This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2934477, IEEE Journal of Biomedical and Health Informatics

Direct Cup-to-Disc Ratio Estimation for Glaucoma Screening via Semi-supervised Learning

Rongchang Zhao, Xuanlin Chen, Xiyao Liu, Zailiang Chen, Fan Guo and Shuo Li*

***Abstract*—Glaucoma is a chronic eye disease that leads to irreversible vision loss. The Cup-to-Disc Ratio (CDR) serves as the most important indicator for glaucoma screening and plays a significant role in clinical screening and early diagnosis of glaucoma. In general, obtaining CDR is subjected to measuring on manually or automatically segmented optic disc and cup. Despite great efforts have been devoted, obtaining CDR values automatically with high accuracy and robustness is still a great challenge due to the heavy overlap between optic cup and neuroretinal rim regions. In this paper, a direct CDR estimation method is proposed based on the well-designed semi-supervised learning scheme, in which CDR estimation is formulated as a general regression problem while optic disc/cup segmentation is cancelled. The method directly regresses CDR value based on the feature representation of optic nerve head via deep learning technique while bypassing intermediate segmentation. The scheme is a two-stage cascaded approach comprised of two phases: unsupervised feature representation of fundus image with a convolutional neural networks (MFPPNet) and CDR value regression by random forest regressor.**

The proposed scheme is validated on the challenging glaucoma dataset Direct-CSU and public ORIGA, and the experimental results demonstrate that our method can achieve a lower average CDR error of 0.0563 and a higher correlation of around 0.726 with measurement before manual segmentation of optic disc/cup by human experts. Our estimated CDR values are also tested for glaucoma screening, which achieves the areas under curve of 0.905 on dataset of 421 fundus images. The experiments show that the proposed method is capable of state-of-the-art CDR estimation and satisfactory glaucoma screening with calculated CDR value.

***Index Terms*—Cup-to-disc ratio (CDR), representation learning, direct estimation, glaucoma screening, semi-supervised learning**

I. Introduction

OPTIC nerve head (ONH) assessment is one of the most clinically significant screening technique for glaucoma, which is a chronic eye disease leading to irreversible vision loss with the progressively damaged of the optic nerve. Based on the ONH assessment, one identifies the glaucomatous from normal cases via manual measurement of ONH geometric structures. In this procedure, some measurements are proposed as clinical conditions for glaucoma screening, such as vertical Cup-to-Disc Ratio (CDR) [1], [2], disc diameter, rim area and ISNT rule. The CDR, calculated as the ratio of the vertical cup diameter (VCD) to the vertical disc diameter (VDD) as shown in Fig.1, is clinically accepted as the significant and most used indicator for glaucoma screening since its precision has been proved for many years.

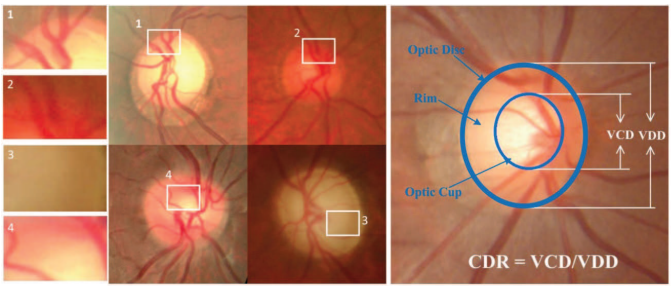


Figure 1. Illustration of the challenges of CDR estimation. Left: Issues, such as extremely low contrast with the surrounding tissues, serious interference of retinal vessel, unintelligible boundary and heavy overlap between optic cup and rim, are hard to handle with existing segmentation-based algorithm. Right: the CDR value to be estimated in this work.

In practice, obtaining reliable CDR value is subjected to measuring on segmented optic disc and cup, which usually obtained by manually contouring the borders of optic disc/cup or manual correction of contours generated by segmentation algorithms [3]–[6]. However, manually contouring of the optic disc/cup borders is time-consuming and subjective to personal experiences. Due to the lack of sharp border information of optic disc/cup, the CDR value of the same subject often varies among different clinicians. Recently, great efforts have been devoted into automatizing the procedure [7]–[11]. Numerous automated segmentation methods are proposed as a prerequisite to segregate disc/cup regions from the complex surroundings with clear borders, including statistical shape model [12]–[15], multiview and multimodal approaches [16], [17], superpixel-based methods [18], [19], and deep learning methods [10], [20]–[22]. Usually, these methods require strong prior information and user interaction to increase its accuracy.

Although the segmentation-based methods obtained effective performance by leveraging state-of-the-art machine learning especially deep learning techniques, accurate measurement of CDR value is still a challenging task (Fig.2 (a)) due to 1) heavy overlap and extremely weak contrast between optic cup and neuroretinal rim regions which make the automated segmentation algorithm can not distinguish the boundaries of optic cup in fundus image; 2) the great variability of shape and inhomogeneity in appearance of optic disc which leads to critical inconsistency of the measured CDR compared with the actual one; 3) insufficient pixel-level labels (e.g. segmentation

This work was supported by the National Science Foundation of China (61702558), Hunan Science and Technology Project (2017WK2074), and National Key R&D Program of China (2017YFC0840104). Asterisk indicates corresponding author.

R. Zhao, X. Chen, X. Liu, Z. Chen and F. Guo are with School of Computer Science and Engineering, Central South University, 410083, China (e-mail: byrons.zhao@gmail.com)

S. Li is with the Department of Medical Imaging, Western University, London, ON N6A 3K7, Canada. S. Li is also with the Digital Imaging Group of London, London, ON N6A 3K7, Canada. (e-mail: slishuo@gmail.com).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2934477, IEEE Journal of Biomedical and Health Informatics

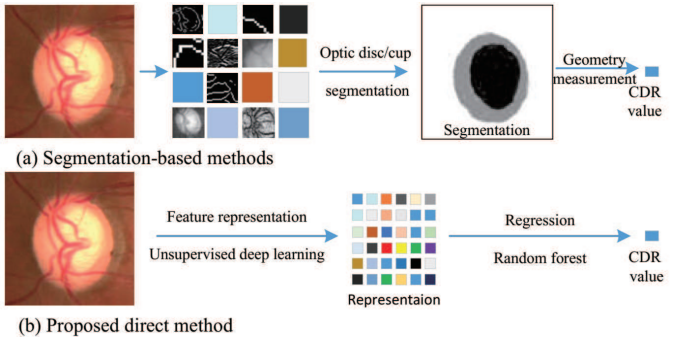


Figure 2. (a) Traditional segmentation-based methods measure CDR from the segmented result which requires strong prior; (b) The proposed regression-based method which estimates CDR values directly from fundus image while skips the intermediate procedure of segmentation. The proposed method follows a two-stage cascaded approach to unsupervised fundus image representation learning from abundant unlabeled data

mask) which is not enough to learn the outstanding model for optic disc/cup segmentation and CDR measurement; and 4) the error introduced by the intermediate steps which seriously affects the accuracy of the calculated CDR value. Even more challenges arise from the presence of various pathologies.

