{X} \in \mathbb{R}^{h \times w \times m}$ is ready to be decomposed, where $m = m_1 + m_2$. In our model, both abnormal and normal fundus images are used for low-rank decomposition. However, since the normal images generally would have much more effects on the decomposition results of a certain abnormal fundus image, thus our method can be regarded as a weakly supervised method [9].

To accurately detect diverse lesions, we need to decompose the given tensor \mathcal{X} into a low-rank tensor \mathcal{B} and a sparse tensor \mathcal{F} that \mathcal{B} is formed by fundus image backgrounds of abnormal images and all normal images, and \mathcal{F} is formed by

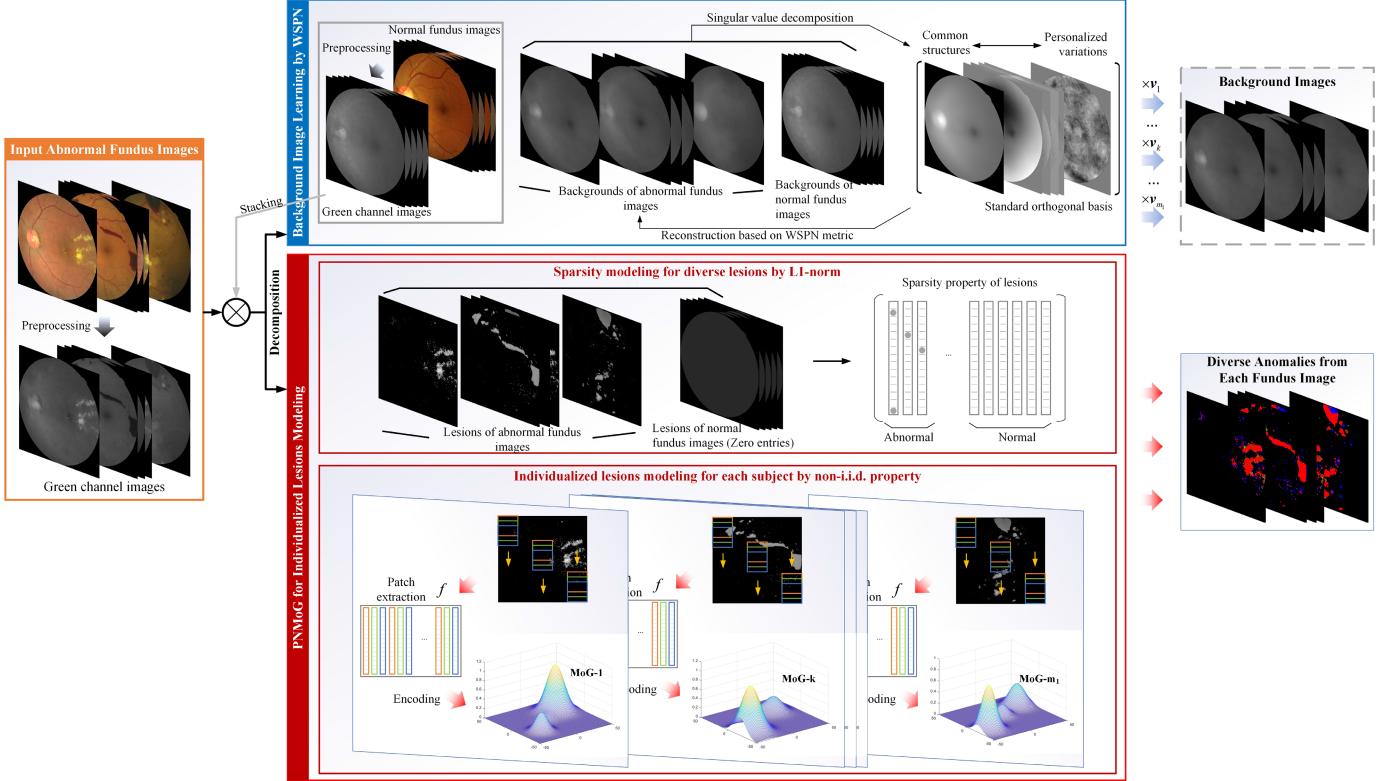


Fig. 2. Illustration of the proposed method. The anomaly fundus images are decomposed into the background images and lesion images. The former is modeled by the background image learning module by WSPN metric, and the latter is constrained by PNMoG module for individualized lesions modeling with patch-based non-i.i.d. statistical regularization for each subject and sparsity lesions constraint by l_1 -norm.

diverse lesions from abnormal images and zero entries from all normal images, as shown below:

$$\mathcal{X} = \mathcal{B} + \mathcal{F} \quad (1)$$

To more accurately generate tensors \mathcal{B} and \mathcal{F} from the tensor \mathcal{X} , the fundus image backgrounds and diverse lesions need to be modeled appropriately, so as to suitably constrain the low-rank decomposition of \mathcal{X} .

1) PNMoG for Individualized Lesions Modeling: In fundus images, the lesions exhibit two characteristics. First, the complexity of the lesions comes from two aspects. On one hand, the distribution of lesions in one fundus image is complex, thus we prefer MoG to approximate and adapt this complexity. On the other hand, the statistical distributions of lesions in different fundus images are also individualized and variant. Thus, instead of a single MoG (i.i.d. MoG), we propose the non-i.i.d. MoG to individually model the variant statistical distributions of lesions in different fundus images, i.e. lesions in each fundus image are modeled by different MoGs.

Second, the lesions manifest patch-like structures rather than pixel-wise patterns. That indicates the patches are more proper to act as units to depict the lesions. Here, we use a joint probability distribution to conduct the statistical calculation:

$$\mathcal{F}_l^{(c)} \sim \mathcal{N}(\mathcal{F}_l^{(c,1)}, \mathcal{F}_l^{(c,2)}, \dots, \mathcal{F}_l^{(c,pq)}) \quad (2)$$

where $\mathcal{N}(\cdot)$ is the joint probability distribution of Gaussian, \mathcal{F}_l denotes the l^{th} image in \mathcal{F} , $\mathcal{F}_l^{(c)}$ is the c^{th} patch in \mathcal{F}_l ,

$\mathcal{F}_l^{(c,i)}$ is the i^{th} pixel in $\mathcal{F}_l^{(c)}$, p and q are the height and the width of a patch and pq is the number of pixels in a patch.

Combined with non-i.i.d. MoG, we propose the patch-based non-i.i.d. MoG to model both the individualized distribution of lesions and their patch-like structures as follows:

$$\mathcal{F}_l \sim \sum_{n_p} \log \sum_k \pi_{kl} \mathcal{N}(f(\mathcal{F})_{n_p l} | \mu_{kl}, \Sigma_{kl}) \quad (3)$$

where $\mathcal{N}(\cdot)$ is with mean value vector $\mu_{kl} \in \mathbb{R}^{pq \times 1}$ and covariance matrix $\Sigma_{kl} \in \mathbb{R}^{pq \times pq}$, $\pi_{kl} \geq 0$ is the mixing coefficient with $\sum_k \pi_{kl} = 1$, k is the number of Gaussian components, f is a map to transform patches in \mathcal{F} into vectors:

$$f : \mathbb{R}^{h \times w \times m_1} \rightarrow \mathbb{R}^{pq \times N_p \times m_1} \quad (4)$$

where N_p is the number of patches extracted from one fundus image. $f(\mathcal{F})_{n_p l} \in \mathbb{R}^{pq \times 1}$ is the n_p^{th} vector from the l^{th} layer in $f(\mathcal{F})$, corresponding to the patch $\mathcal{F}_l^{(n_p)}$, $l = 1, \dots, m_1$.

