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 can accurately segment the optic disc (*OD*) and optic cup (*OC*) from color fundus images, 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 regularisation-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 due to the inherent geometric associations. Thirdly, we exploit the task-specific domain knowledge via the oval shape 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 is made available [1].**

***Index Terms*— Weakly and Semi-supervised Learning, Graph Convolutional Network, Optic Disc and Cup Segmentation**

Y. Meng, H. Zhang and Y. Zheng are with the Department of Eye and Vision Science, University of Liverpool, Liverpool, L7 8TX, United Kingdom.

Y. Zhao is with the Cixi Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, 315201, China.

D. Gao is with the School of Computing, University of Portsmouth, Portsmouth, PO1 3HE, United Kingdom.

G. Patri, S. Madhusudhan are with the St Paul's Eye Unit, Liverpool University Hospitals NHS Foundation Trust, Liverpool, L7 8XP, United Kingdom.

B. Hamill, T. Peto are with the School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, BT9 7BL, United Kingdom.

Corresponding author: Yitian Zhao (yitian.zhao@nimte.ac.cn); Yalin Zheng (yalin.zheng@liverpool.ac.uk).

[1]`https://github.com/smallmax00/Dual_Adaptive_Graph_Reasoning`

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 method [3] directly regresses the *vCDR* values from fundus images, however, it leads 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 by the ratio of vertical cup diameter divided by vertical disc diameter. Consequently, accurate segmentation of *OD & OC* is critical for the *vCDR* measurement, so as to glaucoma assessment. Recently, numerous deep learning-based segmentation models [1], [2], [4]–[10] 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 propose 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 exploiting the geometric associations and domain knowledge, we demonstrate the framework's effectiveness for the segmentation of *OD & OC* and also glaucoma assessment *w.r.t vCDR* estimation in colour fundus images.

Previous segmentation methods concentrated on learning the intensity features of the input image; 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

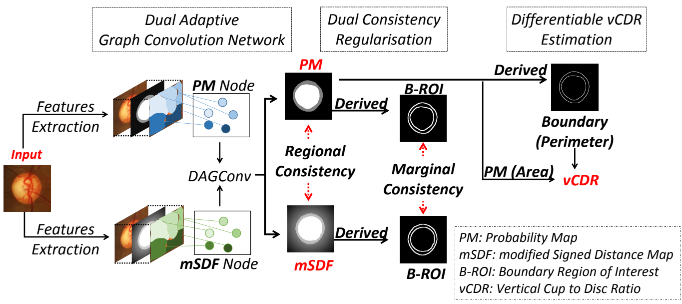

Fig. 1. The overview of the proposed network, where three major contributions such as *DAGCN*, dual consistency regularisation and differential *vCDR* estimation are shown.

and region, they ignore the inherent geometric association between these learned representations, which are critical for improving segmentation performance [9], [11]. To be more precise, segmentation probability map (*PM*) features emphasize the global homogeneity of pixel-level semantics and contextual information at the object level. Local boundary characteristics, such as boundary region of interest (*B-ROI*), describe the spatial variations on both sides of the boundary contour. The signed distance function (*SDF*) representations emphasise 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 semantic level akin to *PM*. Intuitively, the geometric associations between them appear 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 investigate how to accompany the feature complementary 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 construct the dual graph via initializing the adjacency matrix in a data-dependent way. The estimated vertex embeddings of *mSDF* and *PM* contribute to the dual adjacency matrices adaptively according to the geometric associations between them. We implement two matrices to quantify the distance and relations 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 overlooked the fact that the ultimate goal of such a learning pipeline is to estimate the *vCDR* to aid in glaucoma assessment. Undoubtedly, they also overlooked the potential supervision value of *vCDR* in *OD & OC* segmentation task. As a result, they wasted the underlying weak supervision label and required an offline post-processing step to calculate *vCDR* given the estimated diameters of the *OD & OC*. On the contrary, we exploit 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 differentiable *vCDR* estimating layer is proposed for the end-to-end training. Thus, our model does not need any offline post-process to generate *vCDR* but gains more weakly-supervised guidance without further annotations. A novel design like this ensures that the proposed model learns the well-defined goal and gains more supervision from the ground truth on both the region and boundary of objects. The underlying geometric associations between the region and boundary characteristics are usually underestimated by approaches to segmenting biomedical images, despite human graders' instinctive use of both domains. This paper demonstrates how to rationally leverage geometric knowledge and associations of *OD & OC* in terms of region and boundary on several aspects, such as adaptive graph construction and reasoning, semi-supervised consistency learning, and differentiable weakly-supervised *vCDR* estimation. 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:

- 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 propose 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.
- Extensive experiments on six large-scale datasets demonstrate that our method outperforms state-of-the-art semi-supervised approaches for segmentation of the optic disc and optic cup, and estimation of *vCDR* for glaucoma assessment in colour fundus images, respectively. Our model performs consistently well in segmentation and *vCDR* assessment, demonstrating its robustness and generalizability.