Direct quantification [23]–[26] without segmentation has great potential to estimate the CDR value since it predicts ultimate CDR based on the powerful representation of ONH structure while bypasses the ill-conditioned segmentation. However, existing direct methods are unsuitable for handling the challenges of CDR estimation due to several drawbacks. They 1) rely on a well-designed classifier to discriminate the fundus image as glaucomatous or normal case based on the feature representation by feature engineering, which cannot provide the clinical condition or interpretative index for physicians and patients of how the ONH assessment is made; 2) are still far from being a satisfactory tool for CDR estimation since complex model is lack of sufficient labeled data to learn the optimal parameters while general model only perform well on small datasets due to overfitting.

To overcome those issues, we propose a general method for CDR value estimation directly from fundus image while skipping the intermediate procedure of segmentation (as shown in Fig.2 (b)). This method relies on a semi-supervised learning scheme which is composed of unsupervised representation learning (MFPPNet) and supervised CDR value regression. The proposed scheme benefits in the following aspects: 1) an effective regression-based approach is proposed to directly estimate CDR value from fundus image, which eliminates the CDR error introduced by intermediate segmentation algorithm; 2) a deep feature representation is learned with well-designed MFPPNet to represent the task-oriented geometrical structure and appearance of ONH closely related to CDR value in fundus images; 3) an effective unsupervised learning strategy is designed to optimize the parameters of CNNs-based MFPPNet with the abundant unlabeled data; 4) random forests are employed as a regressor which provides a efficient way to model the complex relationships between the deep features and CDR values.

The proposed scheme combines the strengths of both unsupervised representation learning and supervised variable regression. The main contributions of this work are as follows:

- We propose an effective regression-based method to directly estimate the CDR value from fundus image for ONH assessment in glaucoma screening. This method provides for the first time a reliable solution for direct quantitative assessment of ONH while bypassing the extraneous error introduced by intermediate segmentation.
- We propose a novel semi-supervised learning scheme for efficient CDR estimation. The proposed framework combines the strengths of both unsupervised (for feature representation) and supervised (for CDR regression) learning to estimate accurate CDR value with abundant incomplete labeled data. This scheme can be easily transformed to other applications with small amount of labeled training data in medical image analysis.
- We propose a novel multi-scale convolutional network (MFPPNet) for unsupervised fundus image representation learning from unlabeled data.

The remainder of this paper is organized as follows. Section II details the proposed MFPPNet, including the network architecture and the objective function. Section III gives detailed procedure of CDR estimation using regression forests. The experiments and results are reported and analyzed in Section IV and V respectively. Conclusions is given in Section VI.

II. Unsupervised Representation Learning by MFPPNet

The proposed model (MFPPNet, Fig.3) is a unsupervised representation learning framework composed of a densely connective network for basic feature embedding, a pyramid pooling module for multi-scale contextual information exploitation, and a fully-connected fusion layer for fixed-length feature representation. To fully leverage the benefit of complementary information from plenty of incomplete labeled data, our method conducts a multiple objective optimization by aggregating multi-scale pyramid pooling and unsupervised representation learning. The main components of MFPPNet are presented in detail.

A. MFPPNet architecture

Dense connectivity. The dense connectivity [27] is adopted in our network for basic feature extraction since this connectivity pattern allows the network to reuse and bypass existing features from prior layers and ensures high accuracies in later layers. In each Dense Block, let x_i denotes the output of the i-th layer, dense connectivity can be described as $x_i = \mathcal{F}_i([x_1, x_2, ..., x_{i-1}], \mathcal{W}_i)$. It connects each layer with all subsequent layers and allows later layers to bypass features optimized for the short-term. Our MFPPNet implements 3 dense blocks, and the transition layer is inserted between adjacent dense blocks to adjust the resolution of feature maps, which consists of a batch normalization layer and an 1×1 convolutional layer followed by a 2×2 average pooling layer. Each dense block having an equal number of layers is defined following the design in DenseNets [27], and we set the number of output channels of the three scales to 6, 12 and 24, respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2934477, IEEE Journal of Biomedical and Health Informatics

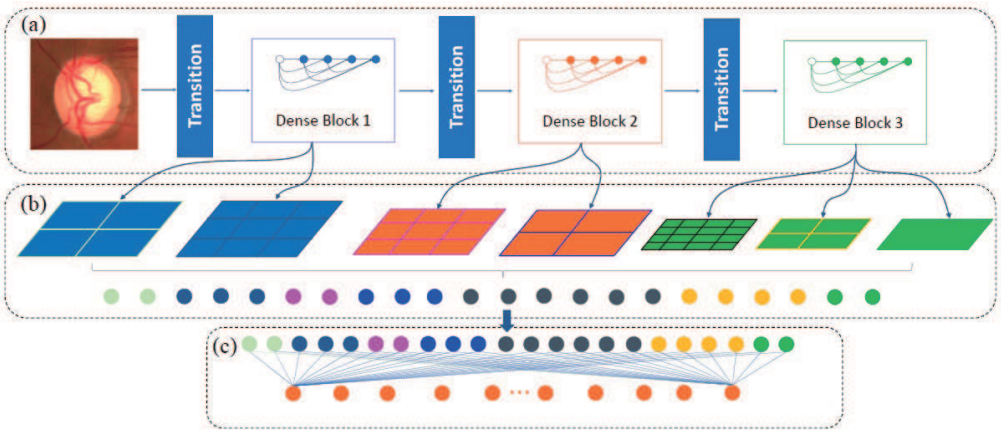


Figure 3. Overview of our proposed MFPPNet, which composed of three distinguished components: (a) densely connective encoder for basic feature extraction, (b) feature pyramid pooling module for multi-scale contextual information exploitation, and (c) fully-connected fusion layer for fixed-length feature representation. The loss is backpropagated to optimize the network parameters by minimizing objective function.

Feature pyramid pooling. In recent models, such as SPP-net [28], spatial pyramid pooling often performed to obtain multi-scale feature representation, and achieve state-of-the-art performance. Inspired by those models, we employ a new-designed feature pyramid pooling strategy to characterize the global structure and local contextual information of ONH. The backbone network extract feature information of different scale and increase respective field effectively via the outputs of different dense blocks. The feature pyramid pooling module fuses features from three different pyramid scales by implementing 7 parallel filters with different receptive fields. To better extract contextual information from different pyramid scales, we implement 4×4, 2×2, 1×1 average pooling on the output of the dense block3, while 3×3 and 1×1 average pooling on the output of the dense block2 and dense block1, respectively. The feature pyramid pooling module integrates multi-scale local contextual cues and high-level global structure to produce effective feature representation characterizing the significant patterns, such as OD/OC borders, shapes, regions, and so on.