Assembling all of the images in \mathcal{F} , the PNMoG model is proposed as follows:

$$\Omega_1(\mathcal{F}) = - \sum_l \sum_{n_p} \log \sum_k \pi_{kl} \mathcal{N}(f(\mathcal{F})_{n_p l} | \mu_{kl}, \Sigma_{kl}) \quad (5)$$

where $\Omega_i(\cdot)$ denotes the i^{th} constraint.

Additionally, in tensor \mathcal{X} , diverse lesions also exhibit sparse characteristics. Thus, their corresponding tensor \mathcal{F} has a small l_1 -norm value. Accordingly, we have:

$$\Omega_2(\mathcal{F}) = \|\mathcal{F}\|_1 \quad (6)$$

2) WSPN for Background Learning: The existence of NPV is the major obstacle of accurate background learning. Accurate modeling for fundus image backgrounds means NPV regions should also be encompassed into fundus image backgrounds. The WSPN model, which provides a relaxation form for low-rank constraint by optimizing the weights and exponential components, is used as a similarity measurement of \mathcal{B} to encompass the NPV, as shown below [27]:

$$\Omega_3(\mathcal{B}) = \|\mathcal{B}_{(3)}\|_{\omega, S_p}^p \quad (7)$$

where $\mathcal{B}_{(3)}$ denotes the unfold operation along the 3rd-mode of tensor \mathcal{B} , specifically, this operation will transform tensor $\mathcal{B} \in \mathbb{R}^{h \times w \times m}$ into a matrix $\mathbf{B} \in \mathbb{R}^{hw \times m}$, $\|\mathbf{A}\|_{\omega, S_p} = (\sum_i \omega_i \delta_i(\mathbf{A})^p)^{\frac{1}{p}}$, $\delta_i(\mathbf{A})$ is the i^{th} singular value of matrix \mathbf{A} , ω_i is the weight of the $\delta_i(\mathbf{A})$, p controls the low-rank degree of \mathbf{A} . When ω equals 1 and p equals 1, WSPN will degenerate to the sum of $\delta_i(\mathbf{A})$, which is the well-known nuclear norm.

3) The PNMoG-WSPN Model: By integrating Eq.(5), Eq.(6) and Eq.(7), we can formulate diverse lesions detection problem as the final PNMoG-WSPN model with parameters $\Theta = \{\mathcal{B}, \mathcal{F}, \mu, \Sigma, k\}$:

$$\begin{aligned} & \min \|\mathcal{B}_{(3)}\|_{\omega, S_p}^p + \beta \|\mathcal{F}\|_1 \\ & -\lambda \sum_l \sum_{n_p} \log \sum_k \pi_{kl} \mathcal{N}(f(\mathcal{F})_{n_p l} | \mu_{kl}, \Sigma_{kl}) \\ & s.t. \mathcal{X} = \mathcal{B} + \mathcal{F} \end{aligned} \quad (8)$$

where β and λ are the trade-off parameters of the sparsity module and the PNMoG module, respectively. The classic Robust PCA decomposition is only a special case of this formula.

III. OPTIMIZATION APPROACH

The Expectation-Maximum (EM) algorithm [28] is readily utilized to optimize the proposed model. The algorithm iterates between updating all of the Gaussian components (E step) and optimizing the parameters in Θ (M step).

A. E Step

A latent variable $z_{n_p l, k}$, where $z_{n_p l, k} \in \{0, 1\}$ and $\sum_k z_{n_p l, k} = 1$, is introduced to represent the assignment of $f(\mathcal{F})_{n_p l}$ to a specific component of PNMoG. The posterior probability of component k takes the form:

$$\kappa_{n_p l, k} = E(z_{n_p l, k}) = \frac{\pi_{kl} \mathcal{N}(f(\mathcal{F})_{n_p l} | \mu_{kl}, \Sigma_{kl})}{\sum_k \pi_{kl} \mathcal{N}(f(\mathcal{F})_{n_p l} | \mu_{kl}, \Sigma_{kl})} \quad (9)$$

B. M Step

This step needs to minimize the following object function with respect to parameters in Θ :

$$\begin{aligned} & \min \|\mathcal{B}_{(3)}\|_{\omega, S_p}^p + \beta \|\mathcal{F}\|_1 \\ & -\lambda \sum_l \sum_{n_p} \sum_k \kappa_{n_p l, k} \log(\pi_{kl} \mathcal{N}(f(\mathcal{F})_{n_p l} | \mu_{kl}, \Sigma_{kl})) \\ & s.t. \mathcal{X} = \mathcal{B} + \mathcal{F} \end{aligned} \quad (10)$$

This minimization problem can be solved by the alternative optimization strategy to iteratively update the parameters $\mathcal{B}, \mathcal{F}, \mu, \Sigma$.

1) Update μ, Σ : The closed-form updating equation [28] for both of the two parameters can be easily deduced as:

$$\begin{aligned} N_{kl} &= \sum_{n_p} \kappa_{n_p l, k}, \pi_{kl} = \frac{N_{kl}}{\sum_k N_{kl}}, \mu_{kl} = \frac{\sum_{n_p} \kappa_{n_p l, k} f(\mathcal{F})_{n_p l}}{\sum_{n_p} \kappa_{n_p l, k}}, \\ \Sigma_{kl} &= \frac{1}{N_{kl}} \sum_{n_p} \kappa_{n_p l, k} (f(\mathcal{F})_{n_p l} - \mu_{kl}) (f(\mathcal{F})_{n_p l} - \mu_{kl})^T \end{aligned} \quad (11)$$

We apply alternating direction method of multipliers (ADMM) [29] algorithm to solve the subproblem. The Lagrangian form of Eq.(10) is given as follows:

$$\begin{aligned} & \min -\lambda \sum_l \sum_{n_p} \sum_k \kappa_{n_p l, k} \log(\pi_{kl} \mathcal{N}(f(\mathcal{F})_{n_p l} | \mu_{kl}, \Sigma_{kl})) \\ & + \|\mathcal{B}_{(3)}\|_{\omega, S_p}^p + \beta \|\mathcal{F}\|_1 + \frac{\nu}{2} \left\| \mathcal{X} - \mathcal{B} - \mathcal{F} + \frac{1}{\nu} \mathcal{Y} \right\|_F^2 \end{aligned} \quad (12)$$

where \mathcal{Y} is the Lagrangian operator, ν is the updating step for each iteration.

2) PNMoG Simplification: Due to the format difference of $f(\mathcal{F})$ and \mathcal{F} , directly optimizing Eq.(12) is difficult. In order to provide a solution and accelerate the optimization, here we build a connection between $f(\mathcal{F})$ and \mathcal{F} . Assuming each pixel within a patch is i.i.d., we take the l^{th} image in \mathcal{F} as an example to demonstrate the simplification procedure.