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 [6], [12]–[14] have considered segmentation as a dense pixel classification task. For example, the classic *U-net* [12] 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.* [14] 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 [15] 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 enhanced long-range reasoning abilities, the aforementioned methods achieved promising results in the *OD & OC* segmentation task. They are inefficient because stacking local cues does not always accurately represent long-range context relationships [9]. On the contrary, we benefit from the long-range information aggregating ability of the graph-based models to address this issue. On the other hand, in order to comprehend scenes or global contexts, these approaches must learn the object's position, boundary, and category using high-level semantic awareness and regional location information [16]. They, however, are primarily concerned with learning image intensity features, thus lacking the consideration of regional position information at the pixel level [17]. As a result, object boundary predictions are inaccurate. Differently, we explicitly consider the boundary and region correlations and geometric associations via semi-supervised paradigms such as region and boundary consistency regularisation and via weakly-supervised paradigms such as differentiable *vCDR* estimation from the *OD & OC* region and boundary predictions.

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 [7], [10]. Recent methods, such as [7]–[10], [18]–[21], have explicitly or implicitly taken into account the geometry dependency between the regions and boundaries of an object of interest in *OD & OC*. Specifically, *Zhang et al.* proposed *ET-Net* [18] for *OD & OC* segmentation, where an edge attention mechanism is proposed to explicitly emphasise the object boundary during the feature learning. *Meng et al.* proposed an aggregated hybrid network [9] to jointly learn the relationship between region and boundary of *OD & OC*, conducting an accurate boundary localization. On the other hand, *Luo et al.* [20] and *Xue et al.* [21] 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, emphasising 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. The aforementioned methods all benefited from the incorporation of geometry information of objects' boundary in the task of *OD & OC* segmentation tasks in a fully-supervised manner. In this work, we take another direction and exploit the geometric boundary information in a semi-supervised manner. Specifically, we explored the boundary consistency regularisation between the *PM* segmentation and the *mSDF* regression in a semi-supervised way. Such boundary consistency regularisation learning can boost our proposed model to learn accurate *OD & OC* boundaries and leverage geometric consistency with a large number of unlabeled data.

Other boundary-based methods [4], [11] 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*) [11] evaluation measurement, which quantifies boundary quality in segmentation tasks. *Wu, et al.* [4] proposed an oval shape constraint-based loss function to regularise 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 [20], [22], [23] achieved high-quality segmentation results. Numerous semi-supervised methods [24]–[28] have recently been developed that incorporate unlabeled data through unsupervised consistency regularisation. In general, there are majorly three different types of unsupervised consistency regularisations. Firstly, they introduced data-level of noises into the unlabeled samples and enforced consistency between the model predictions on the original and perturbed data [24], [29]. Secondly, the feature-level of perturbations are added into various output branches, such as multi-output channels [22], [27], or different levels of decoders [23], [25], to enforce consistency between the model predictions among different output branches. Thirdly, similar prediction distributions on the entire unlabeled dataset can be simply enforced using adversarial regularisation [26], [28], which also belongs to the data-level of regularisation. However, on the other hand, the consistency regularisation of task-level in semi-supervised learning has rarely been explored, until very recently in different computer vision tasks, such as crowd counting [30], 3D object detection [31], and 3D medical image segmentation [20]. 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 [32]. For example, [30], [20] and [31] all shared a similar idea that the dual task's outputs can be aligned into the same presentation space, then an unsupervised loss is applied to regularise the consistency. In this work, we also learn a dual-task level of geometric consistency on the *OD & OC* segmentation. Apart from that, we integrate the boundary quality into the task-level of consistency regularisation. Specifically, we estimate the boundary ROI mask from the *PM* segmentation and *mSDF* regress outputs, respectively. Then the supervised and unsupervised losses are applied to learn more accurate boundary segmentation results with the help of labeled and unlabeled data. Our ablation study results in Section V-A prove that such boundary regularisation can boost the boundary quality, which is essential in the medical image segmentation task.

