Dual Consistency Enabled Weakly and Semi-Supervised Optic Disc and Cup Segmentation With Dual Adaptive Graph Convolutional Networks

Yanda Meng, Hongrun Zhang, Yitian Zhao, Dongxu Gao, Barbra Hamill, Godhuli Patri, Tunde Peto, Savita Madhusudhan, and Yalin Zheng

***Abstract*— Glaucoma is a progressive eye disease that results in permanent vision loss, and the vertical cup to disc ratio (*vCDR*) in colour fundus images is essential in glaucoma screening and assessment. Previous fully supervised convolution neural networks segment the optic disc (*OD*) and optic cup (*OC*) from color fundus images and then calculate the *vCDR* offline. However, they rely on a large set of labeled masks for training, which is expensive and time-consuming to acquire. To address this, we propose a weakly and semi-supervised graph-based network that investigates geometric associations and domain knowledge between segmentation probability maps (*PM*), modified signed distance function representations (*mSDF*), and boundary region of interest characteristics (*B-ROI*) in three aspects. Firstly, we propose a novel Dual Adaptive Graph Convolutional Network (*DAGCN*) to reason the long-range features of the *PM* and the *mSDF* *w.r.t.* the regional uniformity. Secondly, we propose a dual consistency regularization-based semi-supervised learning paradigm. The regional consistency between the *PM* and the *mSDF*, and the marginal consistency between the derived *B-ROI* from each of them boost the proposed model's performance due to the inherent geometric associations. Thirdly, we exploit the task-specific domain knowledge via the oval shapes of *OD* & *OC*, where a differentiable *vCDR* estimating layer is proposed. Furthermore, without additional annotations, the supervision on *vCDR* serves as weakly-supervisions for segmentation tasks. Experiments on six large-scale datasets demonstrate our model's superior performance on *OD* & *OC* segmentation and *vCDR* estimation. The implementation code has been made available.https://github.com/smallmax00/Dual_Adaptive_Graph_Reasoning**

***Index Terms*— Weakly and semi-supervised learning, graph convolutional network, optic disc and cup segmentation.**

I. Introduction

GLAUCOMATOUS damage to the optic nerve head can be assessed on colour fundus images, by measuring the relative size of the optic disc (*OD*) and the optic cup (*OC*) in the vertical direction of the image [1]. Traditionally, a widely adopted method is to calculate the vertical cup to disc ratio (*vCDR*) [2]. Few of the current methods directly regresses the *vCDR* values from fundus images [3]. However, it has lead to the difficulty and uninterpretability in learning [1]. A common pipeline is to segment *OD* and *OC* regions respectively, after which the *vCDR* is calculated as the ratio between the vertical cup diameter and vertical disc diameter. Consequently, accurate segmentation of *OD* & *OC* is critical for the *vCDR* measurement, in turn for the glaucoma assessment. Recently, numerous deep learning-based segmentation models [1], [2], [4], [5], [6], [7], [8] have been proposed, significantly improving the *OD* & *OC* segmentation accuracy. However, most of them use a fully supervised paradigm, where a large number of manual delineation labels by clinicians or trained experts are required as the ground truth prior to training the model. The manual annotations are also hugely subjective, time-consuming, laborious, and costly. Solving this problem depends on automated and precise segmentation algorithms that can exploit a large number of unlabeled images without the need for manual delineations. To this end, we proposed a newly designed weakly/semi-supervised learning mechanism that is integrated with our proposed Dual Adaptive Graph Convolutional Network (*DAGCN*). With the critical novelty of

Manuscript received 23 July 2022; accepted 23 August 2022. Date of publication 31 August 2022; date of current version 2 February 2023. *(Corresponding authors: Yitian Zhao; Yalin Zheng.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the North West Multi-centre Research Ethics Committee of UK Biobank under Approval No. 11/NW/0382.

Yanda Meng, Hongrun Zhang, and Yalin Zheng are with the Department of Eye and Vision Science, University of Liverpool, Liverpool L7 8TX, U.K. (e-mail: yanda.meng@liverpool.ac.uk; hongrun.zhang@liverpool.ac.uk; yalin.zheng@liverpool.ac.uk).

Yitian Zhao is with The Affiliated People's Hospital of Ningbo University, Ningbo 315041, China, and also with the Cixi Institute of Biomedical Engineering and the Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Science, Ningbo 315201, China (e-mail: yitian.zhao@nimte.ac.cn).

Dongxu Gao is with the School of Computing, University of Portsmouth, Portsmouth PO1 3HE, U.K. (e-mail: dongxu.gao@port.ac.uk).

Barbra Hamill and Tunde Peto are with the School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast BT9 7BL, U.K. (e-mail: b.hamill@qub.ac.uk; t.peto@qub.ac.uk).

Godhuli Patri and Savita Madhusudhan are with the St Paul's Eye Unit, Liverpool University Hospitals NHS Foundation Trust, Liverpool L7 8XP, U.K. (e-mail: godhuli.patri@liverpoolft.nhs.uk; savita.madhusudhan@liverpoolft.nhs.uk).

Digital Object Identifier 10.1109/TMI.2022.3203318

exploiting the geometric associations and domain knowledge, we have demonstrated the framework's effectiveness for the segmentation of *OD & OC* and also glaucoma assessment *w.r.t. vCDR* estimation in colour fundus images.

The previous segmentation methods concentrated on learning the intensity features of the input images; they would normally rely on a single task such as dense probability map classification, boundary localization, or signed distance function regression. Despite human graders' instinctive use of both image intensity features and spatial relationships between object's boundary and region, they ignore the inherent geometric association between these learned representations, which are critical for improving segmentation performance [7], [9]. To be more precise, segmentation probability map (*PM*) features emphasize on the global homogeneity of pixel-level semantics and contextual information at the object level. The local boundary characteristics, such as boundary region of interest (*B-ROI*), describes the spatial variations on both sides of the boundary contour. The signed distance function (*SDF*) representations emphasize on the global geometry-aware signed distance *w.r.t.* the object contours. Notably, in this work, we propose a modified signed distance function (*mSDF*) that has similar attributes to the SDF but indicates more coherent signals at the semantic level akin to *PM*. More specifically, the sign label is reversed from the *SDF* to the proposed *mSDF* (*e.g.* +, -) for the inner and outer regions of objects in order to make the learned *mSDF* features need to be coherent with the *PM* features for the construction of the dual graph adjacency matrix. Intuitively, the geometric associations between them appears to complement one another during model learning, such as regional and marginal consistency via spatial area and boundary uniformity, thereby improving segmentation performance. To accomplish this, we propose a semi-supervised learning paradigm to construct dual consistency regularizations on both object's region and boundary via the three aforementioned tasks. Additionally, we investigated the method to accompany the feature complementing rationally between *PM* segmentation and *mSDF* regression tasks at semantic and spatial levels. For example, the proposed novel *DAGCN* leverages the advantage of the graph-based model's long-range information propagation and cross-domain feature update capabilities. Specifically, we adaptively constructed the dual graph via initializing the adjacency matrix in a data-dependent way. The estimated vertex embeddings of *mSDF* and *PM* contributed to the dual adjacency matrices adaptively according to the geometric associations between them. We implemented two matrices to quantify the distance and relationship among different vertices so as to achieve adaptive graph construction and reasoning. On the other hand, previous *OD & OC* segmentation-based glaucoma assessment methods have chased high segmentation accuracy but have overlooked the fact that the ultimate goal of such a learning pipeline is to estimate the *vCDR* in order to aid in glaucoma assessment. As a result, the underlying weak supervision label of *vCDR* in *OD & OC* segmentation task is understudied. The previous methods adopted an offline post-processing step to calculate the *vCDR* given the estimated diameters of the *OD & OC*. On the contrary, we have exploited the domain-specific knowledge between the boundary and region in terms of the perimeter and area of an oval shape of *OD & OC*, where a new differentiable *vCDR* estimating layer is proposed for the end-to-end training. Thus, our model does not only avoid any offline post-process to generate *vCDR* but also gains more weakly-supervised guidance without further annotations. Such a novel design ensured that the proposed model learns the well-defined goals and gains more supervision from the ground truth on both the regions and boundaries of objects. The overview pipeline of our work is depicted in Fig. 1, please refer to Fig. 2 for more details. In summary, this work makes the following contributions:

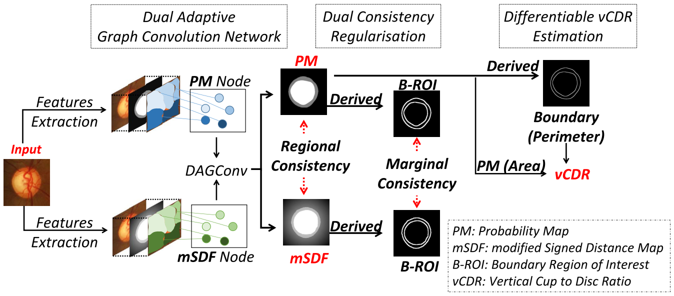


Fig. 1. Overview of the proposed network, where three major contributions, *DAGCN*, dual consistency regularization and differential *vCDR* estimation, are shown.

- We proposed a dual adaptive graph convolutional network (*DAGCN*) to reason the cross-domain segmentation probability maps and modified signed distance function representations. The information propagation and message exchange *w.r.t.* geometric associations and semantic context were exploited to learn a comprehensive graph representation and adaptive structure.
- We proposed a dual consistency-based paradigm on region and boundary geometric associations in a semi-supervised manner. The enforced consistency on regional and marginal features leads the learned model to a generalizable characteristic learning via leveraging a large amount of unlabeled data.
- For the first time, we exploited the task-specific domain knowledge in terms of perimeter and area of the oval-shaped *OD & OC*, and proposed to estimate the *vCDR* in a differentiable way. Thus, without any further laborious annotations, the supervision on *vCDR* serves as weakly-supervised guidance on the accurate *OD & OC* region and boundary segmentation.

II. Related Works

A. Pixel-Wise Medical Image Segmentation

Convolution Neural Network (*CNN*) has found widespread use in the segmentation of medical images. Existing CNN-based methods [2], [10], [11] have considered segmentation as a dense pixel classification task. For example, the classic *U-net* [10] employs a skip-connection between the encoder and decoder to minimize information loss. In recent years, it has been used as a baseline model for medical image segmentation tasks. Recently, Gu *et al.* [11] proposed to capture high-level information while preserving spatial information on *OD & OC* segmentation task. However, due to the limited receptive field of standard *CNN*, dense atrous convolutions were incorporated [12] to enlarge the receptive regions for long-range context reasoning. Similarly, *M-Net* [2] requires multi-scale input and side-output mechanisms with deep supervision, to achieve multi-level receptive field fusion for aggregating long-range relationships. With the assistance of the enhanced long-range reasoning abilities, the aforementioned methods achieved promising results in the *OD & OC* segmentation task. They are however inefficient as the stacking of local cues as it does not always accurately represent long-range context relationships [7]. On the contrary, we benefit from the long-range information aggregating ability of the graph-based models to address this issue.

B. Geometry-Aware Medical Image Segmentation

It is well established that boundary knowledge is essential in acquiring geometric features in segmentation tasks. When it comes to medical image segmentation, the boundary accuracy is often more critical than that of the regional pixel-wise coverage [5], [8]. Recent methods, such as [5], [6], [7], [8], [13], and [14], have explicitly or implicitly taken into account the geometry dependency between the regions and boundaries of an object of interest in *OD & OC*. Specifically, Meng *et al.* proposed an aggregated hybrid network [7] to jointly learn the relationship between region and boundary of *OD & OC*, conducting an accurate boundary localization. On the other hand, Luo *et al.* [13] and Xue *et al.* [14] adopted *SDF* to represent the target mask in segmentation tasks as it enables the network to learn a distance-aware representation *w.r.t* the object boundary, emphasizing the spatial perception of the input images. Similarly, we proposed to learn a *mSDF* regression task in this work to exploit the geometry-aware feature learning. Also, it is integrated into the proposed dual consistency semi-supervised paradigm at the task level, leading to a coherent semantic and spatial information integration with *PM* segmentation task in the proposed graph-based model.