The main diagonal of Σ_{kl} is denoted by $\text{diag}(\Sigma_{kl})$. The l^{th} image of PNMoG module of Eq.(12) can be re-formulated as:

$$\sum_k \sum_{n_p} \kappa_{n_p l, k} \left\langle \left(f(\mathcal{F})_{n_p l} - \mu_{kl} \right)^2, \text{diag}(\Sigma_{kl}^{-1}) \right\rangle \quad (13)$$

In Eq.(13), $f(\mathcal{F})$ can be represented by \mathcal{F} with a weight matrix which is only subject to the patch size p and q . Define:

$$\mathcal{W}_k^{(1)}(h, w, l) = \sum_p \sum_q \frac{\kappa_{(h-p+1, w-q+1, l), k}}{\vartheta_{pq, l, k}^2} \quad (14)$$

and

$$\mathcal{W}_k^{(2)}(h, w, l) = \sum_p \sum_q \frac{\kappa_{(h-p+1, w-q+1, l), k} \mu_{pq, l, k}}{\vartheta_{pq, l, k}^2} \quad (15)$$

where $\vartheta_{i, l, k}^2$ is the i^{th} entry in $\text{diag}(\Sigma_{kl})$. The simplified optimization formula can be obtained as follows:

$$\begin{aligned} & \min \frac{\lambda}{2} \sum_k \left\| \sqrt{\mathcal{W}_k^{(1)}} * \left(\mathcal{F} - \frac{\mathcal{W}_k^{(2)}}{\mathcal{W}_k^{(1)}} \right) \right\|_F^2 \\ & + \|\mathcal{B}_{(3)}\|_{\omega, S_p}^p + \beta \|\mathcal{F}\|_1 + \frac{\nu}{2} \left\| \mathcal{X} - \mathcal{B} - \mathcal{F} + \frac{1}{\nu} \mathcal{Y} \right\|_F^2 \end{aligned} \quad (16)$$

where $(*)$ denotes the Hadamard product.

Algorithm 1 PNMoG-WSPN

Input:

The set of abnormal fundus images: $\mathcal{X}_1 \in \mathbb{R}^{h \times w \times m_1}$;
The number of components of Mixture of Gaussian: k ;
The background low-rank constraint parameters p and ω ;
The trade-off parameters β and λ ;

Output:

The reconstructed background image tensor \mathcal{B} and the lesion tensor of abnormal fundus image \mathcal{F} ;

1: Initialization:

2: $\mathcal{B}, \mathcal{F}, \mu, \Sigma, \nu, \mathcal{Y}$.

3: **while** not convergence **do**

4: (E step) Evaluate $\kappa_{n_p l, k}$ by Eq.(9);

5: (M step) Evaluate parameters μ, Σ by Eq.(11);

6: Refresh the PNMoG coefficients $\mathcal{W}_1, \mathcal{W}_2$ by Eq.(14) and Eq.(15).

7: Update \mathcal{B} by Eq.(17);

8: Update \mathcal{F} by Eq.(18);

9: Update Lagrangian multiplier \mathcal{Y} by Eq.(19) and ν by $\nu = \rho\nu$.

10: **return** \mathcal{B}, \mathcal{F}

3) Update \mathcal{B} : This step needs to solve the minimization subproblem $\mathcal{B} = \arg \min_{\mathcal{B}} \|\mathcal{B}_{(3)}\|_{\omega, S_p}^p + \frac{\nu}{2} \|\mathcal{B} - \mathcal{Z}\|_F^2$, where $\mathcal{Z} = \mathcal{X} - \mathcal{F} + \frac{1}{\nu} \mathcal{Y}$. As suggested in [30], the GST algorithm is readily used to solve this problem:

$$\delta_i^{(t+1)} = |\sigma_i| - \frac{1}{\nu} \omega_i p \left(\delta_i^{(t)} \right)^{p-1} \quad (17)$$

where $\mathcal{Z}_{(3)} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}$ with $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_r)$, $\delta_i^{(t)}$ is the i^{th} singular value of $\mathcal{B}_{(3)}$ in the t^{th} iteration.

4) Update \mathcal{F} : The solution of \mathcal{F} can be solved by the soft-threshold method:

$$\mathcal{F} = \text{sign}(\mathcal{F}) \cdot \max(|a| - \zeta, 0) \quad (18)$$

where $a = \frac{\lambda \sum_k \mathcal{W}_k^{(2)} + \nu(\mathcal{X} - \mathcal{B}) + \mathcal{Y}}{\lambda \sum_k \mathcal{W}_k^{(1)} + \nu}$, $\zeta = \frac{\beta}{\lambda \sum_k \mathcal{W}_k^{(1)} + \nu}$.

5) Update Lagrangian Operator: The Lagrangian operator \mathcal{Y} is updated by:

$$\mathcal{Y} = \mathcal{Y} + \nu (\mathcal{X} - \mathcal{B} - \mathcal{F}) \quad (19)$$

for each iteration, $\nu = \rho\nu$.

Our optimization process is shown in Algorithm 1.

IV. EXPERIMENTS

A. Experimental Datasets

Three independent datasets are used to evaluate the performance of the proposed method, including 1) *Kaggle dataset* and 2) *Messidor dataset* for evaluating the performance of diverse lesions detection, and 3) an *Artificial dataset* for evaluating the performance of background reconstruction.

1) Kaggle Dataset: The *Kaggle dataset* is obtained from [8] and [9] to evaluate the detection performance of diverse lesions from complex fundus images. *Kaggle dataset* consists of 193 fundus images, where 26 different types of lesions with more than 17000 lesion regions exist. The size of lesions ranges from 2 pixels to more than 1800 pixels. Lesions of high to low contrast are contained in this dataset.

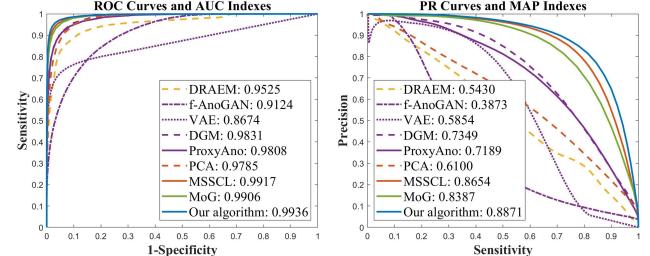


Fig. 3. The ROC and PR curves of *Kaggle dataset*.

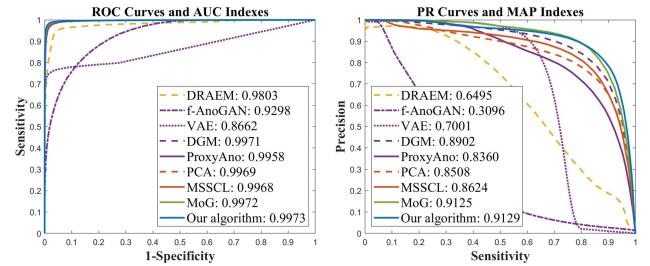


Fig. 4. The ROC and PR curves of *Messidor dataset*.

2) Messidor Dataset: The *Messidor dataset* is obtained from [8] and [9] to evaluate the detection performance of lesions from simple fundus images. *Messidor dataset* consists of 110 fundus images, where 6 types of lesions (mainly retinopathy and macular edema [31]) with more than 7500 lesion regions contained in this dataset.

Low quality images are rejected in the above datasets. Diverse lesions in the above two datasets are manually annotated at the pixel level by an experienced ophthalmologist.

3) Artificial Dataset: We establish an *Artificial dataset* that a pseudo-abnormal fundus image set is generated by the following steps: (1) the lesion regions in green channels of fundus images from *Kaggle dataset* are extracted, and (2) the extracted lesion regions of 193 abnormal images are replaced the corresponding positions of collected 193 normal fundus images. The normal fundus images act as the groundtruths of fundus image backgrounds of this dataset.