Fully-connected fusion. To adopt our MFPPNet with fixed dimensional feature vectors, we use a fully-connected layer for final feature fusion and production. The fully-connected layer maps outputs of feature pyramid pooling module into a fixed-length vector for fundus image representation.

In summary, the learning of MFPPNet is an end-to-end non-linear mapping which projects the input fundus image into the binary descriptor by multiple projection functions. What needs to be explained is that the descriptor is designed as binary to enable our objective function available for unsupervised training with unlabeled data (section II-B). Mathematically, $\mathcal{F}(x; \mathcal{W})$ represents the non-linear mapping that composed of n cascaded projection functions and can be written as:

$$\mathcal{F}(x; \mathcal{W}) = f_n(...f_2(f_1(x; w_1); w_2)...; w_n) \tag{1}$$

where f_i represents the i-th projection function which maps the input x_i and w_i into its output x_{i+1}. The output of the objectives layer is $\mathcal{F}(x; \mathcal{W})$. Then the proposed deep model outputs the binary descriptor $\mathbf{b}$ as:

$$\mathbf{b} = 0.5 \times (sign(\mathcal{F}(x; \mathcal{W})) + 1) \tag{2}$$

where $sign$ is the signum function, $sign(k) = 1$ if $k > 0$ and -1 otherwise. Eq.(2) aims to output the resulting binary descriptor in the vector form. In the training stage, the stochastic gradient descent and back-propagation technique are used to learn a set of parameters $\mathcal{W} = (w_1, w_2, ..., w_n)$ that quantizes the input fundus image x into binary descriptor $\mathbf{b}$, and a set of constraint conditions are enforced on the binary descriptor to make the descriptor discriminative and efficient.

B. The Objective Function for Unsupervised Training

To realize unsupervised learning with plenty of unlabeled data, outputs of multi-layer projection functions is constrained when the training algorithm optimizes the parameters $\mathcal{W}$. We enforce a learning objectives on the outputs of the network to achieve the training procedure in an unsupervised manner. The objectives is also a fully-connected layer which transforms the real-value outputs of previous layer into the discriminative feature descriptor with constraint conditions. Significantly, the objective exacts constraints on the parameters optimizing procedure, which enable the pixel-wised label information unnecessary. To achieve representation learning with unlabeled fundus images, the output descriptor is optimized based on two criterions. Firstly, the learned descriptor should precisely represents the geometrical structure and appearance of the optic disc with the activations of the last projection function. Hence, the loss due to the feature representation should be as less as possible after the last projection function. Secondly, the descriptor should have stronger capacity for information representation. In this way, the learned descriptor is discriminative and compact for information representation in the local optic disc with incomplete label information. To fulfill the two criterions, an objective function is designed to optimize parameters $\mathcal{W}$. We formulate the objectives function as following:

$$\min_{\mathcal{W}} \mathcal{L}(\mathcal{W}) = \alpha_1 \mathcal{L}_1(\mathcal{W}) + \alpha_2 \mathcal{L}_2(\mathcal{W}) \tag{3}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2934477, IEEE Journal of Biomedical and Health Informatics

Discrimination. The feature descriptor **b** is expected to represent the complex mixture of hidden patterns related to CDR value of fundus image x_n, so the first criterion for the objective function is discrimination. Therefore, the approach to keep the descriptor discriminative is to minimize the loss between the descriptor and the original fundus image. In the learning, the parameters $\mathcal{W}$ is optimized using the following loss function:

$$\min_{\mathcal{W}} \mathcal{L}_1(\mathcal{W}) = \sum_{n=1}^{N} \|(\mathbf{b} - 0.5) - \mathcal{F}(x_n; \mathcal{W})\|^2 \tag{4}$$

where N is the training data number. Ideally, to keep the binary descriptor informative is to minimize the loss by rewriting Eq.(2) as $\mathbf{b} - 0.5 = \mathcal{F}(x_n; \mathcal{W})$. Therefore, it indicates that the quantification loss should be as less as possible after the binarization operation.

Efficiency. To increase the efficient of the descriptor **b**, we maximize the representation capacity of the binary descriptor based on the entropy conception in information theory. In the binary descriptor, the higher the entropy is, the more information the descriptor represents. Ideally, each bit in the binary descriptor with 50% probability of being one or zero makes the entropy maximum. Therefore, we keep the binary descriptor to be evenly distributed near the distribution with average of 0.5 as the following objective function:

$$\min_{\mathcal{W}} \mathcal{L}_2(\mathcal{W}) = \sum_{m=1}^{M} \|(\mu_m - 0.5)\|^2 \tag{5}$$

where M is the bit length of the binary descriptor and μ_m is the average of the m-th bin of the descriptor string.

$$\mu_m = \frac{1}{N} \sum_{n=1}^{N} \mathrm{b}_n(m) \tag{6}$$

where N is the data number and $\mathrm{b}_n(m)$ is the binary value at m-th bin of the descriptor $\mathbf{b}_n$ from the n-th fundus image.

C. Representation Learning Algorithm

i) Fundus image crop and data augmentation. Since the CDR value measures the geometrical structures and appearance of optic disc/cup in fundus image, optic disc images are imputed to our MFPPNet to learn its discriminative representation. In our experiments, the method of template matching as used in [13] is adopted to locate and crop the fundus image region bounding the optic disc. In addition, each cropped fundus image is normalized with the resolution of 224×224, and the mean value over all the pixels is subtracted from each pixel to remove the influence of illumination variation across images.

To avoid overfitting on image data, we employ data augmentation to artificially enlarge the dataset using image rotation and horizontal reflections. At the data augmentation time, we perform horizontal reflection and image rotation with 10, 5, 0, -5, 10 degrees to create a larger dataset, and then train our model with the enlarged dataset.

ii) Training and testing algorithm. Algorithm 1 summarizes the detail procedure of training and testing for our fundus image representation algorithm. The algorithm contains four parts: data preparation, initialization, training and representation. For a excellent behavior, our MFPPNet is constructed based on the 3-blocks DenseNet [27]. MFPPNet is initialized with parameters pre-trained from ImageNet dataset. Pre-training has been proved to be effective for regularization towards the parameter space and support better generalization. In the training stage, we use stochastic gradient descent (SGD) and back-propagation technique to train our network, therefore the optimized weight $\mathcal{W}$ can be solved by the proposed objectives function. Unlike the supervised approaches that minimize the difference between the output and the label information, our algorithm achieves the back-propagation to minimize the objective function. The representation results can be got by the forward process after the training. The bit length of our binary descriptor is 1024, the fundus images are normalized to 256×256 and then center-cropped to 224×224 as the input.