Other boundary-based methods [4], [9] integrate the region and boundary geometry constraint into the loss function or evaluation measurement. For example, *Cheng et al. proposed a Boundary Intersection-over-Union (BIoU) [9] evaluation measurement, which quantifies boundary quality in segmentation tasks. Wu et al. [4]* proposed an oval shape constraint-based loss function to regularize the contour shape of the predicted *OD & OC* during learning. Similarly, we exploited the boundary and region relationship in terms of perimeter and area of oval shape to estimate the *vCDR* in a differentiable way. The underlying geometry association of the oval shape of *OD & OC* was researched and specially designed in this work.

C. Weakly and Semi-Supervised Medical Image Segmentation

By learning directly from a small set of labeled data and a large set of unlabeled data, the semi-supervised learning frameworks [13], [15], [16] achieved high-quality segmentation results. Numerous semi-supervised methods [17], [18] have recently been developed that incorporate unlabeled data through unsupervised consistency regularization. In general, there are majorly two different types of unsupervised consistency regularizations, *i.e.* a data-level of perturbations [17], [18], [19] and a feature-level of perturbations [15], [16]. However, on the other hand, the consistency regularization of task-level in semi-supervised learning has rarely been explored, until very recently in different computer vision tasks, such as crowd counting [20], 3D object detection [21], and 3D medical image segmentation [13]. To be more precise, various levels of information from different task branches can complement one another during training, whereas divergent focuses can lead to inherent prediction perturbation [22]. For example, [13], [20] and [21] all shared a similar idea that the dual task's outputs can be aligned into the same presentation space, and then an unsupervised loss is applied to regularize the consistency. In this work, we have also demonstrated a dual-task level of geometric consistency on the *OD & OC* segmentation. Apart from that, we have integrated the boundary quality into the task-level of consistency regularization.

On the other hand, weakly supervised methods [23], [24], [25], [26] segmented images using image-level of labels [24], bounding boxes [23], points [25], scribbles [26] rather than pixel-by-pixel annotation, which alleviated the burden of annotation. They all focused on the data-driven learning-based way of general coarse labels. For example, given the image-level labels, Wu *et al.* [24] proposed an attention mechanism on the top of the class activation maps [27] to improve 3D brain lesion localization. The estimated lesion regions and normal tissues were then used to train the 3D brain lesion segmentation network. Differently, for the first time, we integrated the task-specific domain knowledge into the proposed weakly supervised paradigm, where the oval shape of the *OD & OC* is exploited in the segmentation task. As a result, our model could estimate the *vCDR* end-to-end on the basis of *OD & OC* segmentation. At the same time, the information gained from *vCDR* ground truth could weakly-supervise the segmentation process for the both region and boundary of *OD & OC*.

D. Graph Reasoning in Segmentation

In the recent years, the graph-based models [7], [8], [28], [29], [30] have gained popularity for the segmentation tasks due to their inherent ability to propagate information over long distances and update feature information. *Meng et al. proposed RBA-Net [5]* and *CABNet* [6] to regress the *OD & OC* boundaries by aggregated *CNN* and Graph Convolutional Network (*GCN*), which learns the long-range features and directly regresses vertex coordinates in a Cartesian system. The methods described above made use of a Graph Neural Network (*GNN*) to address the challenge of intra-domain long-range feature propagation because messages passing between graph nodes have semantic and spatial characteristics that are similar to one another. Contrary to this, our method treats extracted pixel-level *PM* features and geometry-aware *mSDF* representations as distinct graph nodes and employed *GNN* to learn their inter-domain relationship. In particular, the geometric associations between them were exploited.

Additionally, methods such as [5], [28], [29], [30], and [6] used *Laplacian* smoothing-based graph convolution [31], provide specific benefits in the sense of global long-range information reasoning. They estimated the initial graph structure from a data-independent *Laplacian* matrix defined by randomly initialized adjacency matrix [29], [30] or hand-crafted adjacency matrix [5], [6], [28], [31]. However, one may enable a model to learn a specific long-range context pattern [8], [32], which is less related to the input features, and thus we considered them as a data-independent non-adaptive graph convolution. Differently, as seen in previous works that the graph structure could be estimated with the similarity matrix from the input data [32], we estimated the initial adjacency matrix in a data-dependent way. The constructed dual graph in this work had two distinct structures, which were adaptively learned from the input features of *PM* and *mSDF* features. Hence, our model was capable of adaptively learning an input-related long-range context pattern, which improved the model segmentation performance; please read *Ablation Study* (Section V-A) for more details.

III. Methods

A. Dual Adaptive Graph Convolutional Network

1) Graph Node Initialization: A backbone network was used to extract the multi-level features. The deep- and shallow-layer features from different levels complemented one another. For example, the deep-layer features contained extensive semantic region information, while the shallow-layer features retained sufficient spatial boundary information. Thus, for initializing the dual graph vertices, we used the feature aggregation module that is similar to [8] on relative deep-level and low-level features. Specifically, the backbone feature maps of 16 × 16, 32 × 32, and 64 × 64 were aggregated with 1 × 1, 3 × 3 convolutions and bilinear up-sampling operations. Reader are referred to Feature Aggregation Module (*FAM*) in [8] for more details. As a result, following the feature aggregation module, the output feature maps for *PM* (R_{pm}) and *mSDF* (R_{mSDF}) have the same sizes of *64 × 64 × 2*. We then referred them to as the initialised *PM* node embeddings and *mSDF* node embeddings, respectively.

2) Classic Graph Convolution: We first revisited the classic graph convolution and their graph construction process *w.r.t* the adjacency matrix. Given a graph *G = (V, E)*, normalised *Laplacian* matrix is defined as $L = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$, where I is the identity matrix, A is the adjacency matrix, and D is a diagonal matrix that represents each vertex's degree in V, such that $D_{ii} = \sum_j A_{i,j}$. The *Laplacian* of the graph is a positive semi-definite symmetric matrix, so L can be diagonalized by the Fourier basis $U \in \mathbb{R}^{N \times N}$, such that $L = U \Lambda U^T$. Thus, the spectral graph convolution of i and j can be defined as $i * j = U((U^T i) \odot (U^T j))$ in the Fourier space. The columns of U are the orthogonal eigenvectors $U = [u_1, \ldots, u_n]$, and $\Lambda = diag([\lambda_1, \ldots, \lambda_n]) \in \mathbb{R}^{N \times N}$ is a diagonal matrix with eigenvalues that are not negative. Due to the fact that U is not a sparse matrix, this operation is computationally inefficient. To solve this, it was proposed that the convolution operation on a graph can be defined by formulating spectral filtering [33] with a kernel g_θ using a recursive Chebyshev polynomial in Fourier space. The filter g_θ is parameterized in terms of an order K Chebyshev polynomial expansion, such that $g_\theta(L) = \sum_k \theta_k T_k(\hat{L})$, where $\theta \in \mathbb{R}^K$ is a vector of Chebyshev coefficients, and $\hat{L} = 2L/\lambda_{max} - I_N$ represents the rescaled *Laplacian*. $T_k \in \mathbb{R}^{N \times N}$ is the Chebyshev polynomial of order K. In Kipf and Welling [31], further simplified the graph convolution as $g_\theta = \theta(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}})$, where $\hat{A} = A + I$, $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$, and θ is the only Chebyshev coefficient left. The corresponding graph *Laplacian* adjacency matrix $\hat{A}$ is hand-crafted, which leads the model to learn a specific long-range context pattern rather than the input-related one [32]. As a result, we refered to the classic graph convolution as data-independent non-adaptive graph convolution.

3) Dual Adaptive Graph Convolution: This section adopts the similar graph structure *w.r.t* adjacency matrix from our previous works [8]. We extended it into a dual adaptive graph, perfectly fitting the proposed semi-supervised paradigm with dual consistency regularization. Given the initialized *PM* nodes $R_{pm} \in \mathbb{R}^{N \times C}$ and *mSDF* nodes $R_{mSDF} \in \mathbb{R}^{N \times C}$, we constructed the input-dependent adaptive adjacency matrix for the dual adaptive graph (G_{pm} and G_{mSDF}), where C is the channel size; $N = H \times W$ is the number of spatial locations of input feature, which is referred to as the number of vertices.

We illustrate G_{pm} as an example and elaborate the graph construction process as below. Firstly, we implemented two matrices ($\tilde{\Lambda}^c$ and $\tilde{\Lambda}^s$) to perform channel-wise attention on the dot-product distance between input vertex embeddings and to quantify spatially weighted relations between different vertices, respectively. For example, $\tilde{\Lambda}^c(R_{pm}) \in \mathbb{R}^{C \times C}$ is the matrix containing channel-specific information about the dot-product distance of the input vertex embeddings.; $\tilde{\Lambda}^s(R_{pm}) \in \mathbb{R}^{N \times N}$ is a spatially weighted matrix that quantifies the relationships between different vertices.

$$\tilde{\Lambda}^c(R_{pm}) = \Big(MLP\big(Pool_c(R_{pm})\big)\Big)^{\mathrm{T}} \cdot \Big(MLP\big(Pool_c(R_{pm})\big)\Big), \quad (1)$$

where $Pool_c(\cdot)$ denotes the global max pooling for each vertex embedding; $MLP(\cdot)$ is a multi-layer perceptron with one hidden layer. On the other hand,

$$\tilde{\Lambda}^s(R_{pm}) = \Big(Conv\big(Pool_s(R_{pm})\big)\Big) \cdot \Big(Conv\big(Pool_s(R_{pm})\big)\Big)^{\mathrm{T}}, \quad (2)$$

where $Pool_s(\cdot)$ represents the global max pooling for each position in the vertex embedding along the channel axis; $Conv(\cdot)$ is a 1 × 1 convolution layer. In this way, the data-dependent adaptive adjacency matrix $\bar{A}$ is given by spatial and channel attention-enhanced input vertex embeddings. We initialized the input-dependent adaptive adjacency matrix $\bar{A}$ as:

$$\begin{aligned}\bar{A} &= \psi(R_{pm}, W_\psi) \cdot \tilde{\Lambda}^c(R_{pm}) \cdot \psi(R_{pm}, W_\psi)^T \\ &\quad + \phi(R_{pm}, W_\phi) \cdot \phi(R_{pm}, W_\phi)^T \odot \tilde{\Lambda}^s(R_{pm}),\end{aligned} \quad (3)$$

where $\cdot$ represents matrix product; $\odot$ denotes Hadamard product; $\psi(R_{pm}, W_\psi) \in \mathbb{R}^{N \times C}$ and $\phi(R_{pm}, W_\phi) \in \mathbb{R}^{N \times C}$ are both