B. Methods in Comparison

Eight related methods of three categories, including (1) five deep learning methods, (2) two background reconstruction methods, and (3) one background separation method are compared with the proposed method. The five deep learning methods include the ProxyAno [19], DGM [20], VAE [32], f-AnoGAN [17], and DRAEM [21]. The two background reconstruction methods include Sparse coding-based method (MSSCL) [8] and PCA-based method (PCA) [7]. The background separation method is the MoG method (MoG) [9].

C. Quantitative Indicators

The Area Under the Receiver Operating Characteristic (ROC) curve (AUC) and Area Under the Precision-Recall (PR) curve (MAP) are used as the indicators. Please note that the MAP indicator is more appropriate to reflect the real detection performance for diverse anomalies since the anomalies only

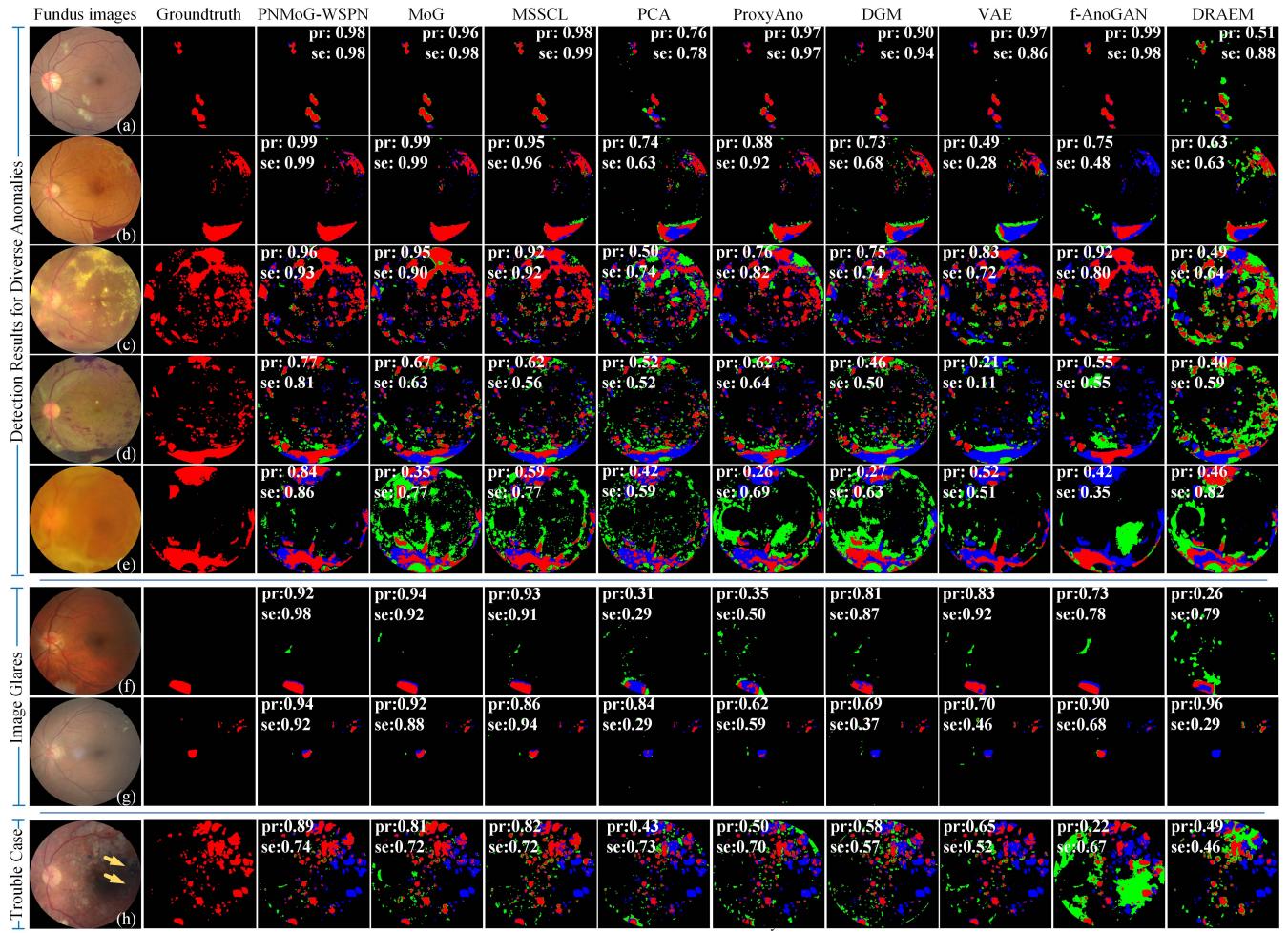


Fig. 5. The binary results of lesions detection of the proposed method with comparison methods. The red regions represent the lesions, the green and blue ones represent the false-positives and false-negatives, respectively. “pr” denotes the value of precision, “se” denotes the value of sensitivity.

occupy a very small proportion of the whole fundus images, which indicates the unbalance of positive-negative samples.

For most anomalies in fundus images, the boundaries are always fuzzy. To alleviate the partial volume effects by boundaries, the lesion regions of detected results and groundtruths are dilated by two pixels to conduct the performance [33].

D. Default Parameters Setting

In the proposed method, the coefficient vector ω of Eq.(7) is defined following the *non-descend* order given by negative interrelation to the singular values. We define $\omega_i = (\sigma_1(\mathcal{B}_{(3)}) / \sigma_i(\mathcal{B}_{(3)}))^\tau$, where the exponent τ is set to 0.1 in our experiments. The component k of Mixture of Gaussian is set to 3, and the patch size is set to 2×2 .

For each comparison method, the reported parameters are used as the default settings for the corresponding models. And the detection results from the comparison methods refer to the primary detection/separation results without any post-processing. All detection results are scaled to $[0, 255]$ to ensure the fairness for evaluation and comparison.

E. Quantitative Evaluations

The proposed method is experimentally compared with the comparison methods on detecting diverse lesions in *Kaggle*

dataset and *Messidor dataset*, as shown in Fig.3 and Fig.4, respectively.

For *Kaggle dataset* mainly containing complex fundus images, the proposed method outperforms both reconstruction-based MSSCL and PCA models and separation-based MoG model, evidenced by achieving the highest AUC value (0.9936) and MAP value (0.8871). And the performance of PCA model is the least satisfactory due to its simplicity. The performance of the unsupervised deep learning-based methods in comparison does not achieve the expected results on detecting diverse lesions.

For *Messidor dataset* that mainly contains simple fundus images, the overall performance of all the methods is improved. While the proposed method achieves the leading performance with 0.9973 AUC value and 0.9129 MAP value, the separation-based MoG obtains the comparable performance. As contrast, the performance of reconstruction-based methods and deep learning methods is yet to be improved.

F. Lesions Detection Performance Comparison

Fig.5 illustrates the detection results from eight representative fundus images Fig.5(a)-(h) including the anomalies ranging from simple cases with large and high-contrast lesions to complex cases with multi-sizes and multi-contrast lesions.