On the other hand, weakly supervised methods [33]–[35] segmented images using image-level of labels [36], bounding boxes [34], [37], points [38], scribbles [35], [39], or image-level tags [33] rather than the pixel-by-pixel annotation, which alleviates the burden of annotation. They all focus on the data-driven learning-based way of general coarse labels . For example, given the image-level labels, *Wu et al.* [33] proposed an attention mechanism on the top of the class activation maps [40] to improve 3D brain lesion localization. The estimated lesion regions and normal tissues were then used to train the 3D brain lesion segmentation network. Similarly, [39] proposed a framework for weakly supervised cell segmentation based on scribble-level annotations. They used a combination of pseudo-labeling and label filtering to generate reliable labels with weak supervision. The pseudo labels were improved by leveraging the predictors' consistency over iterations. The aforementioned methods follow the same pipeline of generating pseudo segmentation labels from coarse ground truth, then train the model with the pseudo labels. Differently, for the first time, we integrate 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. Along with the estimated region and boundary predictions of *OD & OC*, we proposed a novel *vCDR* estimation layer in a differentiable way. As a result, our model can estimate the *vCDR* end-to-end on the basis of *OD & OC* segmentation. At the same time, the information gained from *vCDR* ground truth can weakly-supervise the segmentation process for both region and boundary of *OD & OC*.

D. Graph Reasoning in Segmentation

In recent years, graph-based models [9], [10], [41]–[44] have gained popularity for segmentation tasks due to their inherent ability to propagate information over long distances and update feature information. *Yao et al.* proposed a *GNN* network [45] to investigate the 3D geometrical relationship between vertices in conjunction with mesh representation in an organ segmentation task. With the nature of *GNN*, long-range shape information can be updated and passed among vertices to maintain a consistency constraint. Along the same lines, *Voxel2mesh* [46] learned a deformable mesh representation through *GNN* to propagate the voxel features along the edges of the graph model that had been created. Accordingly, a series of works [47]–[49] explored the surface-based segmentation pipeline because of the *GNN* models' feature extraction ability on Non-Euclidean data. Similarly, *Meng et al.* proposed *RBA-Net* [7] and *CABNet* [8] to regress the *OD & OC* boundaries by aggregated *CNN* and *GCN*, which learns the long-range features and directly regresses vertex coordinates in a Cartesian system. The methods described above made use of a *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 employs *GNN* to learn their inter-domain relationship. In particular, the geometric associations between them are exploited.

Additionally, methods such as [7], [8], [42]–[46] use *Laplacian* smoothing-based graph convolution [50], 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 initialised adjacency matrix [41], [43]–[46] or hand-crafted adjacency matrix [7], [8], [42], [50]. However, one may make a model to learn a specific long range context pattern [10], [51], which is less related to the input features, thus we regard them as data-independent non-adaptive graph convolution. Differently, as seen in previous works that the graph structure can be estimated with the similarity matrix from the input data [51], [52], we estimate the initial adjacency matrix in a data-dependent way. The constructed dual graph in this work has two distinct structures by adaptively learned from the input features of *PM* and *mSDF* features. In this way, our model is capable of adaptively learning an input-related long-range context pattern, which improves 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 Initialisation: A backbone network is used to extract the multi-level features. The deep- and shallow-layer features from different levels complement one another. For example, deep-layer features contain extensive semantic region information, while shallow-layer features retain sufficient spatial boundary information. Thus, for initializing the dual graph vertices, we used the feature aggregation module that is similar to our previous work [10] on relative deep-level and low-level features. 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 $64 \times 64 \times 2$. We then referred them to as the initialised *PM* node embeddings and *mSDF* node embeddings, respectively.