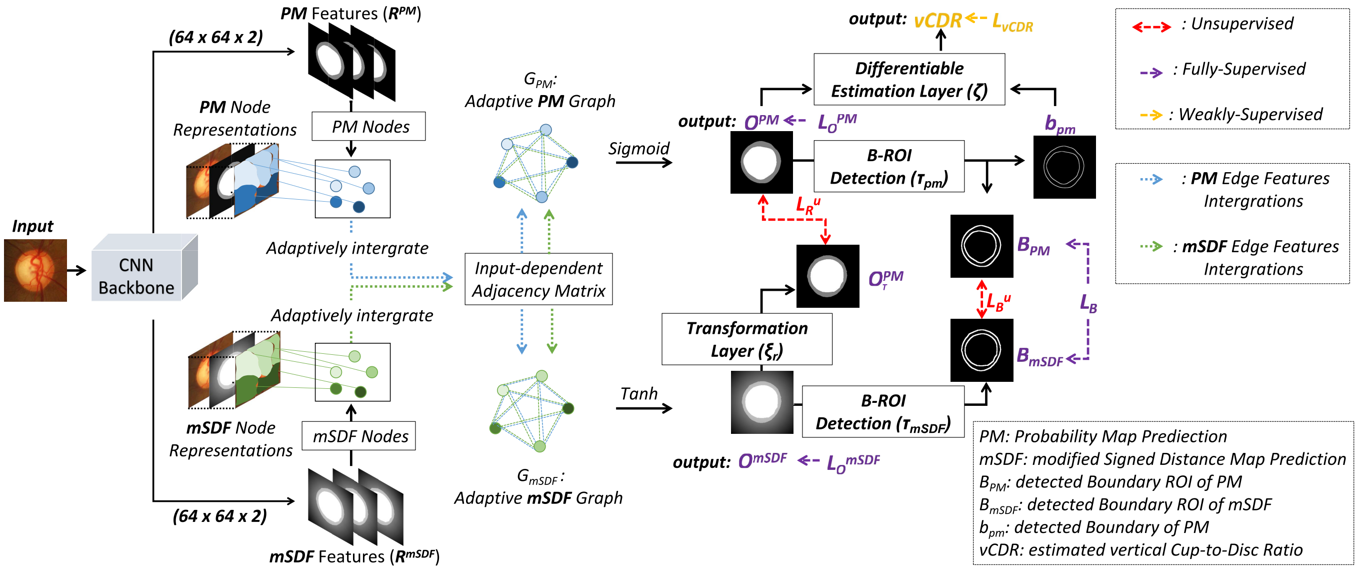

PM Features (R^{PM})
(64 x 64 x 2)
PM Node Representations
PM Nodes
Input
CNN Backbone
Adaptively intergrate
Adaptively intergrate
mSDF Node Representations
mSDF Nodes
(64 x 64 x 2)
mSDF Features (R^{mSDF})
G_{PM}: Adaptive PM Graph
Input-dependent Adjacency Matrix
G_{mSDF}: Adaptive mSDF Graph
Sigmoid
Tanh
output: vCDR ← L_{vCDR}
Differentiable Estimation Layer (ζ)
output: O^{PM} ← L_O^{PM}
B-ROI Detection (τ_{pm})
b_{pm}
L_R^u
O_τ^{PM}
Transformation Layer (ξ_r)
B_{PM}
L_B^u
L_B
B_{mSDF}
B-ROI Detection (τ_{mSDF})
output: O^{mSDF} ← L_O^{mSDF}
: Unsupervised
: Fully-Supervised
: Weakly-Supervised
: PM Edge Features Intergrations
: mSDF Edge Features Intergrations
PM: Probability Map Prediection
mSDF: modified Signed Distance Map Prediction
B_{PM}: detected Boundary ROI of PM
B_{mSDF}: detected Boundary ROI of mSDF
b_{pm}: detected Boundary of PM
vCDR: estimated vertical Cup-to-Disc Ratio

Fig. 2. Overview of the proposed *DAGCN* model (best viewed in color). O^{PM} and O^{mSDF} both have two channels to represent the output of *OC* and *OD* and we overlapped them for better visualization. L_O^{PM}, L_O^{mSDF}, L_B are the supervised *PM*, *mSDF* and *B-ROI* loss functions; L_{vCDR} is the weakly-supervised *vCDR* loss for *OD* & *OC* segmentation; L_{R^u} and L_{B^u} are the unsupervised region and *B-ROI* consistency losses.

linear embeddings (1 × 1 convolution); W_ψ and W_ϕ are learnable parameters. Secondly, we exploited the geometric association between *PM* and *mSDF* through integrating *mSDF* into the built *Laplacian* matrix $\tilde{L}$, which allowed us to adaptively built the graph according to their own constraints. Specifically, we fuse it into the spatial-wise weighted matrix $\tilde{\Lambda}^s(R_{pm})$. The geometry-aware spatial weighted matrix $\tilde{\Lambda}^s_g(R_{pm}, R_{mSDF})$ is given as follows:

$$\tilde{\Lambda}^s_g(R_{pm}, B_{mSDF}) = Conv\Big(Pool_s(R_{pm})\Big) \cdot \Big(Conv\Big(Pool_s(R_{pm} + R_{mSDF})\Big)\Big)^T \tag{4}$$

where $Conv(\cdot)$ is a 1 × 1 convolution layer. In this way, the semantic features of the object's foreground were emphasized by geometry-aware features of *mSDF*. As this is the case, the proposed adaptive graph convolution could take the spatial characteristics into account when reasoning the correlations between different regions. Then, the geometry-aware input-dependent adjacency matrix $\tilde{A}$ will be given as:

$$\tilde{A} = \psi(R_{pm}, W_\psi) \cdot \tilde{\Lambda}^c(R_{pm}) \cdot \psi(R_{pm}, W_\psi)^T + \zeta(R_{pm}, W_\zeta) \cdot \zeta(R_{pm}, W_\zeta)^T \odot \tilde{\Lambda}^s_g(R_{pm}, R_{mSDF}), \tag{5}$$

where $\zeta(R_s, W_\zeta) \in \mathbb{R}^{N\times C}$ is 1 × 1 convolution; W_ζ is learnable parameter. With the constructed $\tilde{A}$, the normalized *Laplacian* matrix is given as $\tilde{L} = I - \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$, where I is the identity matrix, $\tilde{D}$ is a diagonal matrix that represents the degree of each vertex, such that $\tilde{D}_{ii} = \sum_j \tilde{A}_{i,j}$. We calculated degree matrix $\tilde{D}$ with the same way that is used in [8] and [32], to override the computation overhead. Given computed $\tilde{L}$, with R_{PM} as the input vertex embeddings, we formulate the single-layer *DAGConv* as:

$$Y = \sigma(\tilde{L} \cdot R_{pm} \cdot W_G) + R_{pm}, \tag{6}$$

where $W_G \in \mathbb{R}^{C\times C}$ denotes the trainable weights of the *DAGConv*; σ is the ReLu activation function; Y is the output vertex features. Moreover, we add a residual connection to reserve the features of input vertices.

Please note that the graph construction and convolution process of G_{mSDF} is similar to G_{pm}, where the only difference is to replace R_{PM} to R_{mSDF} or reverse the position of R_{PM} and R_{mSDF}, from Eq. 1 to Eq. 6. In that case, the semantic features of *PM* is adaptively integrated into the geometry-aware *mSDF* during the graph construction of G_{mSDF}. As a result, the proposed *DAGCN* consists of two adaptive graphs (G_{pm} and G_{mSDF}), to reason the pixel-wise *PM* features and geometry-aware *mSDF* representations respectively and concurrently, with the benefits of their underlying geometric associations.

After the *DAGConv* (Eq. 6) in graph G_{pm} and graph G_{mSDF}, we apply bilinear up-sampling layers to scale the feature map in dual graph to the same size as input image. Then the *Sigmoid* and *Tanh* activation function were used to generate the *PM* output (O^{PM}) and *mSDF* output (O^{mSDF}) respectively. We then applied *Dice* loss (L_O^{PM}) and *MSE* loss (L_O^{mSDF}) on O^{PM} and O^{mSDF} respectively for all of the labeled input data, to supervise the dual regional predictions.

B. Dual Consistency Regularization of Semi-Supervised Manner

1) Modified Signed Distance Function (mSDF): Given O^{PM} and O^{mSDF}, we explored the geometric association between them and build the unsupervised dual consistency regularization losses via two differentiable transformation layers (ξ_r and τ). As mentioned above, various levels of information from different task branches can complement one another during training, whereas divergent focuses can lead to inherent prediction perturbation. The dual consistency regularization imposed the regional and marginal consistency in the task level in a semi-supervised manner. Given a target object

(*OD* or *OC*), the mSDF is defined as:

$$mSDF(x) = \begin{cases} 1, & x \in B_{in} \\ 0, & x \in \Delta B \\ -\inf_{y \in \Delta B} \|x-y\|_2, & x \in B_{out} \end{cases} \tag{7}$$

where $\|x-y\|_2$ represent the Euclidean distance between pixel x and y. Besides, B_{out}, B_{in} and ΔB denote the outside, inside, and boundary of the object, respectively. In other words, the absolute value of $mSDF(x)$ represented the distance between the point and the nearest point on the object's boundary, whereas the sign indicates whether the point is inside or outside the object. The differences between standard *SDM* and our proposed *mSDF* are twofold. Firstly, the *mSDF* has a reversed sign label against *SDF* because the learned *mSDF* features are used to build adjacency matrix along with *PM* features to learn a dual adaptive graph (*DAGCN*), it needs to have the similar feature space to the *PM* features before activation function (*e.g.* $R^{mSDF}(x) \rightarrow +\infty$, if $x \in B_{in}$). Secondly, we set the distance value of the inside region of *mSDF* to 1, for the ease of building regional consistency (Eq. (8)) between *PM* and *mSDF*. However, the proposed *mSDF* still has the similar attribute as the standard *SDF* to learn distance-aware spatial features. In this way, dual tasks can acquire the coherent semantic features, meanwhile the *mSDF* regression task benefits from the distance-aware spatial information supervision.

2) Regional Consistency: As for region-wise consistency, similar to [13], [20], and [14], we proposed a transformation layer to convert the O^{mSDF} to O^{PM} in a differentiable way. To be precise, the region-wise transformation layer ξ_r is defined as:

$$\xi_r(z) = 2 * Sigmoid(K \cdot ReLu(z)) - 1, \tag{8}$$

where z denotes the *mSDF* value at pixel x; K is a very large value; *Sigmoid* and *ReLu* are the non-linear activation functions. The larger K value indicates a closer approximation, and it is adopted as 5000 in this work. With Eq. 8, we could obtain the transformed segmentation maps O_T^{PM}, for example, $O_T^{PM} = \xi_r(O^{mSDF})$. For all of the unlabeled input, we applied a *Dice* loss (L_{R^u}) between O^{PM} and O_T^{PM} to enforce the unsupervised regional consistency regularization.

3) Marginal Consistency: We derived the spatial gradient of O^{PM} and O^{mSDF} as the estimated contours concerning the boundary-wise consistency. Previous studies [7] and [9] have proven that such narrow contours with a width of one pixel are challenging to optimize due to the highly unbalanced foreground and background, resulting in weakened consistency regularizations. Rather than focusing exclusively on the thin contour locations, we considered the *ROI* within a certain distance (boundary width) of the corresponding estimated contours. A simple yet efficient *B-ROI* detection layer (τ) is proposed for O^{PM} and O^{mSDF}. For example, τ_{PM} and τ_{mSDF} are defined as:

$$\tau_{PM} = O^{PM} + Maxpooling2D(-O^{PM}), \tag{9}$$

$$\tau_{mSDF} = \xi_r(O^{mSDF}) + Maxpooling2D\big(-\xi_r(O^{mSDF})\big), \tag{10}$$

The *Maxpooling2D* operation conducts the same feature map size as its input. It is worth noting that the output width of τ can be determined by varying the kernel size, stride, and padding value of the Maxpooling2D operation. We empirically set the output boundary width of τ_{PM} and τ_{mSDF} to 4 pixels in this work. After τ_{PM} and τ_{mSDF}, we referred to such *B-ROI* of O^{PM} and O^{mSDF} as B_{PM} and B_{mSDF}, respectively. Ideally, B_{PM} and B_{mSDF} should be close enough to one another. Thus, a *Dice* loss (L_{B^u}) between B_{PM} and B_{mSDF} was applied to enforce the unsupervised marginal consistency regularization of unlabeled data. Meanwhile, we apply a *Dice* loss (L_B) on both B_{PM} and B_{mSDF} to supervise the dual boundary predictions of labeled data.