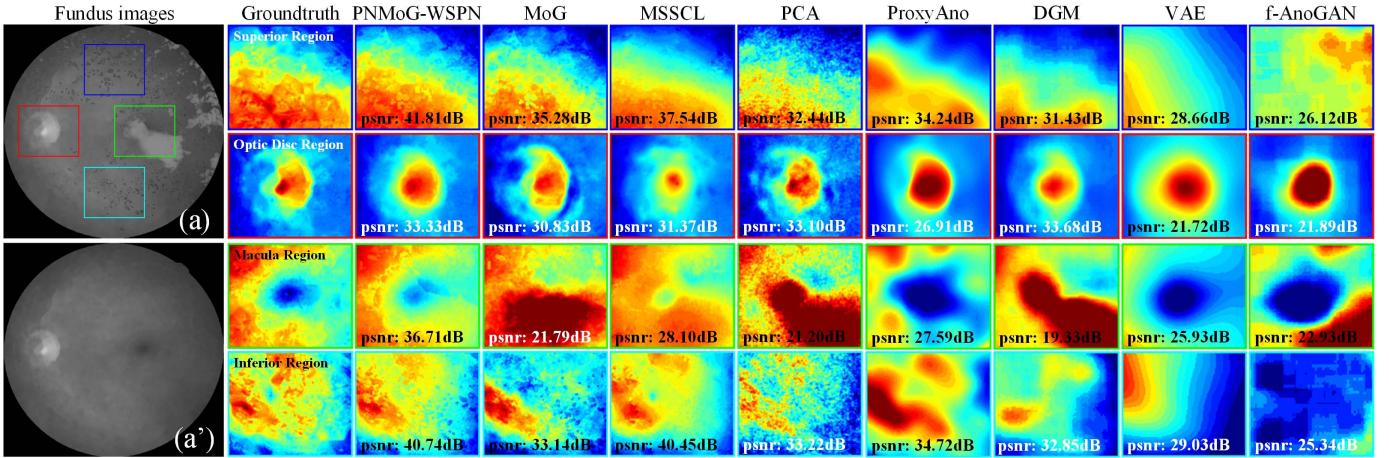


Fig. 6. The background image reconstruction performance of the proposed method and comparison methods. (a) and (a') are the pseudo-abnormal fundus image and its groundtruth. The groundtruth regions and the reconstructed regions are transformed to the pseudo color graphs.

For the simple case (a), the proposed method, MoG, MSSCL, ProxyAno, DGM and f-AnoGAN all achieve comparable performance. For cases (b)-(c), the proposed method, MoG and MSSCL perform well. PCA and the deep learning methods degenerate in these cases. For the most complex cases (d)-(e), while the comparison methods generate more false-positives and false-negatives, the proposed method performs consistently better in detecting lesions while suppressing false-positives.

Cases (f)-(g) mainly contain image glares. For case (f) with a pure image glare, the proposed method, MoG, and MSSCL methods all have comparable performance in detecting the image glare however with some false-positives. For case (g) with both image glares and lesions, the proposed method outperforms others, demonstrating its capacity on detecting more complex cases with mixed lesions and glares.

Case (h) is a challenging case with retinal scars and exhibiting a large number of low-contrast lesions, as annotated by the yellow arrows. All methods experience performance down-turn and the detection results suffer from false-negatives leading to low sensitivity values. As contrast, the proposed method outperforms in suppressing the false-positives.

G. Fundus Image Background Reconstruction Comparison

To qualitatively and quantitatively evaluate the fundus image background reconstruction performance, we have tested all methods which can generate image backgrounds on *Artificial dataset*. Four representative regions of fundus images are used to evaluate the performance, including the optic disc regions (annotated by red blocks), the macula regions (green blocks), the superior regions (blue blocks) and the inferior regions (cyan blocks). One fundus image is used to illustrate the reconstruction results with PSNR indexes of all methods, as shown in Fig.6.

For all of the four regions, the proposed method achieves the most similar manifestation with the groundtruth regions and the remarkable PSNR values. In comparison, the MoG method achieves the comparable reconstruction accuracy in the optic

disc and superior regions, while facing challenges in reconstructing the macula and inferior regions. The MSSCL method erroneously encompasses some high-contrast lesions in the reconstructed macula region of Fig.6. The PCA method shows punctate results in the reconstructed regions. The ProxyAno has the structurally alike reconstruction results comparing with groundtruth images, however, losing fine details. The DGM has difficulty in suppressing lesions in macular region. While the VAE suffers from losing and smoothing out the fine details in the reconstruction results. The f-AnoGAN distorts the reconstructions in all the four regions.

Quantitative comparisons in terms of four performance metrics, including PSNR, MIND [34], VIF [35] and SSIM [36], are listed in Table.I. As shown in Table.I, the proposed method outperforms the comparison methods in all the four metrics. Since these metrics focus on both details and structures, the superiority of the proposed method in accurately reconstructing the fundus image backgrounds and discriminating NPV from lesions can be substantiated.

H. The Effect of Different Patch Sizes

The performance of the proposed method under different patch sizes is listed in Table.II. For computational convenience, we have re-sampled different patch sizes into 2×2 to conduct the experiments. Table.II demonstrates the overall robustness of the performance over different patch sizes. However, for the smaller patch sizes, i.e., 2×2 and 3×3 , the proposed method achieves higher performance. The underlying reason is that the smaller patch size is able to represent both small and big lesions, while the bigger patch size may include and be influenced by the surrounding background when representing the small lesions, which hence results in the decreased performance for small lesions.

I. Time and Memory Cost Comparison

The practical training and running time, and memory cost of the proposed method and the comparison methods are listed in Table.III. The batch size of all deep learning methods is set to 1 and the number of normal instances used in all methods

TABLE I
QUANTITATIVE COMPARISONS OF THE FUNDUS IMAGE BACKGROUND RECONSTRUCTION OF ARTIFICIAL DATASET

Regions	Metrics	Our Method	MoG	MSSCL	PCA	ProxyAno	DGM	VAE	f-AnoGAN
Optic Disc	PSNR	36.36±5.27	34.71±2.19	34.01±4.78	36.07±1.88	27.01±2.42	35.53±1.92	24.23±2.99	21.02±1.48
	MIND	66.58±12.22	93.29±3.26	74.51±5.24	91.82±2.01	98.17±2.42	97.02±2.47	100.08±1.91	105.78±1.43
	VIF	0.76±0.12	0.44±0.09	0.56±0.09	0.26±0.08	0.33±0.06	0.34±0.05	0.37±0.10	0.13±0.04
	SSIM	0.96±0.02	0.94±0.01	0.95±0.02	0.93±0.02	0.92±0.02	0.94±0.01	0.91±0.02	0.86±0.01
Macula	PSNR	48.42±4.55	41.22±4.15	47.91±4.42	39.56±4.30	35.71±2.46	40.72±4.48	32.55±3.04	23.75±1.49
	MIND	61.35±13.42	100.95±4.50	64.11±9.88	97.14±1.40	104.96±2.75	107.95±1.36	106.67±2.76	107.95±1.36
	VIF	0.99±0.15	0.55±0.16	0.95±0.11	0.12±0.05	0.42±0.01	0.31±0.08	0.37±0.12	0.09±0.03
	SSIM	0.99±0.00	0.98±0.02	0.99±0.00	0.96±0.02	0.98±0.01	0.98±0.02	0.98±0.01	0.93±0.01
Superior	PSNR	48.77±3.90	41.92±2.55	47.94±3.96	40.33±2.74	37.74±3.09	41.50±2.89	33.22±3.79	26.79±1.59
	MIND	61.45±13.47	104.09±3.76	67.20±9.26	96.82±0.87	109.14±1.67	109.85±1.26	111.93±0.83	111.42±0.76
	VIF	1.09±0.19	0.58±0.19	1.00±0.13	0.07±0.03	0.38±0.12	0.21±0.05	0.30±0.11	0.07±0.02
	SSIM	0.99±0.00	0.98±0.00	0.99±0.00	0.96±0.01	0.98±0.01	0.98±0.01	0.98±0.01	0.96±0.01
Inferior	PSNR	48.12±4.43	41.06±2.16	46.27±3.43	39.78±2.36	35.88±3.19	40.67±2.42	33.34±3.29	27.81±1.97
	MIND	59.98±14.10	102.05±4.03	67.24±7.83	96.07±0.94	108.30±2.09	108.73±1.76	111.70±0.94	110.90±0.78
	VIF	1.12±0.15	0.60±0.19	0.98±0.11	0.08±0.03	0.44±0.12	0.25±0.06	0.31±0.13	0.09±0.03
	SSIM	0.99±0.00	0.98±0.00	0.99±0.00	0.96±0.01	0.98±0.01	0.98±0.01	0.97±0.01	0.96±0.01