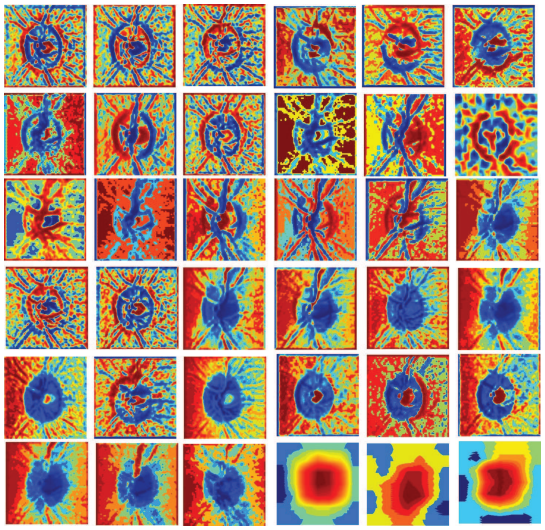


Figure 4. Visual illustration of typical feature representations for the same fundus image. The first three rows are from the DenseBlock 1, the following two rows are from DenseBlock 2, and the last row is from DenseBlock 3 with different pooling scales. Notice that these feature maps favor some specific structures responsive to OC/OC shapes, regions, border cues and other appearances, and the last three maps show a different features because its biggest receptive filed than others in the network.

Algorithm 1:

Input: Training data $X = \{x_1, x_2, ..., x_n\}$
Output: Parameters $\mathcal{W}$ **for deep neural network**
STEP 1: Data Preparation
Optic disc location and data augmentation.
STEP 2: Initialization
Initialize $\mathcal{W}$ with pre-trained weights from ImageNet;
STEP 3: Training
while *iter* $<$ *max-iter* **do**
Fix $\mathcal{W}$ update b_n using (1) and (2);
while *iter1* $<$ *max-iter1* **do**
Fix b_n update $\mathcal{W}$ by minimizing the sum of (3);
return $\mathcal{W}$;
STEP 4: Representation
Input a test image Y, forward propagate the data
through the network with trained weights $\mathcal{W}$, and
get outputs for binary descriptor **b**, using (1) and (2).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2934477, IEEE Journal of Biomedical and Health Informatics

Fig. 4 gives a visual illustration of typical feature representations of the different dense blocks. It can be seen that these feature maps favor some specific structures responsive to optic disc geometrical characteristics, which is closely related to the CDR value considered in this work. By looking at the filters of the different block (Fig.4), one may notice that these feature maps becomes more efficient to obtain CDR relevant information such as OD/OC shapes, regions, border cues.

III. CDR Estimation using Regression Forests

In this section, CDR estimation is formulated as a regression problem which learns a general mapping to associate the given inputs with their corresponding continuous predictions. Here our aim is to associate the binary descriptor with a continuous label, i.e., the CDR value in fundus image. Formally, given a multi-variate input $\mathbf{b}$ we wish to associate a continuous label $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^n$. More generally, we wish to estimate the probability density function $p(\mathbf{y}|\mathbf{b})$. Actually, the input is represented as a bit string or multi-dimensional binary feature vector $\mathbf{b} = (b_1, b_2, ..., b_d) \subseteq \mathbb{R}^d$.

Random forests [29] combine the idea of bagging and randomization to aggregate the efficient prediction. Random forests can effectively deal with feature selection and point estimation, since it has been attracting increasing attention in medical image analysis. A regression forest is a collection of randomly trained regression trees (Fig.5) and a regression tree splits a complex nonlinear regression problem into a set of smaller problems which can be more easily handled by simpler models [30]. The binary decision tree is adopted in this work as the regression tree to predict the CDR value independently trained based on a set of labeled training data. Like most ensemble approaches, the forest output for the CDR estimation is the average of all tree outputs:

$$p(\mathbf{y}|\mathbf{b}) = \frac{1}{T}\sum_{t=1}^{T} p_t(\mathbf{y}|\mathbf{b}) \tag{7}$$

where T is the number of regression tree, and $p_t(\mathbf{y}|\mathbf{b})$ is the output of the t-th regression tree.

i) Training. In this work, we build the single regression tree using the binary decision tree. Each internal node is associated with a split function and its result decides which child node (left or right) the input data $\mathbf{b} \in \mathbf{B}$ is sent to. Forest training is achieved by optimizing an objective function defined over a training set $\mathbf{B}_0 \subset \mathbf{B}$ and associated CDR values. Each regression tree is constructed with a randomly selected training subset. Therefore, a split node j is optimized as

$$\theta_j = \arg\max_{\theta \in \mathcal{T}} I(\mathbf{B}_j, \theta) \tag{8}$$

where $\mathcal{T}$ represents the space of all split parameters, $\mathbf{B}_j \subset \mathbf{B}$ is a subset of training samples associated with the j-th node, and θ_j is the parameters of the j-th node which can be optimized by the minimizing a least-squares error function as shown in (5). Here, we employ the information gain associated with the j-th as the objective function:

$$I(\mathbf{B}_j, \theta) = \mathbf{H}(\mathbf{B}_j) - \sum_{i \in \{L,R\}} \frac{|B_j^i|}{|B_j|}\mathbf{B}_j^i \tag{9}$$

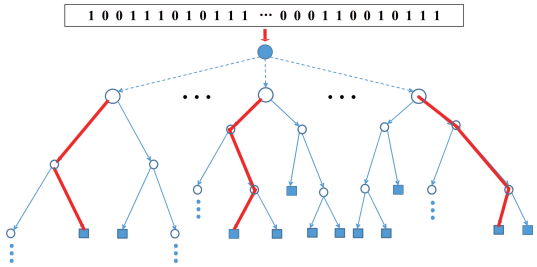


Figure 5. Random Regression Forest for CDR Estimation. The regressor is composed of multiple decision trees and trained with the labeled CDR values.

where $\mathbf{H}$ is the average entropy for the training subset $\mathbf{B}_j$.

ii) Prediction. Given a unseen binary descriptor of fundus image, the objective of our random forest model is to predict the CDR value of the fundus image. As shown in Fig.5, the feature descriptor $\mathbf{b}$ is passed through each regression tree and send to the left or right child node from the root of each tree T_i. The branch is decided by the split function, and the prediction result is made when the data reaches a leaf node of the tree. After the prediction of all the regression tree in the forest, the estimated CDR value is calculated by averaging results of all the regression trees as shown in (7).

iii) CDR estimation. Overall, the CDR estimation can be implemented based on the semi-supervised learning scheme which composed of two phases: unsupervised feature representation of fundus image and CDR value regression. The CNNs-based feature representation and random forest-based regressor are cascaded as a integral architecture for CDR value estimation. The scheme takes the fundus image as input and learns the robust representation with the trained MFPPNet, and then associates the binary descriptor with its corresponding CDR value based on random forest regressor. In this way, the semi-supervised learning scheme couples the CDR estimation with the representation of appearance and structure of OD/OC and forms an unified complete procedure. Therefore, our semi-supervised learning scheme delivers a task-driven process and thence ensures accurate estimation of CDR values.

IV. Experiments

In this section, experimental methods are detailed after the presentation of the key information about the test dataset and experimental configurations. In addition, a set of evaluation criteria used to evaluate the estimation accuracy between the estimated results and the ground truth are introduced.