2) Classic Graph Convolution: We first revisit the classic graph convolution and their graph construction process *w.r.t* the adjacency matrix. Given a graph *G = (V, E)*, the normalised *Laplacian* matrix is defined as $L = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$, 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 diagonalised 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 $U = [u_1, ..., u_n]$, and $\Lambda = diag([\lambda_1, ..., \lambda_n]) \in \mathbb{R}^{N \times N}$ is a diagonal matrix with eigenvalues that are not negative. Due to the fact that

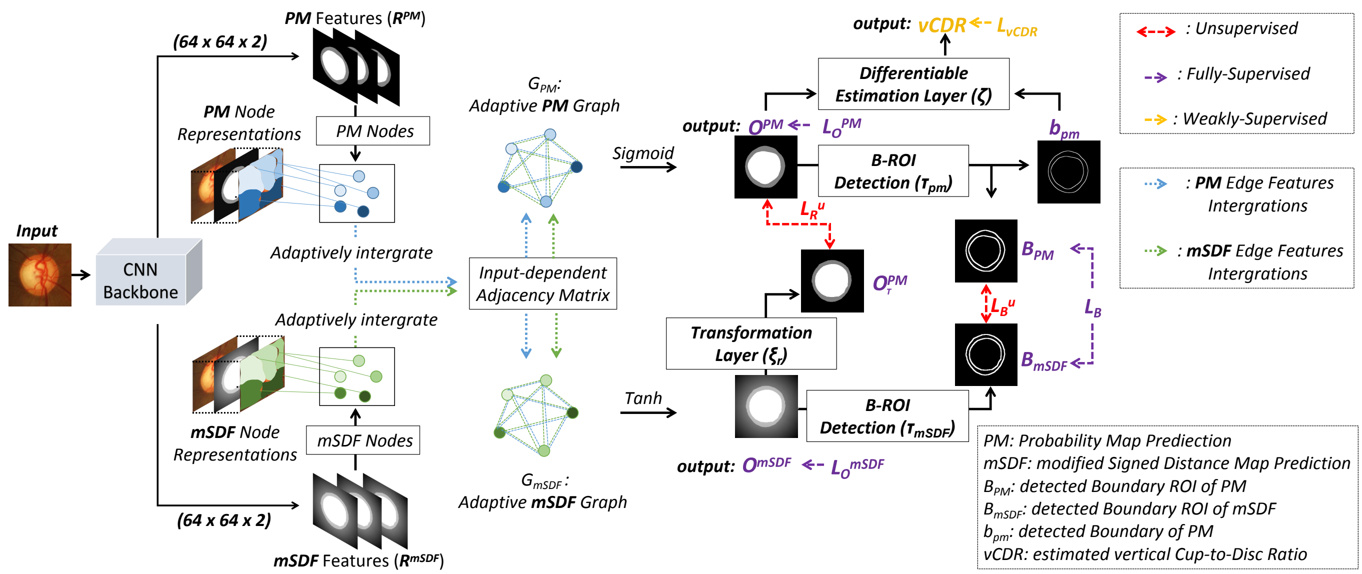

Fig. 2. Overview of the proposed *DAGCN* model (best viewed in color). O^{PM} and O^{mSDF} both have two channels to represent the output of *OC* and *OD* and we overlapped them for better visualisation. L_O^{PM}, L_O^{mSDF}, L_B are the supervised *PM*, *mSDF* and *B-ROI* loss functions; L_{vCDR} is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

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 [53] 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 [50], *Kipf et al.* 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 [51]. As a result, we refer 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 [10]. We extend it into a dual adaptive graph, perfectly fitting the proposed semi-supervised paradigm with dual consistency regularisation. Given the initialised *PM* nodes $R_{pm} \in \mathbb{R}^{N\times C}$ and *mSDF* nodes $R_{mSDF} \in \mathbb{R}^{N\times C}$, we construct 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 vertexes.

We illustrate G_{pm} as an example and elaborate the graph construction process as below. Firstly, we implement 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.

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

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,

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

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 initialise the input-dependent adaptive adjacency matrix $\bar{A}$ as:

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

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 linear embeddings (1×1 convolution); W_ψ and W_ϕ are learnable parameters. Secondly, we exploit the geometric association between *PM* and *mSDF* through integrating *mSDF* into the built *Laplacian* matrix $\tilde{L}$, which allows 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:

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

where $Conv(\cdot)$ is a 1×1 convolution layer. In this way, the semantic features of the object's foreground are emphasized by geometry-aware features of *mSDF*. As this is the case, the proposed adaptive graph convolution can 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:

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

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 normalised *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 [10], [51], 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 are used to generate the *PM* output (O^{PM}) and *mSDF* output (O^{mSDF}) respectively. We then apply *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 Regularisation of Semi-supervised Manner

1) Modified Signed Distance Function (mSDF): Given O^{PM} and O^{mSDF}, we explore the geometric association between them and build the unsupervised dual consistency regularisation 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 regularisation imposes 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:

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

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 $mSDF(x)$ represents 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. 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 [20], [21], [30], we propose 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 can obtain the transformed segmentation maps O_T^{PM}, for example, $O_T^{PM} = \xi_r(O^{mSDF})$. For all of the unlabeled input, we apply a *Dice* loss (L_{R^u}) between O^{PM} and O_T^{PM} to enforce the unsupervised regional consistency regularisation.