The degree of similarity between two comparison images is positively correlated with the values of PSNR, VIF and SSIM, while is negatively correlated with the value of MIND.

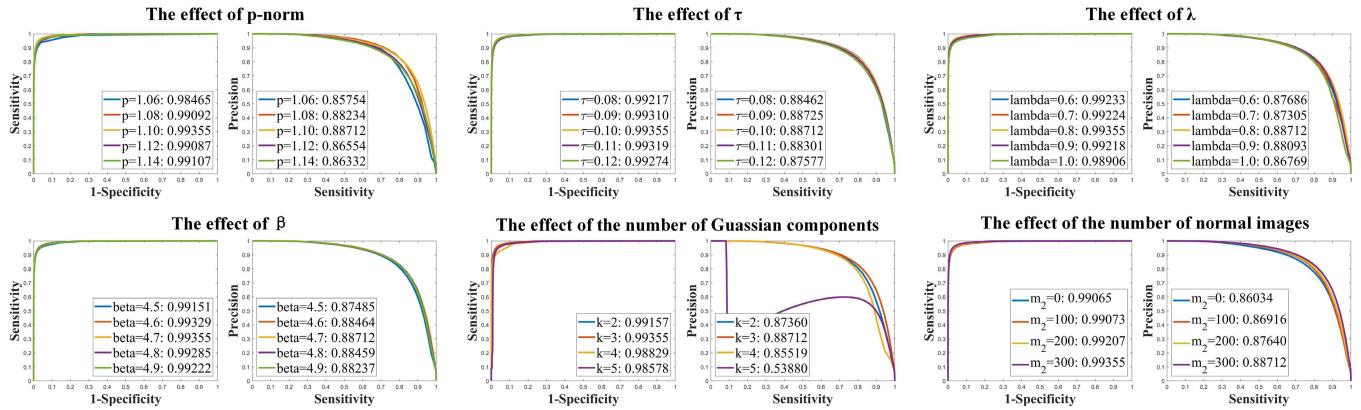


Fig. 7. The ROC and PR curves corresponding to the different values of the parameters in PNMoG-WSPN model.

TABLE II
PERFORMANCE UNDER DIFFERENT PATCH SIZES

Patch Size	2×2	3×3	4×4	5×5
AUC	0.9936	0.9923	0.9915	0.9912
MAP	0.8871	0.8809	0.8708	0.8679

TABLE III
THE TRAINING/RUNNING TIME AND MEMORY COST

Groups	Methods	Training time	Running time	Memory cost
Deep Learning	DRAEM	5h	22.8s	5Gb
	f-AnoGAN	4h	2.2s	3Gb
	VAE	5h	4.5s	3Gb
	DGM	2h	2.5s	3Gb
Background Reconstruction	ProxyAno	52h	13.0s	5Gb
	PCA	2min	5.0s	2.8Gb
Background Separation	MSSCL	1.8h	50min	2.2Gb
	MoG	-	18mins	2.6Gb
	Ours ($m_2=1000$)	-	15min	2.0Gb
Background Separation	Ours ($m_2=300$)	-	6min	1.5Gb

is set to 1000. All experiments are run on a workstation with Intel (R) Core (TM) i9-10920X CPU @ 3.50 GHz with

GeForce RTX 3090 GPU. The proposed method and MoG method do not require a training phase. While PCA and deep learning methods require the shortest running time counted in seconds, their detection performance is less competitive compared with the proposed method. Particularly, the proposed method achieves favorable results even under 300 normal instances, which requires a shorter running time (6 minutes) and a smaller memory cost (1.5 Gb).

J. Robustness and Generalizability of the Parameters

The robustness of the six parameters in the proposed method is tested in *Kaggle dataset* and shown in Fig.7. These six parameters include (1) p and τ for controlling the low-rank degree of fundus image backgrounds, (2) λ and β for the trade-off of the weights of PNMoG module and l_1 -norm, respectively, (3) k for the number of components of Gaussian, and (4) m_2 for the number of normal fundus images. The parameters are overall robust around their default settings in *Kaggle dataset*. The parameter k influences the weight of PNMoG as shown in Eq.(16), which explains the rapidly drop of MAP value when $k = 5$.

Among the six parameters, five parameters share the consistent setting across both *Kaggle dataset* and *Messidor dataset*,

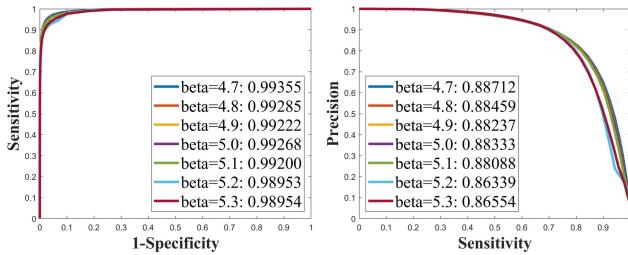


Fig. 8. The generalization ability of β on *Kaggle* dataset.

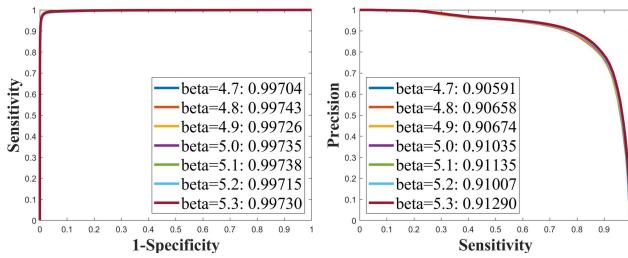


Fig. 9. The generalization ability of β on *Messidor* dataset.

TABLE IV
THE ABLATION EXPERIMENTS

Method	AUC	MAP
l_1 -norm+Nuclear-Norm (Robust PCA, baseline)	0.9851	0.8702
l_1 -norm+WSPN	0.9932	0.8737
$MoG+UV^T$	0.9906	0.8387
$MoG+l_1$ -norm+WSPN	0.9921	0.8753
PNMoG+ l_1 -norm+Nuclear-Norm	0.9916	0.8774
PNMoG+ l_1 -norm+WSPN	0.9936	0.8871

while, the parameter β has different optimal values (4.7 and 5.3) in *Kaggle dataset* and *Messidor dataset*. Here, we show the generalization of β in the value range of [4.7, 5.3]. As shown in Fig.8 and Fig.9, in both *Kaggle dataset* and *Messidor dataset*, the performance of the proposed method remains robust when altering β in the value range of [4.7, 5.3] with a step of 0.1. Initializing the model around the default settings for all six parameters can achieve the favorable results.