C. Differentiable vCDR Estimation of Weakly Supervised Manner

Because the shapes of *OD* & *OC* are oval-like [1], previous methods resort to offline post-process the segmentation predictions with ellipse fitting to improve the segmentation accuracy [2], or to calculate *vCDR* using the approximated diameters of the *OD* & *OC* in the long axis [5], [6], [7]. However, they only use *vCDR* as an evaluation tool for glaucoma assessment but overlook the underlying supervision value of it in *OD* & *OC* segmentation task. Additionally, in the real world setting of clinical ophthalmology and ophthalmic image reading centres, clinicians and graders prefer to calculate the *vCDR* value with manually measured diameters of the *OD* & *OC* on the long axis, rather than to delineate the contour of *OD* & *OC* then calculating the *vCDR*, to save time. This results in a large number of labeled data with *vCDR* scalars; however, they have not been exploited in the computer vision community yet. For example, one of the datasets we used in this work (*UKBB*) contains 117,832 images with *vCDR* ground truth labeled. To address this issue, we took advantage of the specific domain knowledge between the boundary and region in terms of the perimeter and area of an oval-like shape to approximate the *vCDR* in a differentiable way.

To be precise, the *vCDR* is defined as the ratio of dividing the measured diameters of the cup by disc in the long axis. While such ratio can also be estimated given the size of perimeter and the area of *OD* and *OC*. According to the *Euler's Method* [34], the area (A_o) and perimeter (P_o) of the oval shape are defined as:

$$A_o = \pi \cdot a \cdot b, \tag{11}$$

$$P_o = \pi \cdot \sqrt{2(a^2 + b^2)}. \tag{12}$$

where a and b denote the semi-axis of the long and short axis of oval shape, respectively. We approximated A_o with the summed pixel value of O^{PM}, which can be regarded as the area of oval shape in pixel level. Furthermore, we derived the spatial gradient of O^{PM} via the *B-ROI* detection layer (τ_{PM}), to detect the boundary (b_{pm}) with width = 1. Then the summed pixel values of b_{pm} was approximately regarded as P_o. With Eq. 11 and Eq. 12, we could approximate a with A_o

and P_o, such as:

$$a = \sqrt{\frac{(P_o)^2 + \sqrt{(4\pi A_o + (P_o)^2) \cdot |((P_o)^2) - 4\pi A_o)|}}{4\pi^2}}), \tag{13}$$

where $|\cdot|$ is used to prevent sqrt from returning a negative value during the initial learning period. Given Eq. 13, we could calculate the *OD* long semi-axis (a^{OD}) and the *OC* long semi-axis (a^{OC}) with the respective P_o and A_o. Then, a *vCDR* estimation layer ζ was defined as:

$$\zeta(vCDR) = \frac{a^{OC} + e^{-6}}{a^{OD} + e^{-6}}, \tag{14}$$

where, e^{-6} is added to avoid dividing by zero errors. Given the prediction of *vCDR*, we apply a *MSE* loss (L_{vCDR}) between the prediction and ground truth to fully supervise the *vCDR* estimation and weakly-supervise the *OD* & *OC* segmentation.

IV. Experiments

A. Datasets

1) SEG Dataset: following the previous methods [7], [8], we pooled 2,068 images from five public available datasets (Refuge [1], Drishti-GS [36], ORIGA [37], RIGA [38], RIM-ONE [39]). These five datasets provided the fundus images and the ground truth masks, then we generated the corresponding ground truth of O^{mSDF}, B_{PM}, B_{mSDF} and *vCDR* with Eq. 7, 9, 10 and 14. Following the previous methods [7], [8], 613 fundus images were randomly selected as the test dataset, leaving the other 1,315 images for training and 140 images for validation.

2) UKBB Dataset: The UK Biobank [1] is a large-scale population-based biomedical database and research resource that contains detailed health information on half a million participants from the United Kingdom. Retinal colour photographs were acquired in a subset of participants that were scanned using the TOPCON 3D OCT 1000 Mk2 camera (Topcon Inc, Japan). The color fundus photographs have been graded for various eye diseases by NetwORC UK, a network of three UK Ophthalmic Reading Centers (Moorfields, Queen University of Belfast, and Liverpool) to support further scientific research on this invaluable dataset. First and foremost, the accredited graders evaluated the image quality to determine whether it was sufficient for measuring the *vCDR*. Then *vCDR* was calculated by dividing the measured diameter of the cup by the measured diameter of the disc in the long-axis or vertical direction. There were 117,832 fundus images with *vCDR* scalars are available, of which 38,421 were randomly selected as the weakly/semi-supervised training dataset, and the rest 79,411 are used as the test datasets.

B. Experimental Settings and Evaluation Metrics

We cropped the image of 256 × 256 pixels in the same way of [5], [7] and [8]. To avoid over-fitting, we adopted an on-the-fly data augmentation strategy. Specifically, we randomly flipped the training dataset with a probability of 0.5. We used stochastic gradient descent with a momentum of 0.9 to optimize the overall parameters. We trained the model for 10,000 iterations for all the experiments, with a learning rate of 1e-2 and a step decay rate of 0.999 every 100 iterations. The batch size was set as 56, consisting of 28 labeled and 28 unlabeled images. A backbone network [40] is used for ours and all the compared methods. The network was trained end-to-end; all the training processes were performed on a server with four *GEFORCE RTX 3090 24GiB GPUs*, and all the test experiments were conducted on a workstation with *Intel(R) Xeon(R) W-2104 CPU* and *Geforce RTX 2080Ti GPU* with 11GB memory. We used the output of the *PM* as the segmentation result. A fixed threshold of 0.5 is employed to obtain a binary mask from the probability map. Given the previously discussed loss function terms, we defined the overall loss function as:

$$Loss = L_O^{PM} + L_O^{mSDF} + L_B + \beta * (L_{R^u} + L_{B^u} + L_{vCDR}) \tag{15}$$

where β is adopted from [1] as the time-dependent Gaussian ramp-up weighting coefficient to trade-off between the supervised loss, unsupervised loss and weakly-supervised loss. This avoids the network getting stuck in a degenerated solution during the initial training period because no meaningful prediction of the unlabeled data, as well as *vCDR*, are obtained.

We reported Dice similarity score (*Dice*) as the region segmentation accuracy metrics; Boundary Intersection-over-Union (*BIoU*) [9] as the boundary segmentation metrics; and Mean Absolute Error (*MAE*) in pixel level, Pearson's correlation coefficients [41] (*Corr*), Bland-Altman analysis [42] as the *vCDR* estimation metric. 95% confidence intervals were generated by using 2,000 sample bootstrapping. Note that the Pearson's correlation coefficients [41] are used to measure the linear association. Paired t-test was used to assess statistical significance of the differences between our model and the compared methods. A p-value of < 0.05 was deemed as statistically significant.

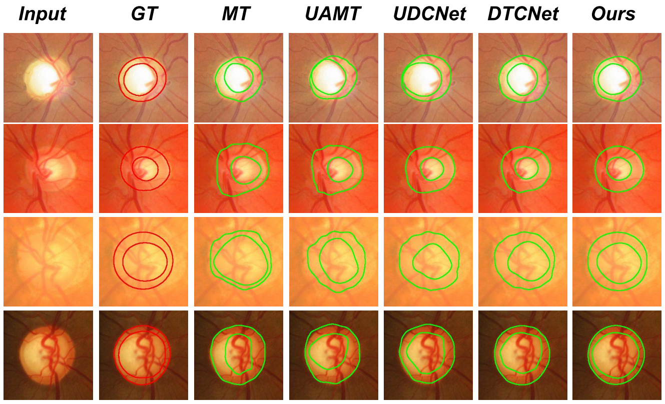


Fig. 3. Qualitative results of *OD* & *OC* segmentation in the *SEG* test dataset. We compare our model with *MT* [19], *UAMT* [17], *UDCNet* [16] and *DTCNet* [13]. Our method can produce more accurate segmentation results than the other methods when compared with the ground truth (*GT*). The boundaries were superimposed on the input image for better visualization of the segmentations.

[1] https://www.ukbiobank.ac.uk/

TABLE I

QUANTITATIVE SEGMENTATION RESULTS OF *OD & OC* AND GLAUCOMA ASSESSMENT ON *SEG* TESTING DATASETS. THE PERFORMANCE IS REPORTED AS *Dice* (%), *BIoU* (%), *MAE*, AND *Corr*. 95% CONFIDENCE INTERVALS ARE PRESENTED IN BRACKETS, RESPECTIVELY. WE COMPARE OUR MODEL WITH PREVIOUS STATE-OF-THE-ART FULLY-SUPERVISED METHODS BY RUNNING THEIR CODES IN THE PUBLIC DOMAIN. THE IMPLEMENTATION OF THE COMPARED STATE-OF-THE-ART SEMI-SUPERVISED WORKS IS MAINLY BASED ON AN OPEN-SOURCE CODEBASE [35]. *Ours (Semi)* ACHIEVES STATISTICALLY SIGNIFICANT IMPROVEMENTS CONSISTENTLY OVER OTHER COMPARED SEMI-SUPERVISED METHODS; PLEASE REFER TO TABLE. II FOR MORE DETAILS. UP AND DOWN ARROWS REPRESENT PROPORTIONAL AND INVERSELY PROPORTIONAL METRIC VALUE AND PERFORMANCE CORRELATIONS

Methods	SEG (OC)		SEG (OD)		*SEG (vCDR)*		*UKBB (vCDR)*	
	Dice (%)↑	*BIoU(%)*↑	*Dice (%)*↑	*BIoU(%)*↑	*MAE* ↓	*Corr* ↑	*MAE* ↓	*Corr* ↑
U-Net [10]	85.3 (82.1, 86.8)	80.1 (77.6, 82.4)	95.0 (93.1, 97.1)	86.2 (84.1, 88.3)	0.089 (0.079, 0.095)	0.685 (0.643, 0.713)	0.150 (0.140, 0.158)	0.301 (0.275, 0.329)
M-Net [2]	86.9 (85.0, 88.0)	82.9 (79.5, 84.7)	96.8 (95.5, 97.6)	88.1 (87.0, 89.3)	0.064 (0.051, 0.073)	0.707 (0.668, 0.741)	0.128 (0.119, 0.140)	0.365 (0.337, 0.390)
GRBNet [7]	89.4 (87.6, 90.8)	85.1 (83.3, 86.8)	97.7 (97.0, 98.7)	91.1 (90.2, 92.0)	0.056 (0.043, 0.067)	0.750 (0.739, 0.764)	0.118 (0.094, 0.134)	0.398 (0.371, 0.415)
RBA-Net [5]	87.8 (85.2, 89.7)	83.8 (81.6, 85.9)	96.1 (95.5, 96.7)	88.9 (88.0, 89.2)	0.062 (0.051, 0.073)	0.713 (0.690, 0.734)	0.126 (0.109, 0.142)	0.369 (0.350, 0.373)
MT [19]	84.1 (81.8, 85.7)	78.2 (77.0, 79.6)	94.3 (94.0, 94.7)	86.5 (85.0, 87.3)	0.091 (0.080, 0.099)	0.683 (0.641, 0.701)	0.145 (0.139, 0.150)	0.307 (0.276, 0.340)
UAMT [17]	85.3 (82.8, 86.9)	80.2 (79.0, 81.7)	95.2 (94.7, 95.6)	86.4 (85.1, 87.7)	0.075 (0.063, 0.081)	0.692 (0.642, 0.723)	0.134 (0.127, 0.139)	0.339 (0.301, 0.361)
URPC [15]	86.1 (83.1, 87.2)	81.2 (79.6, 82.0)	96.0 (95.4, 96.3)	87.3 (85.0, 87.9)	0.067 (0.059, 0.073)	0.701 (0.659, 0.742)	0.126 (0.121, 0.135)	0.361 (0.337, 0.382)
DTCNet [13]	86.1 (83.0, 87.4)	81.1 (79.5, 82.8)	96.1 (95.3, 96.4)	87.0 (85.2, 87.8)	0.065 (0.060, 0.072)	0.703 (0.661, 0.739)	0.126 (0.120, 0.137)	0.364 (0.339, 0.389)
UDCNet [16]	86.2 (83.3, 87.1)	81.4 (79.6, 83.0)	96.2 (95.7, 96.5)	87.1 (85.6, 87.9)	0.067 (0.059, 0.071)	0.714 (0.663, 0.742)	0.127 (0.119, 0.135)	0.389 (0.365, 0.412)
SASSNet [18]	85.8 (82.1, 87.3)	80.6 (78.2, 82.9)	95.7 (94.1, 96.5)	86.5 (85.4, 87.6)	0.070 (0.061, 0.079)	0.695 (0.633, 0.741)	0.139 (0.118, 0.153)	0.340 (0.313, 0.368)
Ours (Semi-100%)	**90.3** (89.6, 90.8)	**87.6** (83.6, 90.8)	**98.4** (98.4, 98.5)	**93.3** (92.1, 94.9)	**0.037** (0.035, 0.041)	**0.894** (0.863, 0.918)	**0.075** (0.073, 0.078)	**0.558** (0.514, 0.583)
Ours (Semi)	**88.2** (87.5, 88.9)	**84.1** (81.0, 87.6)	**97.6** (97.5, 97.8)	**89.9** (88.8, 90.7)	**0.047** (0.044, 0.051)	**0.848** (0.809, 0.879)	**0.097** (0.094, 0.099)	**0.463** (0.447, 0.480)