A. Dataset

In our experiments, a set of (934 in total) fundus images from 934 different eyes and 443 clinical subjects was used to demonstrate the effectivity of our method. These subjects are collected from the second XiangYa hospital and Center for ophthalmic image analysis (Central South University). Among the 934 eyes, 280 are diagnosed as glaucomatous by ophthalmologists, accounted for 30%, which matches the actual results of epidemiological studies. There are only 421 images are given with manual labeled CDR value by trained ophthalmologists to train the regression forest for the CDR

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2934477, IEEE Journal of Biomedical and Health Informatics

estimation, while all the images are used to train the MFPPNet for a discriminative descriptor without any label information. For convenience, we name the dataset as Direct-CSU.

Meanwhile, the public datasets (ORIGA [31]), consisting 650 fundus images, is also used in our experiments to validate the performance of our proposed method. The ORIGA consists of 168 images from all glaucomatous eyes and 482 images from normal eyes. Manual CDRs are computed from the given manual labeled boundaries by experts in ophthalmology of the hospital. A two-fold cross validation strategy is adopted to obtain the results for all 650 images, where 325 images selected randomly are used to train our model.

B. Configurations

The configuration of the MFPPNet is same as 3-blocks DenseNet [27] just adding the feature pyramid pooling module after the DenseNet and replacing the classification layer with the fully-connected objectives layer defined in our paper. The training is implemented with the Pytorch deep learning framework, using four NVidia GeForce TITAN xp GPU on Ubuntu 16.04 Linux OS. Our network is initialized with the pre-trained weights from ImageNet. The stochastic gradient descent (SGD) and back-propagation are adopted to train the whole network. All models are trained for up to 90 epochs with early stopping criteria, where a model snapshot with low validation loss is taken for the final model. Other hyper-parameters are fixed as follows: α_1 = 1.0; α_2 = 0.8; and step learning rate schedule with base learning rate of 0.01, decreased by a factor of 10 every 30 epochs.

In our experiments, we set the number of the tree in random forest as 600, which is a key parameter determining the regression performance and computational burden. When random forest is used to CDR estimation, 10-fold validation approach is employed for evaluation and comparison of the performance of direct estimation in practice. During CDR estimation procedure, each dataset is divided into 10 groups. 9 groups are employed to train the regressor model, and the last group is used for test. This procedure is repeated 10 times.

C. Evaluation Criteria

In this work, CDR values obtained from manual segmentation by ophthalmologists are adopted as the gold standard (ground truth) for CDR estimation and glaucoma screening. We evaluate the proposed method on CDR estimation and glaucoma screening by the three criteria as follows:

i) The correlation with manual CDRs. The correlation coefficient is a number that quantifies the statistical relationships between two observed data values. In this paper, we compute the Pearson's correlation coefficient ρ which measures the linear correlation between the estimated CDRs and the manual CDRs. Here, the correlation coefficient ρ can be calculated by

$$\rho = \frac{\sum_{i=1}^{N}(CDR_i^E - \bar{CDR^E})(CDR_i^M - \bar{CDR^M})}{\sqrt{\sum_{i=1}^{N}(CDR_i^E - \bar{CDR^E})^2 \sum_{i=1}^{N}(CDR_i^M - \bar{CDR^M})^2}} \quad (10)$$

where CDR^E, CDR^M are the CDR values from the estimated and manual respectively, N is the number of each observed data, $\bar{CDR^E}$ and $\bar{CDR^M}$ are the means of the CDR values.

The correlation coefficient of "1" implies the total positive linear correlation existing between the two CDRs. Generally, we try to obtain a value closed to "1" for demonstrating the accuracy of the estimation algorithm.

ii) CDR error. CDR error represents the absolute deviation which measures difference between the estimated and manual CDRs. Here, it is computed as $\delta_i = |CDR_i^E - CDR_i^M|$ where CDR_i^E denotes the i-th manual CDR and CDR_i^M denotes the i-th estimated CDR value with our algorithm. Specially, the mean absolute error (MAE) between the estimated results and the ground truth is adopted, which is defined as

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|CDR_i^E - CDR_i^M| = \frac{1}{N}\sum_{i=1}^{N}\delta_i \quad (11)$$

iii) Glaucoma detection accuracy. The performance for glaucoma screening is evaluated based on the estimated CDR. We report the receiver operating characteristic (ROC) curve and area under the curve (AUC) as the overall measure of the screening strength.

D. Experiments

Extensive experiments are conducted to validate the effectiveness of our direct method from the following aspects.

Firstly, the performance of the proposed method for CDR estimation is test on our datasest (Direct-CSU) and public datasets (ORIGA). The correlation and CDR error are examined with the 10-fold cross validation protocol. Furthermore, the glaucoma detection accuracy is validated by the ROC curve and AUC when the estimated CDRs are utilized as the indictor to diagnose glaucoma. In addition, to demonstrate the capacity of feature representation of MFPPNet, the correlation coefficient is examined with different feature representation, such as hand-crafted, stacked convolutional network and neural network with dense blocks while the same CDR regressor. In our experiments, hand-crafted feature is the combination of pyramidal Gabor feature (PGF), histogram of oriented gradients and appearance, stacked network is a 16 layers convolutional neural network without dense blocks and feature pyramid pooling, and densely connected neural network without feature pyramid. All those neural networks are trained with our objective function.

Secondly, to evaluate the robustness of the proposed method, the tests under different settings are conducted. There is one significant parameters (the length of the feature vector outputted from MFPPNet) maybe affect the performance of our method. In our experiments, the proposed method is tested under different settings.

Thirdly, the comparison experiment is implemented to demonstrate the advantages of our method over existing segmentation-based and reconstruction-based methods. We test the state-of-the-art methods using both segmentation and reconstruction approaches. The segmentation-based method [19], [32]–[34] measures the CDR value after optic disc/cup segmentation, whereas the reconstruction-based method [9] computes the CDR after the segmentation and reconstruction of optic

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2934477, IEEE Journal of Biomedical and Health Informatics

disc image using the reference images and the regularization. In our experiments, we select the most competitive methods of optic disc/cup segmentation for CDR estimation, for example ASM [32], ACM [14], Threshold method [33], R-bends method [34], superpixel-based method [19], and CNN-based method [21][1] The experiments are conducted on the Direct-CSU and ORIGA datasets, respectively. As described above, the manual CDRs and diagnostic outcomes are compared on the three aspects: correlation coefficient, CDR error and glaucoma screening accuracy.

In addition, statistical significance improvement of the proposed method compared with existing segmentation-based and reconstruction-based methods is examined by one-tailed F-test with significance level of 1%. The test results H, p-values, variance ratio and its confidence level are computed to demonstrate the significance improvement of our proposed method. A test result of $H = 1$ indicates that the proposed method achieves significantly lower estimation error variance than its competitor. Here, the test statistic is variance ratio F, which is the ratio of two scaled sums of squares reflecting different sources of variability, formulated as $F = \frac{\sigma_1^2}{\sigma_2^2}$, where σ_1 and σ_2 are variances of the estimation errors for proposed method and one competitor to compare with.