K. Ablation Experiments

Ablation experiments are conducted over *Kaggle dataset* for evaluating the overall performance. As shown in Table.IV, the components for lesions regularization include PNMoG, MoG and l_1 -norm, and the components for backgrounds constraint include WSPN, Nuclear norm and UV^T basis.

The baseline of Robust PCA decomposition is displayed as the 1st row, and the 6th row corresponds to the proposed method. With PNMoG-WSPN model, the MAP value increases from 0.8702 to 0.8871. The effectiveness of PNMoG is demonstrated via the comparison between the 2nd row and the 6th row with the MAP value increasing from 0.8737 to 0.8871. The effectiveness of WSPN module is validated through the comparison between the 5th and the 6th row with the increased MAP value.

V. DISCUSSIONS AND ANALYSES

Experiments and comparisons demonstrate the effectiveness of the proposed method on the anomaly detection for fundus

images. Our major findings include that the PNMoG-WSPN model is able to (1) more precisely detect anomalies via the individualized modeling for diverse lesions and accurate background learning, and (2) outperform in detecting anomalies in complex fundus images.

A. Analyses on the PNMoG-WSPN Model

Firstly, PNMoG module coherently provides individualized lesions modeling and takes into account of lesion's regional patch-like properties to achieve accurate detection of diverse lesions, as illustrated in Fig.3 and Fig.4. As shown in Fig.1, the lesions statistic distributions in different fundus images are individualized due to their different sizes and colors. The non-i.i.d. modeling scheme in PNMoG contributes to the better fitting to these diverse statistical variations, as demonstrated in Fig.5. Meanwhile, the PNMoG module encompasses lesion's regional properties by encoding patches as units to conduct the statistical calculations, and thus, is capable to more completely portray lesions' properties to improve the detection.

Secondly, as a new metric, WSPN is used for constraining low-rank structure decomposition from the fundus images. The advantages of WSPN are two-fold. On one hand, since various lesions generally adhere to sparse structures while normal fundus images and fundus image backgrounds adhere to low-rank structures, based on the low-rank decomposition mechanism, most of various lesions and fundus image backgrounds will be naturally decomposed into sparse images and low-rank images, respectively. On the other hand, since WSPN can provide a relaxation low-rank constraint for the images, normal regions with NPV are more likely to be decomposed into low-rank images rather than sparse images. As demonstrated in Fig.6, with WSPN, the NPV are better discernible from lesions. According to [37], the weak and small lesions usually represent early stages of diseases, thus precisely discriminating the lesions from NPV would be helpful for early stage diagnoses.

Thirdly, the reciprocal reinforcement of background learning (by WSPN) and discrimination of lesions with diverse features (by PNMoG) facilitates the discernment of our method for the diverse anomaly detection task. The WSPN is capable to encompass NPV while isolating most lesions. The precise modeling on lesions by PNMoG further reinforces more accurate learning of fundus image backgrounds by WSPN. By such, suitable constraints and metrics are effectively applied to low-rank decomposition of the fundus images for more accurate anomaly detection, which has been well validated by our thorough experimental comparisons. More recently, the contrastive learning (CL) is proposed for object segmentation or lesions detection [38], [39], [40]. While CL has merits in alleviating the needs on annotation, it is different from our method in terms of the underlying mechanisms. More specifically, CL reciprocally learns the features of the same object through two sub-networks, whereas our method regularizes both the lesions and the backgrounds by different mechanisms.

B. Analyses on Detection in Complex Fundus Images

1) *Analyses on Detection Ability of Lesions in Complex Fundus Images:* Complexity of fundus images is imposed by (1)

TABLE V
THE DETECTION PERFORMANCE OF MULTIPLE TYPES OF LESIONS IN COMPLEX FUNDUS IMAGES

Number of Lesion Types per Image	Number of Images	PNMoG-WSPN	MoG	MSSCL	PCA	ProxyAno	DGM	VAE	f-AnoGAN	DRAEM
1	53	0.7893	0.7085	0.7079	0.4897	0.5632	0.5886	0.5296	0.2564	0.4242
2	52	0.9060	0.8545	0.8836	0.6613	0.7362	0.7480	0.7032	0.3823	0.5299
3	51	0.9252	0.9072	0.9004	0.7101	0.8132	0.8271	0.6168	0.5298	0.5874
4	33	0.9301	0.8993	0.9012	0.7490	0.7692	0.8015	0.6001	0.3548	0.6735
5	4	0.8746	0.6843	0.8042	0.6513	0.7301	0.5626	0.2070	0.4229	0.5568

The 53 fundus images in which each image contains one particular type of diverse lesion types (1st row) consist of 1 fundus image with soft exudates, 13 fundus images with hard exudates, 4 fundus images with macular edemas, 1 fundus image with peripapillary atrophy, 9 fundus images with glares, 4 fundus images with crystallines, 8 fundus images with laser scars, and 13 fundus images with retinal scars.

TABLE VI
THE DETECTION PERFORMANCE OF SPECIFIC TYPES OF LESIONS

Lesion type	Proportion	PNMoG-WSPN	MoG	MSSCL	PCA	ProxyAno	DGM	VAE	f-AnoGAN	DRAEM
Hemorrhage	10.82%	0.7111	0.6540	0.6373	0.3242	0.4178	0.4850	0.0022	0.1773	0.2226
Hard exudate	35.78%	0.8981	0.8600	0.8743	0.6985	0.7744	0.7650	0.7264	0.4149	0.5154
Soft exudate	5.62%	0.9114	0.8164	0.8546	0.5682	0.7596	0.7611	0.6446	0.5936	0.1721
Retinal scar	14.35%	0.7326	0.5694	0.6596	0.4099	0.4774	0.4503	0.3239	0.2989	0.3477
Image glare	6.77%	0.8374	0.8211	0.7520	0.3356	0.5213	0.5759	0.5411	0.3416	0.2448

"Proportion" denotes the ratio of the area of the certain type of lesion to the area of all lesions.

multiple types of lesions in one fundus image, or (2) a wide variety of lesion types in different fundus images. The methods are validated on the experiments for detecting lesions in the complex fundus images in *Kaggle dataset*.

As shown in the last two rows in **Table.V**, when more than four types of different lesions are contained in one fundus image, our method outperforms SOTA methods with more than 3% higher MAP value. Additionally, as shown in the first row of **Table.V**, when the fundus images contain diverse lesion types, where each image only contains one particular type of these lesion types, the MAP value of the proposed method is at least 8% higher than those of the SOTA methods. The performance gains validate the effectiveness of the individualized modeling strategy to better describe and specifically adapt the complexity and diversity of lesions in one fundus image or among different fundus images.

2) Analyses on Detection Ability of Specific Lesions: Five types of lesions as shown in **Table.VI** are annotated at pixel-level in *Kaggle dataset* for evaluating and comparing the detection ability of different methods on specific types of lesions. The MAP values of different methods for the five types of lesions are listed in **Table.VI**.