3) Marginal Consistency: We derive the spatial gradient of O^{PM} and O^{mSDF} as the estimated contours concerning the boundary-wise consistency. Previous studies [9], [11] 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 regularisations. Rather than focusing exclusively on the thin contour locations, we consider 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 :

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

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

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 refer 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} is applied to enforce the unsupervised marginal consistency regularisation 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 [7]–[9]. 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 117832 images with *vCDR* ground truth labeled. To address this issue, we take 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* [54], 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 approximate 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 derive 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} is approximately regarded as P_o. With Eq. 11 and Eq. 12, we can approximate a with A_o and P_o, such as:

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

where $|\cdot|$ is used to prevent sqrt from returning a negative value during the initial learning period. Given Eq. 13, we can 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, the *vCDR* estimation layer ζ can be 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.

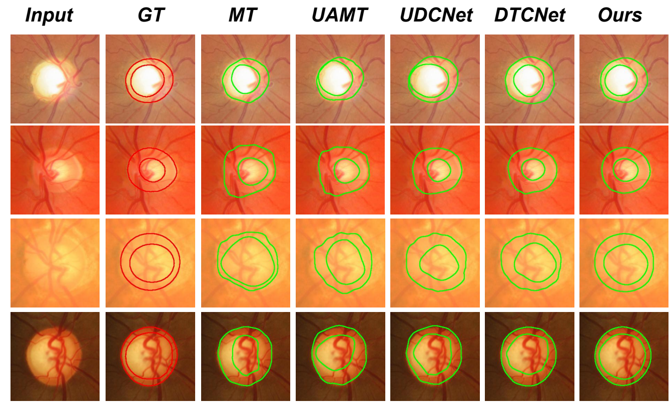

Fig. 3. Qualitative results of *OD & OC* segmentation in the *SEG* test dataset. We compare our model with *MT* [29], *UAMT* [24], *UDCNet* [23] and *DTCNet* [20]. Our method can produce more accurate segmentation results when compared with the ground truth (*GT*). Note that we plot the boundary of the segmentation mask on the input image to better visualise the segmentation comparison.

IV. Experiments

A. Datasets

SEG dataset: following the previous methods [9], [10], we pooled 2,068 images from five public available datasets (Refuge [1], Drishti-GS [56], ORIGA [57], RIGA [58], RIM-ONE [59]). These five datasets provide the fundus images and the ground truth masks, then we generate the corresponding ground truth of O^{mSDF}, B_{PM}, B_{mSDF} and *vCDR* with Eq. 7, 9, 10 and 14. Following the previous methods [9], [10], 613 fundus images were randomly selected as the test dataset, leaving the other 1,315 images for training and 140 images for validation.

UKBB dataset: The UK Biobank [2] 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 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, QUB, 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 is sufficient for measuring the *vCDR*. Then *vCDR* is calculated by dividing the measured diameter of the cup by the measured diameter of the disc in the long-axis or vertical direction. There are 117,832 fundus images with *vCDR* scalars are available, of which 38,421 are randomly selected as the weakly/semi-supervised training dataset, and the rest 79,411 are used as the test datasets.

[2]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 THE BRACKETS, RESPECTIVELY. WE COMPARE OUR MODEL WITH PREVIOUS FULLY-SUPERVISED STATE-OF-THE-ART METHODS BY RUNNING THEIR OPEN-SOURCE CODE. THE IMPLEMENTATION OF THE COMPARED SEMI-SUPERVISED STATE-OF-THE-ART WORKS IS MAINLY BASED ON AN OPEN-SOURCE CODEBASE [55].