V. Results and analysis

In this section, we demonstrate the performance of the proposed CDR estimation method by results and analyses of the above experiments on clinical fundus images. The estimated CDR values are quantitatively evaluated by the correlation coefficient and estimation error with the manually labeled CDR values. Assuredly, the glaucoma detection accuracy is clarified when the estimated CDR acts as the indictor for glaucoma detection. Furthermore, we demonstrate the advantages of our direct estimation method over the state-of-the-art segmentation-based methods.

A. Ablation Analysis of Direct Estimation

1) Estimated Results. The comparison of estimated CDR values with the manual labeled CDRs is depicted in Fig.8(f). The graph shows the correlations between direct estimated CDRs and manually obtained CDRs. Ideally, the distribution of CDR plots is desired as closed as possible to the red straight line as shown in Fig.8. Despite of the challenges in estimation of CDRs, the proposed method achieves a correlation coefficient of **0.7263** and the mean absolute errors of **0.0563**, which are outstanding more accurate than the segmentation-based and reconstruction-based methods.

The feature representation capacity of our DMFPPNet is compared with hand-crafted features, stacked network and densely connected network in Table 1. The results show that benefit from dense blocks and feature pyramid pooling, the hierarchical features of MFPPNet effectively represent the complex hidden patterns for the benefit of estimating CDR values. The visualization of different features is shown in Fig.4. As can be seen, MFPPNet captures the geometrical structures such as optic disc regions and borders, especially optic cup shape, which are closely related to CDR values estimation, while the hand-crafted filters such as PGF and HOG only represent partial edges of optic disc and vessel, which can not accurately describe shape variance of optic disc and cup.

Furthermore, our method achieves a AUC of **0.905** for glaucoma detection. The receiver operating characteristic (ROC) curve is depicted in Fig.6(a), which plots sensitivity against 1-specificity for all possible values. The results confirm feature representation with MFPPNet overcomes the geometrical complexity of the hidden patterns in fundus images, and validate the effectiveness of the estimated CDRs used for glaucoma screening. Significantly, the results demonstrate the effectiveness of our method for direct estimation of CDRs.

2) Performance Evaluation under Different Settings To evaluate the robustness of the proposed method, it is important to conduct the tests under different settings. In our method, the bit-length of the binary descriptor is important to feature representation of fundus image since it determines the capacity of the deep feature of our MFPPNet. In our experiments, we conduct a validation search in 64, 128, 512, 640, 868, 1024, 2048 and 4096 to evaluate the performance under different settings and we found that the bit-length does not have much effect on the performance of the CDR estimation when the bit-length is within a certain range.

In Fig.7, we utilize the statistical box-and-whisker diagrams to show the performance variations with the different bit-length. Each whisker represents the range of CDR values and its errors with all the test images for each bit-length, the bottom end of each whisker represents the optimal performance, and the horizontal red line shows the median value of the performance. From the plots we can see that the variety of the whisker's bottom ends are obvious with the increasing of the bit-length before 1024, which reveals bit-length can reduce estimation errors by increasing representation power for CDR-related geometrical information of fundus. According to the results, we find outstanding performance improvement as the hit-length increases between 64 and 1024, while slight improvement after 1024. In principle, from the architecture of the MFPPNet we can find that the representation performance is confined by the channel number of the outputs of the last convolutional layer. Therefore, the bit-length of the binary descriptor in the following experiments is adopted as 1024.

B. Comparison with Segmentation-based Methods

The proposed method reveals great advantages for CDR estimation and glaucoma screening, when being compared with the most competitive state-of-the-art optic disc/cup segmentation-based and the reconstruction-based method.

The average CDR error (MAE), correlation coefficient, and AUC are given in Table II. The results show that the proposed method achieves more accurate CDR values with lower CDR error, while obtains higher correlation coefficient and AUC for glaucoma screening than the other methods.

[1]The ASM [32] and ACM [14] methods are from the public projects, and the CNN-based method [21] are from the original author, while the rest [9], [19], [33], [34] are implemented by ourselves strictly following the original papers. The default parameters used in the original paper are adopted in our experiment.

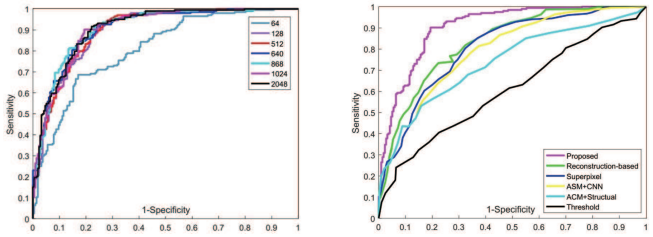
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2934477, IEEE Journal of Biomedical and Health Informatics

Table I
THE PERFORMANCE COMPARISON BETWEEN MANUALLY OBTAINED AND ESTIMATED CDRS WITH DIFFERENT FEATURES.

Dataset	Feature	Dense Block	FPP	MAE	Correlation coefficients ρ	AUC
Direct-CSU	Hand-crafted	×	×	0.1123	0.6418	0.86
	Stacked network	×	×	0.0652	0.6531	0.83
	Densely connected	✓	×	0.0608	0.6912	0.87
	MFPPNet	✓	✓	**0.0563**	**0.7263**	**0.90**
ORIGA	Hand-crafted	×	×	0.1012	0.6312	0.85
	Stacked network	×	×	0.0645	0.6401	0.81
	Densely connected	✓	×	0.0628	0.6621	0.85
	MFPPNet	✓	✓	**0.0606**	**0.6842**	**0.88**

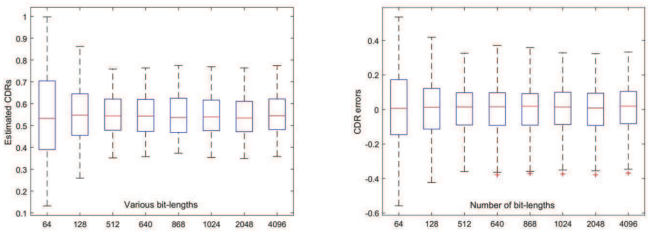
Table II
PERFORMANCE COMPARISON WITH SEGMENTATION-BASED AND RECONSTRUCTION-BASED METHODS.