TABLE II

PAIRED T-TEST RESULTS BETWEEN *Ours (Semi)* AND THE COMPARED SEMI-SUPERVISED METHODS. WE PRESENTED THE *p*-VALUE OF THE MEAN *Dice* OF *OD & OC* SEGMENTATION ON *Seg* TEST DATASET; THE MEAN *MAE* OF *vCDR* ESTIMATION ON *UKBB* TEST DATASET; THE MEAN *AUROC* OF GLAUCOMA DIAGNOSIS ON *ORIGA*, *RIM-ONE*, *Refuge* TEST DATASETS; THE MEAN *Dice* OF *polyps* SEGMENTATION ON COLONOSCOPY POLYPS TEST DATASET. BECAUSE OUR MODEL ACHIEVES CONSISTENTLY BETTER PERFORMANCE THAN THE OTHER COMPARED SEMI-SUPERVISED METHODS ON THE FOUR TASKS, THE *p*-VALUE DEMONSTRATES THAT *Ours (Semi)* ACHIEVES STATISTICALLY SIGNIFICANT IMPROVEMENTS CONSISTENTLY OVER OTHER COMPARED SEMI-SUPERVISED METHODS

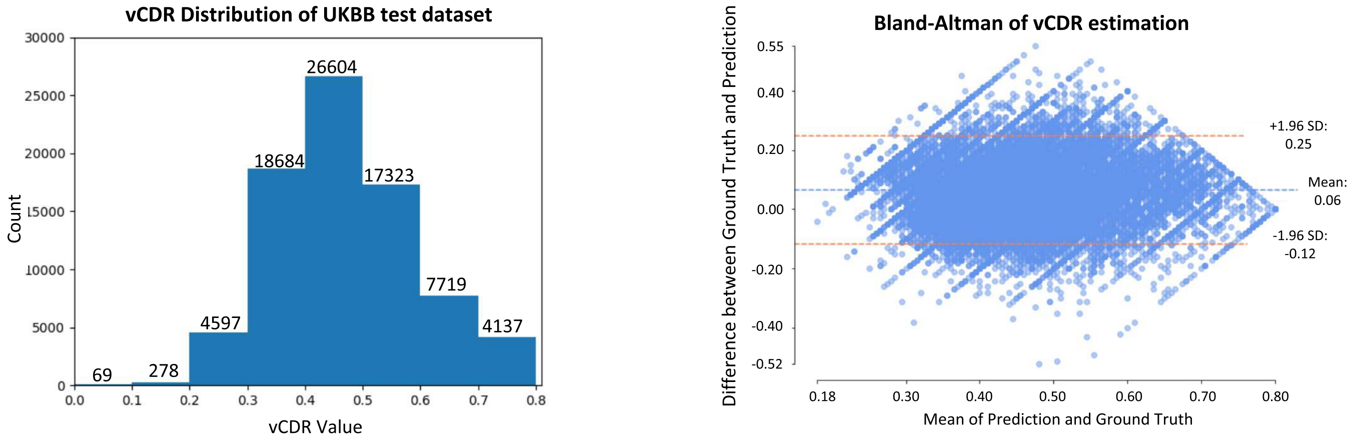
Tasks:	*Ours (Semi)* vs others	*MT* [19]	*UAMT* [17]	*URPC* [15]	*DTCNet* [13]	*UDCNet* [16]	*SASSNet* [18]
OD & OC	*p*-value (on *Dice*)	0.014	0.021	0.039	0.041	0.033	0.019
vCDR	*p*-value (on *MAE*)	0.018	0.029	0.040	0.044	0.036	0.020
Diagnosis	*p*-value (on *AUROC*)	0.009	0.021	0.033	0.037	0.029	0.011
Colonoscopy polyps	*p*-value (on *Dice*)	0.010	0.028	0.041	0.024	0.031	0.013

C. Performance Comparison and Analysis

In this section, we demonstrate the qualitative (Fig. 3) and quantitative (TABLE. I) results of the *OD & OC* segmentation and glaucoma assessment tasks. Specifically, in TABLE. I, we have presented the results of fully-supervised methods on the upper half part, and the rest are semi-supervised methods. All the fully-supervised methods were trained with 100% of the labeled *SEG* training dataset, and all the semi-supervised methods were trained with 5 % of *SEG* training dataset and 100 % of *UKBB* training dataset. In order to conduct complementary experiments, we trained our model with 100 % *SEG* and 100 % *UKBB* training data to fully utilise the available labeled and unlabeled data (*Ours (Semi-100%)*).

1) OD & OC Segmentation: Fig. 3 illustrates qualitative comparison with other semi-supervised methods on *SEG* test dataset. TABLE. I shows the quantitative performance of *Ours* and other methods under fully-supervised and semi-supervised manner, respectively. More experimental results for the data utilization efficiency can be found in Section V-A.

With only 5 % labeled segmentation training data, *Ours (Semi)* obtains an average 92.9 % *Dice* on *OC* and *OD* segmentation, outperforms data-level consistency regularization based methods *MT* [19], *UAMT* [17] by 4.2 % and 2.9 %, outperforms feature-level regularization based methods *URPC* [15] and *UDCNet* [16] by 2.0 % and 1.9 %, and outperforms adversarial regularization based method *SASSNet* [18] by 2.3 %. Paired t-tests on average *Dice* of *OD & OC* segmentation between *Ours (Semi)* and other semi-supervised methods were conducted to evaluate the statistical significance in the difference. The proposed method achieves statistically significant improvements consistently over other compared semi-supervised methods. Readers are referred to Table. II for the details. Distinctively, with sufficient labeled and unlabeled



(A) (B)

Fig. 4. (A): The *vCDR* distribution histogram of the *UKBB* test dataset. In total, there are 79,411 testing images with corresponding *vCDR* ground truth ranging from 0 to 0.8. (B): Bland-Altman plot of *vCDR* estimation for *Ours (Semi)* in *UKBB* test dataset. The x-axis and y-axis represents the mean and difference between ground truth and predicted *vCDR* value, respectively. The mean offsets and the limits of agreement, as well as the 95 % confidence interval on the mean values are shown.

TABLE III
Number of Parameters and *FLOPs* on a 256 × 256 Input Image

	M-net [2]	*RBA-Net* [5]	*GRBNet* [7]	*MT* [19]	*UAMT* [17]	*URPC* [15]	*DTCNet* [13]	*SASSNet* [18]	***Ours (Semi)***
Params (*M*)	27.7	34.3	24.7	26.3	26.3	27.2	26.7	29.5	28.6
FLOPs (*G*)	15.5	130.3	7.1	5.5	5.5	7.3	5.5	10.3	9.1

data, *Ours (Semi-100%)* achieved the best performance of averaged 94.4 % *Dice* on *OD & OC* segmentation, outperforming previous fully-supervised cutting-edge methods, such as *M-Net*, *RBA-Net* and *GRBNet* [7] by 2.7 %, 2.6% and 0.9 %.

2) Clinical Evaluation: vCDR Assessment: TABLE. I illustrates the *vCDR* evaluation results on *SEG* and *UKBB* test dataset respectively. The *UKBB (vCDR)* has 79,411 images, which is much larger than *SEG (vCDR)* (619 images). The performance on *UKBB (vCDR)* could reflect a more realistic situation in the real-world *w.r.t.* data distribution. Specifically, with only 5 % labeled *SEG* training data, *Ours (semi)* achieved the best performance of 0.097 *MAE* and 0.463 *Corr*, which outperforms *DTCNet* [13] by 23.0 % and 53.3 %. Paired t-tests on the *MAEs* of *vCDR* estimation between *Ours (Semi)* and other semi-supervised methods in Table. II were conducted to evaluate the statistical significance in the difference. Please note that, we utilised 38421 images of *UKBB* training dataset with the corresponding *vCDR* ground truth for weakly-supervised *OD & OC* segmentation and fully supervised *vCDR* estimation. On the other hand, with 100 % labeled *SEG* training dataset, *Ours (Semi-100%)* achieved much better performance with 0.075 *MAE* and 0.558 *Corr*, which is 22.7 % and 20.5 % better than *Ours (Semi)*. Additionally, the direct *vCDR* regression-based method [3] with all *UKBB* train data achieved 0.074 *MAE* but only 0.240 *Corr* on the *UKBB* test data. As the distribution of glaucoma patients and health participants are unbalanced, thus such regression model tends to predict closer to the majority of the distribution. The distribution of *vCDR* ground truth in *UKBB* test dataset is shown in Fig. 4 (A) for a better understanding of the data and our model's performance. In total, there were 79,411 test images with corresponding *vCDR* ground truth ranging from 0 to 0.8. It illustrated that the majority of *vCDR* ground truth distribution fell between 0.3 and 0.7. On the other hand, in order to evaluate mean biases and 95 % limit of agreements of estimated *vCDR*, a Bland-Altman plot [42] for *UKBB* test dataset was conducted and shown in Fig. 4 (B). The mean value of the offsets was 0.06, and the 95 % confidence interval was 0.18, which indicated a close agreement and minimal bias between the ground truth and our predictions. The bias occurs mainly for a value within the range of 0.3 to 0.7 in the majority of data distribution. However, our model performs well when *vCDR* is small or big (*e.g.* less than 0.3 or larger than 0.7), where little bias cases are observed.

D. Computational Efficiency

Tab. III demonstrates the number of parameters (*M*) and floating-point operations (*FLOPs*) of the compared models. *Ours (Semi)* and other compared models adopted the same backbone network, thus showing similar model size (*Params*). While, *RBA-Net* [5] has the largest model size and *FLOPs* because it contains several iterative feature aggregation modules, which requires more computations. *Ours (Semi)* contains 28.6*M* parameters and 9.1*G FLOPs*, which is comparable to other compared models.