The proposed method consistently achieves superior performance in all five types of lesions, and shows far better detection performance in three types of lesions (hemorrhages, soft exudates, and retinal scars) in comparison with other methods. However, the detection performance of different methods may vary across different lesion types. For hard exudates and image glares, the proposed method has comparable results with MSSCL and MoG, and they all achieve remarkable performance in comparison with other methods. For hemorrhages and retinal scars, since they have a wide range of intensities, shapes, and sizes, the performance of all methods has degenerated. However, the proposed method retains remarkable improvements compared with other methods. That indicates the non-i.i.d. MoG also outperforming in

modeling certain types of lesions than i.i.d. MoG. The reason may be that a certain type of lesions can slightly vary among different images, the non-i.i.d. MoG is capable to capture and fit these variations.

VI. CONCLUSION

To address the challenges posed by lesions diversities and normal personalized variations (NPV) in anomaly detection, we propose a cohesive PNMoG-WSPN model to simultaneously model the individualized properties of lesions and accurately learn the fundus image backgrounds. The PNMoG adaptively models the lesions' individualized variations in terms of statistical distributions and patch-like structural characteristics, and the WSPN enhances the accuracy of the decomposed fundus image backgrounds with the relaxation low-rank measurement. The experimental comparisons validate that the proposed method outperforms state-of-the-art methods, and contributes to a better detection performance for diverse anomalies in the fundus images.

In the future work, it is worthwhile to invest further research dedication to explore how to adaptively fit the patch sizes and automatically find the optimal parameters for different datasets.

REFERENCES

- [1] I. G. Morgan, K. Ohno-Matsui, and S.-M. Saw, "Myopia," *Lancet*, vol. 379, no. 9827, pp. 1739–1748, 2012.
- [2] T. Fricke et al., "Global prevalence of visual impairment associated with myopic macular degeneration and temporal trends from 2000 through 2050: Systematic review, meta-analysis and modelling," *Brit. J. Ophthalmology*, vol. 102, pp. 855–862, Jul. 2018.
- [3] D. S. W. Ting et al., "Deep learning in ophthalmology: The technical and clinical considerations," *Prog. Retinal Eye Res.*, vol. 72, Sep. 2019, Art. no. 100759.
- [4] S. Sengupta, A. Singh, H. A. Leopold, T. Gulati, and V. Lakshminarayanan, "Ophthalmic diagnosis using deep learning with fundus images—A critical review," *Artif. Intell. Med.*, vol. 102, Jan. 2020, Art. no. 101758.
- [5] C. You et al., "Incremental learning meets transfer learning: Application to multi-site prostate MRI segmentation," 2022, *arXiv:2206.01369*.

- [6] Y. Xia et al., "Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation," *Med. image Anal.*, vol. 65, Oct. 2020, Art. no. 101766.
- [7] B. Chen et al., "Diverse lesion detection from retinal images by subspace learning over normal samples," *Neurocomputing*, vol. 297, pp. 59–70, Jul. 2018.
- [8] B. Chen et al., "Abnormality detection in retinal image by individualized background learning," *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107209.
- [9] R. Wang, B. Chen, D. Meng, and L. Wang, "Weakly supervised lesion detection from fundus images," *IEEE Trans. Med. Imag.*, vol. 38, no. 6, pp. 1501–1512, Jun. 2019.
- [10] A. Rocha, T. Carvalho, H. F. Jelinek, S. Goldenstein, and J. Wainer, "Points of interest and visual dictionaries for automatic retinal lesion detection," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2244–2253, Aug. 2012.
- [11] N. Weiss, D. Rueckert, and A. Rao, "Multiple sclerosis lesion segmentation using dictionary learning and sparse coding," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2013, pp. 735–742.
- [12] D. Meng and F. De La Torre, "Robust matrix factorization with unknown noise," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1337–1344.
- [13] X. Cao, Q. Zhao, D. Meng, Y. Chen, and Z. Xu, "Robust low-rank matrix factorization under general mixture noise distributions," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4677–4690, Oct. 2016.
- [14] W. Wei, L. Yi, Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Should we encode rain streaks in video as deterministic or stochastic?" in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2535–2544.
- [15] G. Pang, C. Shen, L. Cao, and A. Van Den Hengel, "Deep learning for anomaly detection," *ACM Comput. Surveys (CSUR)*, vol. 54, no. 2, pp. 1–38, 2021.
- [16] T. Schlegl, P. Seeböck, S. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Proc. IPMI*, 2017, pp. 146–157.
- [17] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Med. Image Anal.*, vol. 54, pp. 30–44, May 2019.
- [18] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2016, *arXiv:1511.06434*.
- [19] K. Zhou et al., "Proxy-bridged image reconstruction network for anomaly detection in medical images," *IEEE Trans. Med. Imag.*, vol. 41, pp. 582–594, 2022.
- [20] Y. Tang, Y. Tang, Y. Zhu, J. Xiao, and R. M. Summers, "A disentangled generative model for disease decomposition in chest X-rays via normal image synthesis," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101839.
- [21] V. Zavrtanik, M. Kristan, and D. Skocaj, "DRAEM—A discriminatively trained reconstruction embedding for surface anomaly detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 8310–8319.
- [22] F. Nie, H. Huang, and C. Ding, "Low-rank matrix recovery via efficient schatten p -norm minimization," in *Proc. AAAI*, 2012, pp. 1–7.
- [23] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, Mar. 2005.
- [24] J. H. Tan, U. R. Acharya, S. V. Bhandary, K. C. Chua, and S. Sivaprasad, "Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network," *J. Comput. Sci.*, vol. 20, pp. 70–79, May 2017.
- [25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2015.
- [26] A. Telea, "An image inpainting technique based on the fast marching method," *J. Graph Tools*, vol. 9, no. 1, pp. 23–34, 2004.
- [27] Y. Xie, S. Gu, Y. Liu, W. Zuo, W. Zhang, and L. Zhang, "Weighted schatten p -norm minimization for image denoising and background subtraction," *IEEE Trans. Image Process.*, vol. 25, pp. 4842–4857, 2016.
- [28] A. P. Dempster and N. M. Laird, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc., Ser. B*, vol. 39, pp. 1–38, Sep. 1977.
- [29] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Nov. 2010.
- [30] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 217–224.
- [31] E. Decencière et al., "Feedback on a publicly distributed database: The messidor database," *Image Anal. Stereol.*, vol. 33, no. 3, pp. 231–234, 2014.
- [32] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2014, *arXiv:1312.6114*.
- [33] T. Walter, J.-C. Klein, P. Massin, and A. Erginay, "A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina," *IEEE Trans. Med. Image*, vol. 21, no. 10, pp. 1236–1243, Oct. 2002.
- [34] M. P. Heinrich et al., "Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration," *Med. image Anal.*, vol. 16, no. 7, pp. 1423–1435, 2012.
- [35] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [36] 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.
- [37] E. T. D. R. S. R. Group, "Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified airlie house classification: ETDRS report number 10," *Ophthalmology*, vol. 127, no. 4, pp. 99–119, 2020.
- [38] C. You, Y. Zhou, R. Zhao, L. Staib, and J. S. Duncan, "SimCVD: Simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2228–2237, Sep. 2022.
- [39] C. You, W. Dai, L. Staib, and J. S. Duncan, "Bootstrapping semi-supervised medical image segmentation with anatomical-aware contrastive distillation," 2022, *arXiv:2206.02307*.
- [40] C. You, R. Zhao, L. H. Staib, and J. S. Duncan, "Momentum contrastive voxel-wise representation learning for semi-supervised volumetric medical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2022, pp. 639–652.