Method	Direct-CSU			ORIGA		
	MAE	Correlation coefficients ρ	AUC	MAE	Correlation coefficients ρ	AUC
ASM+CNN [21], [32]	0.0956	0.58	0.72	0.0970	0.56	0.70
ACM+Structual [14]	0.1435	0.53	0.69	-	-	-
Threshold [33]	0.1936	0.38	0.60	-	-	-
R_bends [34]	0.1592	0.40	0.68	0.1271	0.38	-
Superpixel [19]	0.0827	0.62	0.78	0.0780	0.59	0.80
Reconstruction-based [9]	0.0651	0.64	0.84	0.0640	0.67	0.83
Proposed	**0.0563**	**0.72**	**0.90**	**0.0606**	**0.68**	**0.88**



(a) ROC with various bit-length (b) ROC by various methods

Figure 6. Compared with existing state-of-the-art methods, the proposed method obtains more robust ROC curve and higher AUC value when the estimated CDRs values used as indicator for glaucoma screening.



(a) estimated CDRs (b) CDR errors

Figure 7. Box-and-whisker plots of the CDR estimation performance of our method with different bit-lengths. The variety of bit-length has little effect on the estimated CDRs when the bit-length is greater than 1024.

Compared with the classical segmentation-based methods, the proposed method reduces the MAE by 41.11% and 31.9% over ASM+CNN [21], [32] and Superpixel [19], respectively. In term of the Pearson correlation, the relative improvement is 24.14% and 16.13%, respectively. Our further analysis reveals

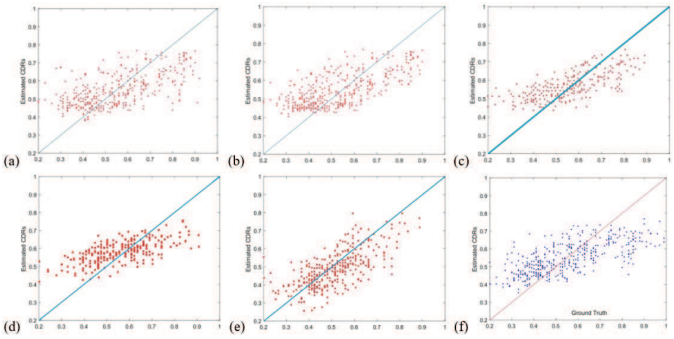


Figure 8. CDR plots by various methods using manual CDRs as ground truth (Direct-CSU dataset as testing). Notice that the distribution of CDR plots from our method is close to the red straight line, which means the estimated CDRs are vary correlated with its ground truth. The proposed method delivers lower estimation error and more compact distribution than other methods. (a)Threshold [33]. (b)ACM+Structual [14]. (c) ASM+CNN [21], [32]. (d) Superpixel [19]. (e) Reconstruction-based [9]. (f) The proposed.

that the accuracy of CDR estimation is heavily sensitive to the segmentation performance in traditional methods, and the dependency on precision location of boundary pixels makes the CDR error of segmentation-based methods increase. Although lots of efforts have been devoted into segmentation of optic disc/cup, it still fails to deliver precision location of optic cup boundary pixels in the inferior and superior regions, especially when disturbance exist. Benefit of direct estimation and deep representation of fundus image, the proposed method skips the intermediate procedure of optic disc/cup segmentation, eliminate the dependency on segmentation results in the inferior and superior regions, and couples the estimation procedure with the representation of complex hidden patterns of optic disc. Compared with the reconstruction-based method [9], the proposed method reduces the MAE by 13.5% and improves the Pearson correlation coefficient by 12.5%. This evidences that

Table III
LEFT-TAILED F-TEST BETWEEN THE PROPOSED METHOD AND EACH OF THE COMPETITOR. OUR PROPOSED METHOD SIGNIFICANTLY OUTPERFORMS ITS COMPETITORS.

Method	H	p-value	CI	$F = \frac{\sigma_1^2}{\sigma_2^2}$
ASM+CNN [21], [32]	1	< 0.01	[0, 0.810]	0.781
ACM+Structual [14]	1	< 0.01	[0, 0.664]	0.610
Threshold [33]	1	< 0.01	[0, 0.432]	0.389
R_bends [34]	1	< 0.01	[0, 0.507]	0.459
Superpixel [19]	1	< 0.01	[0, 0.893]	0.788
Reconstruction-based [9]	1	< 0.01	[0, 0.933]	0.829
Proposed	1	< 0.01	**[0, 0.961]**	**0.854**

the direct estimation through deep representation and indices regressor is a best way to achieve accurate estimation of CDR value. Fig.8 shows scatter plots of the correspondence between manual CDRs and the estimated CDRs with various methods. As can be seen, the plots from the proposed method are evenly compact distributed on both sides of the line with correlation coefficient 1. It reveal that the proposed method can deliver more accurate estimation than its competitors.

In terms of glaucoma screening and diagnosis, the improvement in AUC is 25% and 15.38% compared with ASM+CNN [21], [32] and Superpixel [19], while the improvement (7.14%) is relatively smaller when compared with the reconstruction-based method [9]. It is likely that the screening is more related with the discriminative pattern and geometrical structure changes between glaucomatous and the healthy subjects. Compared with the segmentation-based methods, MFPPNet captures the discriminative features such as geometrical structure, optic disc region and intensity, especially optic cup shape, which are not only closely related to CDR values estimation, but also reveals the changes of optic disc and cup caused by glaucoma. It justifies the benefits of using the discriminative features for CDR value regression in our direct method.

Table III demonstrates the results of left-tailed F-test for our method and other competitors for CDRs on . The test result H, p-value, variance ratio and corresponding confidence interval are demonstrated. In Table III, the variance ratio and confidence interval reveal the extent of different between the existing methods and the proposed method. The test results demonstrate that the proposed method significantly outperforms these competitors.

VI. CONCLUSIONS

A effective method was proposed to estimate CDR value directly from fundus images while skips the intermediate procedure of segmentation used in existing approaches. The semi-supervised learning enhances the expressiveness of image structures and random regression forest builds bridges between image features and corresponding CDRs. When validated on a challenging glaucoma dataset, the proposed direct method achieved accurate estimation with high correlations with manually obtained ones, and it also estimated the CDR value more effectively than results obtained from traditional segmentation-based methods. According to what we have learnt, it is the first time that CDR value is estimated directly from fundus image without segmentation.

The proposed method introduces semi-supervised learning into estimation of glaucoma risk factors to handle the lack of labeled data. The success of the proposed method provides a modified way for low-cost glaucoma screening in polyclinics.

REFERENCES

[1] N. Thakur and M. Juneja, "Survey on segmentation and classification approaches of optic cup and optic disc for diagnosis of glaucoma," *Biomedical Signal Processing and Control*, vol. 42, pp. 162–189, 2018.

[2] J. Guo, G. Azzopardi, C. Shi, N. M. Jansonius, and N. Petkov, "Automatic determination of vertical cup-to-disc ratio in retinal fundus images for glaucoma screening," *IEEE Access*, vol. 7, pp. 8527–8541, 2019.

[3] P. S. Mittapalli and G. B. Kande, "Segmentation of optic disk and optic cup from digital fundus images for the assessment of glaucoma," *Biomedical Signal Processing and Control*, vol. 24, pp. 34–46, 2016.