V. Discussion and Conclusion

A. Ablation Study

We conducted detailed ablation studies with 5 % *SEG* training data and 100 % *UKBB* training data, and all the

TABLE IV
ABLATION STUDY ON GRAPH CONVOLUTIONS. THE PERFORMANCE IS REPORTED AS *Dice* (%), *BIoU* (%), *MAE* AND *Corr* ON TWO TEST DATASETS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Methods	SEG (OC)		SEG (OD)		UKBB (vCDR)	
	Dice (%)↑	*BIoU* (%)↑	*Dice* (%)↑	*BIoU* (%)↑	*MAE* ↓	*Corr* ↑
Classic GCN [31]	85.9	80.4	95.7	85.9	0.149	0.323
w/ Channel	86.8	82.8	95.8	86.8	0.121	0.349
w/ Spatial	87.1	83.0	96.0	87.1	0.109	0.407
w/ Both	87.6	83.4	96.6	87.8	0.108	0.411
w/ *SGR* [28]	87.2	83.6	96.5	87.7	0.105	0.430
w/ *DualGCN* [29]	87.5	83.7	96.6	88.1	0.104	0.427
w/ *GloRe* [30]	87.4	83.6	96.7	88.4	0.106	0.429
Ours (Semi)	**88.2**	**84.1**	**97.6**	**89.9**	**0.097**	**0.463**

TABLE V
ABLATION STUDY ON WEAKLY/SEMI-SUPERVISIONS. THE PERFORMANCE IS REPORTED AS *Dice* (%), *BIoU* (%), *MAE* AND *Corr* ON TWO TEST DATASETS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Methods	SEG (OC)		SEG (OD)		UKBB (vCDR)	
	Dice (%)↑	*BIoU* (%)↑	*Dice* (%)↑	*BIoU* (%)↑	*MAE* ↓	*Corr* ↑
w/o L_{R^u}	86.1	80.9	96.3	86.9	0.146	0.326
w/o L_{B^u}	86.5	81.7	96.5	87.4	0.131	0.345
w/ Both	86.8	82.6	96.8	88.4	0.123	0.348
w/ L_{vCDR}	87.1	82.9	96.7	88.8	0.108	0.415
w/ $L_{B^u}+L_{vCDR}$	87.3	83.3	96.9	88.9	0.106	0.434
w/ $L_{R^u}+L_{vCDR}$	87.4	83.2	97.1	89.1	0.102	0.443
Ours (Label-only)	80.5	70.7	91.6	75.8	0.628	0.118
Ours (Semi)	**88.2**	**84.1**	**97.6**	**89.9**	**0.097**	**0.463**

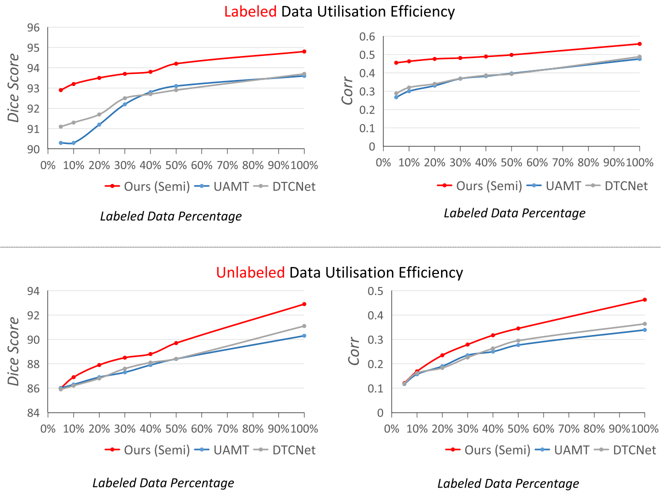


Fig. 5. The mean *OD* & *OC* segmentation performance of our semi-supervised approach with different ratio of labeled data. The performance is reported with *Dice* and *Corr*.

results demonstrated our model's effectiveness. As an illustration, the ablation results for different graph reasoning modules, weakly/semi-supervisions, and the efficiency analysis of data utilization are shown in TABLE. IV, TABLE. V and Fig. 5.

1) Graph Reasoning: In this section, we assessed the efficacy of the proposed *DAGCN*. Notably, we maintained the same dual graph structure while experimenting with various graph construction methods (via adjacency matrix) and graph convolutions. To begin, we use the classic graph convolution [31] to reason about the relationships between the *PM* and the *mSDF*, respectively. Then, we investigated input-dependent graph convolutions in terms of channel attention (*w/ Channel*) and spatial attention (*w/ Spatial*) mechanisms, both separately and concurrently (*w/ Both*). Additionally, we adopt three more powerful graph reasoning modules to demonstrate the superiority of our proposed *DAGCN*. In particular, we use the *SGR* [28], *DualGCN* [29], and *GloRe* module [30] respectively. In detail, the *SR* module exploits knowledge graph mechanism; *DualGCN* investigates the coordinate space and feature space graph convolution; and *GloRe* leverage projection and re-projection mechanism to reason the semantics between different regions. Note that the methods mentioned above belong to single graph reasoning; thus, we have built two separate graphs for *PM* segmentation and *mSDF* regression individually, where there was no associations or geometric associations between the dual graph. TABLE. IV shows that our model achieved more accurate and reliable results than [31] and outperformed the *SGR* [28], *DualGCN* [29], and *GloRe* [30] by 1.1 %, 0.9 % and 0.9 % mean *Dice* on the *SEG* test datasets.

2) Weakly/Semi-Supervision: We performed experiments to evaluate the effectiveness of the proposed dual consistency regularization paradigm in semi-supervised learning and the proposed differentiable *vCDR* estimation module in a weakly-supervised manner. The results are shown in TABLE. V. Specifically, we evaluated the region-wise consistency loss, the boundary-wise consistency loss, and the *vCDR* estimation loss, respectively. We have represented our model that is trained with only 5 % *SEG* training data as *Ours (Label-only)*. Firstly, we have retained the same model structure and eliminate the *vCDR* estimation loss to focus on the dual consistency regularization losses (*w/ Both*). Following that, we have removed the region-wise unsupervised loss (*w/o* L_{R^u}), boundary-wise unsupervised loss (*w/o* L_{B^u}) respectively. Secondly, we removed both of the consistency losses and only applied the weakly-supervised *vCDR* estimation loss (*w/* L_{vCDR}). Then we added the other two unsupervised consistency losses individually (*w/* $L_{B^u}+L_{vCDR}$ and *w/* $L_{R^u}+L_{vCDR}$) to see if the performance were boosted. TABLE. V demonstrates that the proposed unsupervised dual consistency losses and weakly supervised loss could improve the model by 6.6 % and 6.5 % mean *Dice* for segmentation. Particularly, the boundary-wise unsupervised loss can increase the model by 6.2 % *BIoU*, which leads to a better boundary segmentation quality. The weakly supervised loss can bring a large improvement of 82.8 % *MAE* of *vCDR* estimation, which is the ultimate goal for *OD* & *OC* segmentation task *w.r.t* clinic application.

3) Data Utilization Efficiency: In this section, we show more ablation study results on the data utilization efficiency. In detail, we have examined the performance of cutting-edge semi-supervised methods *UAMT* [17], *DTCNet* [13] and *Ours (Semi)* with different ratio of labeled and unlabeled images. We evaluated the segmentation performance on the *SEG* test dataset with *Dice* and the *vCDR* estimation performance on the *UKBB* test dataset with *Corr*, respectively. As for the labeled images, we vary the ratio of labeled segmentation

TABLE VI

QUANTITATIVE COMPARISONS BETWEEN THE GROUND TRUTH *vCDR* VALUES (*GT vCDR*), *Ours (semi)*, *Ours (Semi-100 %)* AND OTHER CUTTING-EDGE SEMI-SUPERVISED METHODS FOR THE GLAUCOMA CLASSIFICATION PERFORMANCE ON ORIGA [37], RIM-ONE [39], AND REFUGE [1] TEST DATASET. THE PERFORMANCE IS REPORTED AS *Precision (%)*, *Specificity (%)*, *Sensitivity (%)*, *AUROC (%)*. 95 % CONFIDENCE INTERVALS ARE PRESENTED IN THE BRACKETS

Methods	***ORIGA*** [37]				***RIM-ONE*** [39]				***Refuge*** [1]			
	Precision(%)	*Specificity(%)*	*Sensitivity(%)*	*AUROC(%)*	*Precision(%)*	*Specificity(%)*	*Sensitivity(%)*	*AUROC(%)*	*Precision(%)*	*Specificity(%)*	*Sensitivity(%)*	*AUROC(%)*
MT [19]	27.8 (21.1, 34.9)	25.5 (18.6, 32.9)	**95.7** (88.9, 100.0)	76.3 (68.7, 83.2)	28.0 (11.5, 46.7)	55.0 (39.5, 70.0)	**87.5** (64.0, 100.0)	88.8 (73.4, 99.3)	38.2 (22.6, 55.6)	81.3 (73.8, 88.1)	**86.7** (66.7,100.0)	95.0 (88.7, 99.4)
UAMT [17]	38.9 (29.4, 18.9)	62.1 (54.4, 69.9)	80.4 (68.1, 91.7)	76.7 (69.0, 83.6)	50.0 (16.7, 83.3)	87.5 (76.3, 97.4)	62.5 (25.0, 100.0)	83.8 (66.7, 96.8)	84.6 (62.5, 100.0)	98.2 (95.4, 100.0)	73.3 (50.0, 93.8)	95.4 (88.6, 99.6)
URPC [15]	44.7 (34.1, 55.4)	69.3 (62.0, 76.6)	82.6 (70.8, 92.7)	80.2 (72.9, 87.1)	41.7 (12.5, 71.4)	82.5 (70.0, 93.9)	62.5 (25.0, 100.0)	84.1 (65.3, 97.6)	90.9 (69.2, 100.0)	**99.1** (97.2, 100.0)	66.7 (40.0, 90.0)	94.7 (85.6, 99.6)
DTCNet [13]	37.4 (27.7, 47.1)	59.5 (51.6, 67.3)	80.4 (68.2, 91.1)	76.2 (67.7, 84.1)	44.4 (11.1, 80.0)	87.5 (76.5, 97.4)	50.0 (14.3, 85.7)	86.6 (72.5, 97.2)	83.3 (58.3, 100.0)	98.2 (95.4, 100.0)	66.7 (40.0, 90.0)	93.7 (83.2, 99.6)
UDCNet [16]	45.3 (34.3, 57.0)	73.2 (65.6, 80.0)	73.9 (60.0, 86.5)	80.2 (72.8, 87.1)	50.0 (14.3, 87.5)	90.0 (80.4,85.7)	50.0 (14.3, 85.7)	87.8 (73.9, 98.4)	83.3 (58.5, 100.0)	98.2 (95.7, 100.0)	60.0 (35.0, 85.7)	93.6 (82.7, 99.6)
SASSNet [18]	38.5 (28.6, 48.7)	63.4 (55.7, 71.2)	76.1 (63.5, 88.4)	77.7 (70.0, 84.9)	40.0 (10.0, 72.7)	85.0 (73.7, 95.0)	50.0 (14.3, 85.7)	85.0 (69.8, 97.3)	76.9 (50.0, 100.0)	97.3 (93.8, 100.0)	66.7 (40.0, 90.0)	94.8 (87.1, 99.3)
GT vCDR	45.5 (35.6, 55.8)	68.6 (60.9, 76.0)	87.0 (76.7, 96.0)	82.9 (76.1, 88.9)	63.6 (33.3, 90.9)	90.0 (80.0, 97.6)	**87.5** (60.0, 100.0)	91.3 (75.4, 100.0)	81.8 (44.4, 87.5)	98.2 (95.5, 100.0)	60.0 (33.3 84.6)	90.9 (81.0, 98.3)
Ours (Semi)	45.7 (35.6, 55.7)	67.3 (59.7, 75.3)	91.3 (88.2, 98.0)	85.5 (78.9, 91.3)	42.9 (16.7, 71.4)	80.0 (66.7, 92.3)	75.0 (40.0, 100.0)	90.6 (78.6, 99.1)	**91.7** (71.4, 100.0)	**99.1** (97.2, 100.0)	73.3 (50.0, 93.8)	**95.7** (90.2, 99.3)
Ours (Semi-100%)	**54.2** (42.5, 65.4)	**78.4** (71.1, 85.1)	84.8 (72.7, 94.4)	**86.5** (80.3, 91.9)	**66.6** (33.3, 100.0)	**92.5** (84.2, 100.0)	75.0 (40.0, 100.0)	**91.6** (78.4, 99.7)	90.0 (66.7, 100.0)	**99.1** (97.2, 100.0)	60.0 (35.0, 85.7)	**95.7** (90.0, 99.7)

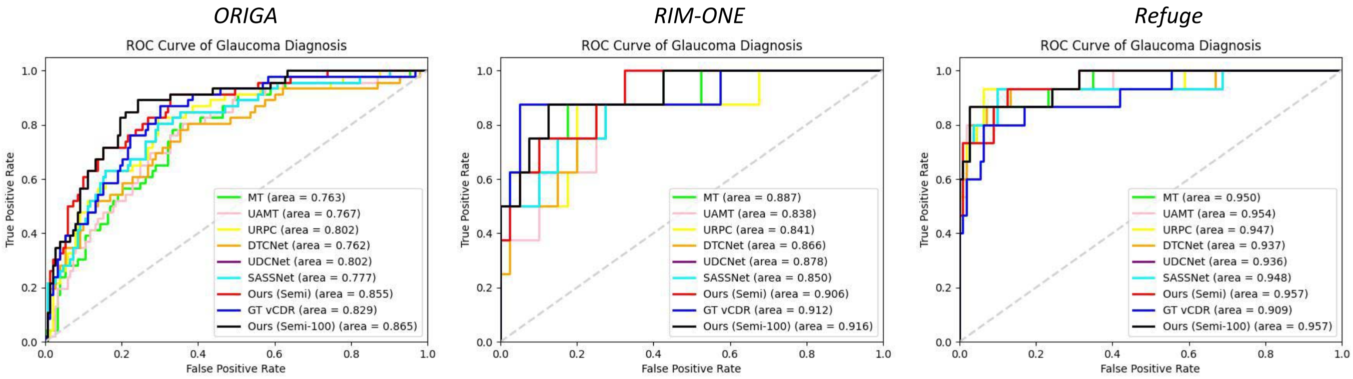


Fig. 6. ROC curves showing the glaucoma classification performance using the Ground Truth *vCDR* values (*GT vCDR*), *Ours (Semi)*, *Ours (Semi-100 %)* and other cutting-edge semi-supervised methods on ORIGA [37], RIM-ONE [39], and Refuge [1] test dataset, respectively.

images from 5 % to 100 % (out of 1315 *SEG* training data) while fixing the number of unlabeled images to be 38421 (100 % *UKBB* training data). The performance are shown in the top of Fig. 5 for the averaged *OD* & *OC* segmentation performance and *vCDR* estimation, respectively. It shows *Ours (Semi)* achieves consistent superior performance over the *UAMT* [17], *DTCNet* [13] on both tasks under different labeled data utilizations. Primarily when less labeled data is used, *Ours (Semi)* suppressed the other two methods by a large margin. On the other hand, for unlabeled images, we varied the ratio of unlabeled segmentation images from 5 % to 100 % (out of 38421 *UKBB* training data) while fixing the number of labeled images to be 73 (5 % *SEG* training data). The performance are shown in the bottom of Fig. 5 for the averaged *OD* & *OC* segmentation performance and *vCDR* estimation, respectively. It shows *Ours (Semi)* achieved consistent superior performance over the *UAMT* [17], *DTCNet* [13] on both tasks under different unlabeled data utilizations, which indicated that our method effectively utilized the unlabeled data. When more unlabeled data is used, *Ours (Semi)* significantly outperformed the other two methods by a large margin.

B. Glaucoma Diagnosis

In order to understand the relevance of the glaucoma diagnosis and the *vCDR* value, we conducted a classification evaluation based on the given glaucoma and healthy participant labels. Among the datasets used in this work, the glaucoma classification labels are available in RIM-ONE [39], Refuge [1], and ORIGA [37] datasets. Their corresponding test datasets with glaucoma and healthy participant labels are used in this section. In detail, there were 40 healthy participants and 8 glaucoma patients in the RIM-ONE test dataset; 112 healthy participants and 15 glaucoma patients in the Refuge test dataset; 153 healthy participants and 46 glaucoma patients in the ORIGA test dataset. We compared the aforementioned semi-supervised methods (in TABLE. I), to evaluate the *vCDR* assessment performance in glaucoma diagnosis. *Precision*, *Specificity*, *Sensitivity* and Area Under the Receiver Operating Characteristic (*AUROC*) were used as the classification metrics. Specifically, a *vCDR* value larger than 0.6 is considered at risk for glaucoma, because the optic nerve damage from increased eye pressure reflected by an increase in the *vCDR* value [45], [46], [47]. TABLE. VI shows the quantitative comparison between ours and previous cutting-edge semi-supervised methods on the three test datasets, respectively. *GT vCDR* represents the glaucoma diagnosis performance using the ground truth *vCDR* values of three test datasets. Specifically, *Ours (Semi-100%)* achieved consistently better *Precision* and *Specificity* than *GT vCDR* and other compared semi-supervised methods on the three test datasets. Fig. 6 demonstrates the *ROC Curve* comparison, where *Ours (Semi-100%)* obtains 86.5 %, 91.6% and 95.7% *AUROC* scores

TABLE VII

QUANTITATIVE SEGMENTATION RESULTS OF POLYPS ON THE TEST DATASET. THE PERFORMANCE IS REPORTED AS *Dice* (%) AND *BIoU* (%). 95% CONFIDENCE INTERVALS ARE PRESENTED IN BRACKETS, RESPECTIVELY

Methods	Fully-Supervised				Semi-Supervised						
	ACSNet [43]	*BI-GCN* [8]	*PraNet* [44]	*GRBNet* [7]	*MT* [19]	*UAMT* [17]	*URPC* [15]	*DTCNet* [13]	*UDCNet* [16]	*SASSNet* [18]	***Ours (Semi)***
Dice (%)	70.1 (67.8, 72.3)	73.2 (70.7, 75.8)	74.0 (72.6, 75.7)	75.7 (73.1, 77.6)	69.6 (67.2, 72.0)	70.9 (68.4, 73.3)	72.6 (70.1, 74.9)	73.1 (70.6, 75.5)	71.5 (69.0, 74.1)	71.6 (69.0, 74.0)	**74.7** (72.1, 77.0)
BIoU (%)	65.2 (62.5, 67.7)	67.5 (65.9, 70.7)	66.0 (63.3, 68.9)	69.3 (67.9, 70.5)	64.3 (61.0, 67.9)	65.4 (61.7, 67.7)	66.8 (62.3, 69.7)	66.8 (63.4, 70.0)	65.9 (62.1, 69.8)	65.4 (62.0, 68.7)	**68.7** (64.2, 71.1)

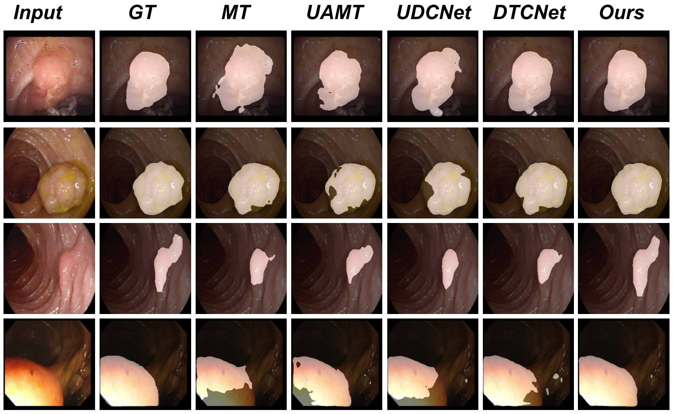


Fig. 7. Qualitative results of colonoscopy polyps segmentation in the polyps segmentation test dataset. We compare our model with *MT* [19], *UAMT* [17], *UDCNet* [16] and *DTCNet* [13]. Our method can produce more accurate segmentation results when compared with the ground truth (*GT*).

on ORIGA, RIM-ONE and Refuge test datasets respectively, which is consistently better than *GT vCDR*. Paired t-test results on *AUROC* of glaucoma diagnosis between *Ours (Semi)* and other semi-supervised methods were conducted using bootstrapping [48] and are shown in TABLE. II, which indicates that our method achieved statistically significant improvements over other semi-supervised methods. Notably, paired T-test between *Ours (Semi-100%)* and *GT vCDR* also suggests that our method outperforms the *vCDR* ground truth with a statistically significant difference in performance ($P < 0.05$). The potential reason for not so perfect *GT vCDR* diagnosis performance and our better *AUROC* performance could be twofold. Firstly, glaucoma patients usually have a higher *vCDR* compared to healthy people; however, there is a significant overlap in *vCDR* between healthy individuals and glaucoma patients [49]. Thus, only relying on *vCDR* value cannot guarantee an accurate glaucoma diagnosis. Instead, it can be used as a strong clinical indicator for suspected disc in clinical practice [45], [46]. Secondly, some of the *OD & OC* ground truth annotations in the aforementioned datasets may not be accurate, thus leading to an inaccurate *vCDR* value, which has also been noted previously [7].

C. Generalizability of Dual Consistency Regularization

In order to verify the generalizability of our proposed dual consistency regularization mechanism in semi-supervised learning, we conducted external experiments on a large-scale colonoscopy polyps segmentation benchmark [44] that has been validated by previous methods [7], [8], [43]. The polyp shapes in the dataset are irregular and complex, which compose a more challenging task than the *OD & OC* segmentation. The dataset contains 2,247 colonoscopy images from five datasets (ETIS [50], CVC-ClinicDB [51], CVC-ColonDB [52], EndoScene-CVC300 [53], and Kvasir [54]). All the images were resized to 256 × 256 pixels. As suggested by [44] in fully-supervised data setting, *i.e.*, 1,450 images from Kvasir [54] and CVC-ClinicDB [51] were selected as the training datasets and the other 635 images from ETIS [50], CVC-ColonDB [52], EndoScene-CVC300 [53] are pooled together for independent testing (**unseen data**). By doing this, the training and test data are from different data sources so as to evaluate the model's generalization capability. Note that 10 % of training datasets were randomly selected as internal validation. For the semi-supervised data setting in this section, we used 50 % of the training data as the labeled data and the rest 50 % training data was used as the unlabeled data. As for the framework structure, the differentiable *vCDR* estimation module was specially designed for the *OD & OC* segmentation tasks with prior knowledge of ellipse shape objects. Thus, we removed it and remained the rest of the structure (*Ours (Semi)*) as our framework in this section. The quantitative results are shown in TABLE. VII, where *Ours (Semi)* achieved 74.7 % *Dice* with only 50 % labeled training data, which is comparable to the previous fully-supervised cutting-edge methods *PraNet* [44] and *GRBNet* [7]. On the other hand, our model outperformed other state-of-the-art semi-supervised methods *UAMT* [17], *URPC* [15], and *UDCNet* [16] by 5.4 %, 2.9 %, and 4.5 % in *Dice*, respectively. Paired T-test on *Dice* of segmentation between *Ours (Semi)* and other semi-supervised methods suggests a statistically significant difference in performance ($P < 0.05$). We have shown the qualitative results comparison in the Fig. 7, where *Ours* could generate a more accurate polyps segmentation performance compared to other semi-supervised methods. This demonstrated that our proposed dual consistency regularization mechanisms could generalise to more complex objects with irregular shapes.