[4] V. G. Edupuganti, A. Chawla, and A. Kale, "Automatic optic disk and cup segmentation of fundus images using deep learning," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 2227–2231, IEEE, 2018.

[5] F. Bokhari, T. Syedia, M. Sharif, M. Yasmin, and S. L. Fernandes, "Fundus image segmentation and feature extraction for the detection of glaucoma: A new approach," *Current Medical Imaging Reviews*, vol. 14, no. 1, pp. 77–87, 2018.

[6] M. S. Haleem, L. Han, J. van Hemert, B. Li, A. Fleming, L. R. Pasquale, and B. J. Song, "A novel adaptive deformable model for automated optic disc and cup segmentation to aid glaucoma diagnosis," *Journal of medical systems*, vol. 42, no. 1, p. 20, 2018.

[7] R. Bock, J. Meier, L. G. Nyúl, J. Hornegger, and G. Michelson, "Glaucoma risk index: automated glaucoma detection from color fundus images," *Medical Image Analysis*, vol. 14, no. 3, pp. 471–481, 2010.

[8] U. R. Acharya, S. Dua, X. Du, C. K. Chua, *et al.*, "Automated diagnosis of glaucoma using texture and higher order spectra features," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 3, pp. 449–455, 2011.

[9] J. Cheng, F. Yin, D. W. K. Wong, D. Tao, and J. Liu, "Sparse dissimilarity-constrained coding for glaucoma screening," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 5, pp. 1395–1403, 2015.

[10] X. Chen, Y. Xu, S. Yan, D. W. K. Wong, T. Y. Wong, and J. Liu, "Automatic feature learning for glaucoma detection based on deep learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 669–677, Springer, 2015.

[11] D. Yadav, M. P. Sarathi, and M. K. Dutta, "Classification of glaucoma based on texture features using neural networks," in *Contemporary Computing (IC3), 2014 Seventh International Conference on*, pp. 109–112, IEEE, 2014.

[12] J. Xu, O. Chutatape, and P. Chew, "Automated optic disk boundary detection by modified active contour model," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 3, pp. 473–482, 2007.

[13] C. Jun, L. Jiang, W. Damon Wing, Kee, and Y. Wong, "Automatic optic disc segmentation with peripapillary atrophy elimination," in *Engineering in Medicine and Biology Society, Embc, 2011 Annual International Conference of the IEEE*, pp. 6224–6227, 2011.

[14] M. C. V. S. Mary, E. B. Rajsingh, J. K. K. Jacob, D. Anandhi, U. Amato, and S. E. Selvan, "An empirical study on optic disc segmentation using an active contour model," *Biomedical Signal Processing and Control*, vol. 18, pp. 19–29, 2015.

[15] S. Zhao, Z. Gao, H. Zhang, Y. Xie, J. Luo, D. Ghista, Z. Wei, X. Bi, H. Xiong, C. Xu, *et al.*, "Robust segmentation of intima–media borders with different morphologies and dynamics during the cardiac cycle," *IEEE journal of biomedical and health informatics*, vol. 22, no. 5, pp. 1571–1582, 2017.

[16] G. D. Joshi, J. Sivaswamy, and S. R. Krishnadas, "Depth discontinuity-based cup segmentation from multiview color retinal images," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1523–1531, 2012.

[17] M. S. Miri, M. D. Abrmoff, K. Lee, M. Niemeijer, J. K. Wang, Y. H. Kwon, and M. K. Garvin, "Multimodal segmentation of optic disc and cup from sd-oct and color fundus photographs using a machine-learning graph-based approach," *IEEE Transactions on Medical Imaging*, vol. 34, no. 9, pp. 1854–1866, 2015.

[18] Y. Xu, L. Duan, S. Lin, X. Chen, D. W. K. Wong, T. Y. Wong, and J. Liu, "Optic cup segmentation for glaucoma detection using low-rank superpixel representation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 788–795, Springer, 2014.

[19] J. Cheng, J. Liu, Y. Xu, F. Yin, D. W. K. Wong, N.-M. Tan, D. Tao, C.-Y. Cheng, T. Aung, and T. Y. Wong, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1019–1032, 2013.

[20] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, "Deep retinal image understanding," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 140–148, Springer, 2016.

[21] Y. Guo, B. Zou, Z. Chen, Q. He, Q. Liu, and R. Zhao, "Optic cup segmentation using large pixel patch based cnns," in *Lecture Notes in Computer Science, the Opthalmic Medical Image Analysis Third Internation Workshop (OMIA2016)*, pp. 129–136, Springer, 2016.

[22] S. Zhao, X. Wu, B. Chen, and S. Li, "Automatic spondylolisthesis grading from mris across modalities using faster adversarial recognition network," *Medical Image Analysis*, p. 101533, 2019.

[23] Z. Wang, M. B. Salah, B. Gu, A. Islam, A. Goela, and S. Li, "Direct estimation of cardiac biventricular volumes with an adapted bayesian formulation," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 4, pp. 1251–1260, 2014.

[24] X. Zhen, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Direct and simultaneous four-chamber volume estimation by multi-output regression," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 669–676, Springer, 2015.

[25] W. Xue, A. Islam, M. Bhaduri, and S. Li, "Direct multitype cardiac indices estimation via joint representation and regression learning," *IEEE Transactions on Medical Imaging*, 2017.

[26] W. Xue, G. Brahm, S. Pandey, S. Leung, and S. Li, "Full left ventricle quantification via deep multitask relationships learning," *Medical image analysis*, vol. 43, pp. 54–65, 2018.

[27] K. Lin, J. Lu, C. S. Chen, and J. Zhou, "Learning compact binary descriptors with unsupervised deep neural networks," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pp. 1183–1192, 2016.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[29] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[30] C. Antonio and J. Shotton, *Decision Forests for Computer Vision and Medical Image Analysis*. Springer London, 2013.

[31] D. Tao, D. W. K. Wong, J. Liu, J. Cheng, M. Baskaran, T. Y. Wong, T. Aung, and Z. Zhang, "Similarity regularized sparse group lasso for cup to disc ratio computation," *Biomedical Optics Express*, vol. 8, no. 8, pp. 3763–3777, 2017.

[32] F. Yin, J. Liu, D. W. K. Wong, N. M. Tan, C. Cheung, M. Baskaran, T. Aung, and T. Y. Wong, "Automated segmentation of optic disc and optic cup in fundus images for glaucoma diagnosis," in *Computer-based medical systems (CBMS), 2012 25th international symposium on*, pp. 1–6, IEEE, 2012.

[33] G. D. Joshi, J. Sivaswamy, K. Karan, and S. Krishnadas, "Optic disk and cup boundary detection using regional information," in *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pp. 948–951, IEEE, 2010.

[34] G. D. Joshi, J. Sivaswamy, and S. Krishnadas, "Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment," *IEEE transactions on medical imaging*, vol. 30, no. 6, pp. 1192–1205, 2011.