D. Limitations

Regarding to the *vCDR* estimation performance in *UKBB* test dataset, *Ours (Semi)* and *Ours (Semi-100 %)* could gain 0.463 and 0.558 *Corr*, respectively. Although our method outperformed the compared cutting-edge semi-supervised methods in TABLE. I, the performance is moderate if applied to real-world clinical applications. The main reason for moderate *Corr* performance could be threefold. Firstly, the limited number of labeled segmentation masks for training would

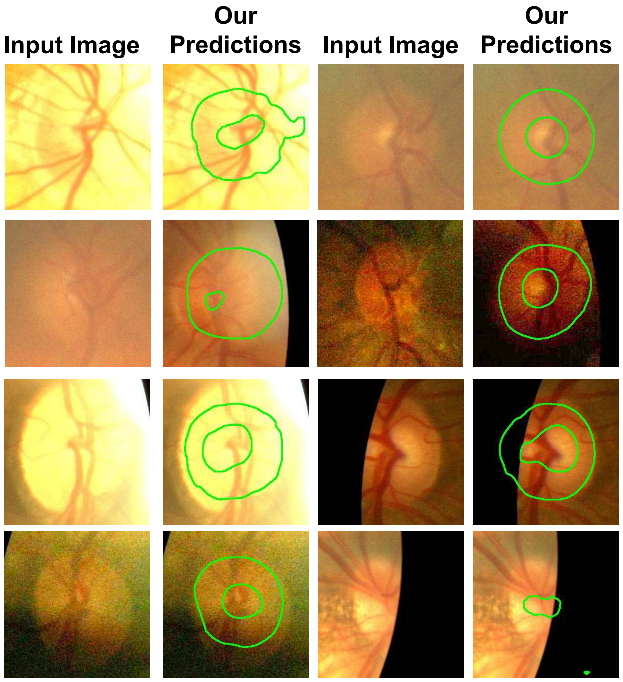


Fig. 8. Examples of the input image and our model's predictions (*Ours (Semi)*) in some challenging cases. The proposed model failed to segment the *OD & OC* if the image quality is considerably poor, such as incomplete *OD & OC* region, blurred area, extremely low-contrast, *etc.*

undoubtedly affect the model's performance. Our model could achieve better *Corr* if given more labeled data. For example, in TABLE. I, *Ours (Semi-100 %)* outperforms *Ours (Semi)* by 20.5 % of *Corr* on *UKBB* test dataset. Secondly, the underlying low-quality input images also lead to limited performance. We considered *vCDR* estimation to be 'failed' if it fell outside 95 % confidence interval of the Bland-Altman plot in Fig. 4 (B). According to these criteria, we showed some of the 'failed' predictions in Fig. 8. It illustrates that our model could not accurately segment the *OC* and *OD* if the image quality was relatively low, such as incomplete *OD & OC* region, blurred area, extremely low-contrast, *etc.*. Thirdly, an extremely unbalanced data distribution could contribute to a moderate *Corr* performance. As the *vCDR* distribution and Bland-Altman plot shown above, the majority of *vCDR* falls between 0.3 to 0.7, where the bias mainly occurs. The glaucoma diagnosis evaluation presented in Section V-B further demonstrates that our method could achieve satisfying diagnosis performance, even when compared to the *vCDR* ground truth.

On the other hand, the designed dual consistency regularization mechanism can be widely applied to other semi-supervised medical image segmentation tasks such as ultrasound fetal head segmentation, *etc.* However, it may not work for highly complex objects, such as curvilinear structures like blood vessels [11], whose region and boundary areas can be challenging to distinguish due to their composite topology and tortuosity. Thus, an inevitable perturbation will be introduced in the marginal and regional consistency regularization, thus impacting the semi-supervised segmentation performance.

E. Conclusion

We have proposed a novel graph-based weakly/semi-supervised segmentation framework. The geometric associations between the pixel-wise probability map features, modified signed distance function representations and object boundary characteristics are exploited in the proposed dual graph model, semi-supervised consistency regularizations, and weakly-supervised guidance. Remarkably, the proposed differential *vCDR* estimation module boosts the proposed model with a significant improvement in glaucoma assessment. Apart from the performance, It has facilitated our model to leverage an extensive data set with no segmentation but only *vCDR* labels. Such data and labels commonly exist in real-world clinical circumstances (*UK-Biobank*); however, they are usually understudied. Our experiments have demonstrated that the proposed model can effectively leverage semantic region features and spatial boundary features for segmentation of optic disc & optic cup and *vCDR* estimation for glaucoma assessment from retinal images. We believe our proposed method can be easily extended to explore geometric associations between more feature representations, such as regions, surfaces, boundaries, and landmarks in different medical image segmentation tasks.

Acknowledgment

The authors would like to thank Prof. Rongchang Zhao for the discussion and share of their models for the comparison.

References

[1] J. I. Orlando *et al.*, "REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Med. Image Anal.*, vol. 59, Jan. 2020, Art. no. 101570.

[2] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1597–1605, Jul. 2018.

[3] R. Zhao, X. Chen, X. Liu, Z. Chen, F. Guo, and S. Li, "Direct cup-to-disc ratio estimation for glaucoma screening via semi-supervised learning," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 4, pp. 1104–1113, Apr. 2020.

[4] J. Wu *et al.*, "Oval shape constraint based optic disc and cup segmentation in fundus photographs," in *Proc. BMVC*, 2019, p. 265.

[5] Y. Meng *et al.*, "Regression of instance boundary by aggregated CNN and GCN," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 190–207.

[6] Y. Meng *et al.*, "CNN-GCN aggregation enabled boundary regression for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, 2020, pp. 352–362.

[7] Y. Meng *et al.*, "Graph-based region and boundary aggregation for biomedical image segmentation," *IEEE Trans. Med. Imag.*, vol. 41, no. 3, pp. 690–701, Mar. 2022.

[8] Y. Meng *et al.*, "BI-GCN: Boundary-aware input-dependent graph convolution network for biomedical image segmentation," in *Proc. 32nd Brit. Mach. Vis. Conf. (BMVC)*, 2021, pp. 1–14.

[9] B. Cheng, R. Girshick, P. Dollar, A. C. Berg, and A. Kirillov, "Boundary IoU: Improving object-centric image segmentation evaluation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15334–15342.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2015, pp. 234–241.

[11] Z. Gu *et al.*, "Ce-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[12] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. ICLR*, 2016, pp. 1–13.

[13] X. Luo, J. Chen, T. Song, and G. Wang, "Semi-supervised medical image segmentation through dual-task consistency," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 10, 2021, pp. 8801–8809.

[14] Y. Xue *et al.*, "Shape-aware organ segmentation by predicting signed distance maps," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, 2020, pp. 12565–12572.

[15] X. Luo *et al.*, "Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, 2021, pp. 318–329.

[16] Y. Li, L. Luo, H. Lin, H. Chen, and P.-A. Heng, "Dual-consistency semi-supervised learning with uncertainty quantification for COVID-19 lesion segmentation from CT images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, 2021, pp. 199–209.

[17] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2019, pp. 605–613.

[18] S. Li, C. Zhang, and X. He, "Shape-aware semi-supervised 3D semantic segmentation for medical images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2020, pp. 552–561.

[19] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1195–1204.

[20] Y. Meng *et al.*, "Spatial uncertainty-aware semi-supervised crowd counting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 15549–15559.

[21] Y. Lu *et al.*, "Taskology: Utilizing task relations at scale," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8700–8709.

[22] A. R. Zamir *et al.*, "Robust learning through cross-task consistency," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11197–11206.

[23] M. Rajchl *et al.*, "DeepCut: Object segmentation from bounding box annotations using convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 674–683, Jun. 2017.

[24] J. Lee, E. Kim, S. Lee, J. Lee, and S. Yoon, "FickleNet: Weakly and semi-supervised semantic image segmentation using stochastic inference," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5267–5276.

[25] I. Laradji *et al.*, "A weakly supervised consistency-based learning method for COVID-19 segmentation in CT images," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 2453–2462.

[26] X. Liu *et al.*, "Weakly supervised segmentation of COVID19 infection with scribble annotation on CT images," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108341.

[27] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2921–2929.

[28] X. Liang, Z. Hu, H. Zhang, L. Lin, and E. P. Xing, "Symbolic graph reasoning meets convolutions," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 1858–1868.

[29] L. Zhang, X. Li, A. Arnab, K. Yang, Y. Tong, and P. H. Torr, "Dual graph convolutional network for semantic segmentation," in *Proc. BMVC*, 2019, pp. 1–18.

[30] Y. Chen, M. Rohrbach, Z. Yan, Y. Shuicheng, J. Feng, and Y. Kalantidis, "Graph-based global reasoning networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 433–442.

[31] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. ICLR*, 2017, pp. 1–14.

[32] X. Li, Y. Yang, Q. Zhao, T. Shen, Z. Lin, and H. Liu, "Spatial pyramid based graph reasoning for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8950–8959.

[33] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. NIPS*, 2016, pp. 3844–3852.

[34] E. Lockwood, "Length of ellipse," *Math. Gazette*, vol. 16, no. 220, pp. 269–270, 1932.

[35] X. Luo. (2020). *Ssl4mis*. [Online]. Available: https://github.com/hilab-git/ssl4mis

[36] J. Sivaswamy, S. R. Krishnadas, G. D. Joshi, M. Jain, and A. U. S. Tabish, "Drishti-GS: Retinal image dataset for optic nerve head(ONH) segmentation," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2014, pp. 53–56.

[37] Z. Zhang *et al.*, "ORIGA-light: An online retinal fundus image database for glaucoma analysis and research," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug. 2010, pp. 3065–3068.

[38] A. Almazroa *et al.*, "Retinal fundus images for glaucoma analysis: The RIGA dataset," *Proc. SPIE*, vol. 10579, Mar. 2018, Art. no. 105790B.

[39] F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "RIM-ONE: An open retinal image database for optic nerve evaluation," in *Proc. 24th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2011, pp. 1–6.

[40] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2Net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, Feb. 2021.

[41] M. M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi Med. J.*, vol. 24, no. 3, pp. 69–71, 2012.

[42] J. M. Bland and D. G. Altman, "Measuring agreement in method comparison studies," *Stat. Methods Med. Res.*, vol. 8, no. 2, pp. 135–160, Apr. 1999.

[43] R. Zhang, G. Li, Z. Li, S. Cui, D. Qian, and Y. Yu, "Adaptive context selection for polyp segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2020, pp. 253–262.

[44] D.-P. Fan *et al.*, "PraNet: Parallel reverse attention network for polyp segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2020, pp. 263–273.

[45] Y. Ikeda *et al.*, "Ten-year of glaucoma transition rate on the basis of optic nerve morphology in normal Japanese subjects," *Investigative Ophthalmol. Vis. Sci.*, vol. 60, no. 9, p. 1968, 2019.

[46] A. O. Amedo *et al.*, "Comparison of the clinical estimation of cup-to-disk ratio by direct ophthalmoscopy and optical coherence tomography," *Therapeutic Adv. Ophthalmol,.*, vol. 11, Mar. 2019, Art. no. 2515841419827268.

[47] P. A. Alhadeff, C. G. De Moraes, M. Chen, A. S. Raza, R. Ritch, and D. C. Hood, "The association between clinical features seen on fundus photographs and glaucomatous damage detected on visual fields and optical coherence tomography scans," *J. Glaucoma*, vol. 26, no. 5, p. 498, 2017.

[48] X. Robin *et al.*, "PROC: An open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinf.*, vol. 12, no. 1, pp. 1–8, Dec. 2011.

[49] H. Hashemi, R. Pakzad, M. Khabazkhoob, M. H. Emamian, A. Yekta, and A. Fotouhi, "The distribution of vertical cup-to-disc ratio and its determinants in the Iranian adult population," *J. Current Ophthalmol.*, vol. 32, p. 226, Jun. 2019.

[50] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado, "Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 9, no. 2, pp. 283–293, 2014.

[51] J. Bernal *et al.*, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Comput. Med. Imag. Graph.*, vol. 43, pp. 99–111, Jul. 2015.

[52] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 630–644, Feb. 2015.

[53] D. Vázquez *et al.*, "A benchmark for endoluminal scene segmentation of colonoscopy images," *J. Healthcare Eng.*, vol. 2017, pp. 1–9, Jul. 2017.

[54] D. Jha *et al.*, "Kvasir-SEG: A segmented polyp dataset," in *Proc. Int. Conf. Multimedia Modeling*. Cham, Switzerland: Springer, 2020, pp. 451–462.