Methods	SEG (OC)		SEG (OD)		*SEG (vCDR)*		*UKBB (vCDR)*	
	Dice (%)↑	*BIoU*(%)↑	*Dice* (%)↑	*BIoU*(%)↑	*MAE* ↓	*Corr* ↑	*MAE* ↓	*Corr* ↑
U-Net [12]	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 [9]	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 [7]	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 [29]	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 [24]	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 [22]	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 [20]	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 [23]	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 [26]	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)

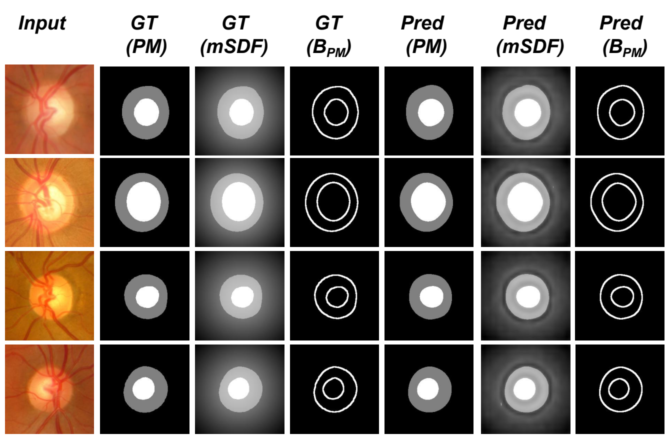

Fig. 4. Figure shows comparison between our model's prediction and the ground truth of the *SEG* test dataset. Our model produces consistent region and boundary predictions compared with the ground truth ***(GT)***.

B. Experimental Setting and Evaluation Metrics

We cropped the image of 256×256 pixels with the same way of [7], [9], [10]. To avoid over-fitting, we adopt the on-the-fly data augmentation strategy. Specifically, we randomly rotated and flipped the training dataset with a probability of 0.5. The rotation ranges from -20 to 20 degrees. We use stochastic gradient descent with a momentum of 0.9 to optimize the overall parameters. We trained the model 10000 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 [60] 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 use 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 [61] 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 degenerate solution during the initial training period. Because no meaningful prediction of the unlabeled data, as well as *vCDR*, are obtained.

We report Dice similarity score (*Dice*) as the region segmentation accuracy metrics; Boundary Intersection-over-Union (*BIoU*) [11] as the boundary segmentation metric; and Mean Absolute Error (*MAE*), Pearson's correlation coefficients [62] (*Corr*) as the *vCDR* estimation metric. 95% confidence intervals were generated by using 2000 sample bootstrapping. Note that the Pearson's correlation coefficients [62] are used to measure the linear association.

C. Performance Comparison and Analysis

In this section, we show qualitative (Fig. 3, Fig. 4) and quantitative (TABLE. I) results of the *OD & OC* segmentation and glaucoma assessment tasks.

***OD & OC* Segmentation** Fig. 3 and Fig. 4 illustrate 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. Specifically in TABLE. I, we present the results of fully-supervised methods on the upper half part, and the rest are semi-supervised methods. All of the fully-supervised methods are trained with 100% of the labeled *SEG* training dataset, and all of the semi-supervised methods are 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%)*). 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 regularisation based methods *MT* [29], *UAMT* [24] by 4.2 % and 2.9 %, outperforms feature-level regularisation based methods *URPC* [22] and *UDCNet* [23] by 2.0 % and 1.9 %, and outperforms adversarial regularisation based method *SASSNet* [26] by 2.3 %. On the other hand, with sufficient labeled and unlabeled data, *Ours (Semi-100%)* achieves 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* [9] by 2.7 %, 2.6% and 0.9 %.

Clinical Evaluation: Glaucoma Assessment TABLE. I illustrates the *vCDR* evaluation results on *SEG* and *UKBB* test dataset respectively. The *UKBB (vCDR)* has 79411 images, which is much larger than *SEG (vCDR)* (619 images). The performance on *UKBB (vCDR)* can reflect more realistic situation in the real-world *w.r.t* data distribution. Specifically, *Ours (semi)* achieved the best performance of 0.097 *MAE* and 0.463 *Corr*, which outperforms *DTCNet* [20] by 23.0 % and 53.3 %. Please note that, we utilised 38421 images of *UKBB* training dataset for weakly-supervised *OD & OC* segmentation. However, they also serve as fully supervision for *vCDR* estimation. Additionally, the direct *vCDR* regression-based method [3] with all *UKBB* train data achieves 0.074 *MAE* but only 0.240 *Corr* on the *UKBB* test data. Because the distribution of glaucoma patients and normal controls are unbalanced, thus such regression model tends to predict closer to the majority of the distribution.

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 results demonstrate 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.

Graph Reasoning In this section, we assess the efficacy of the proposed *DAGCN*. Notably, we maintain 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 [50] to reason about the relationships between the *PM* and the *mSDF*, respectively. Then, we investigate 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* [42], *DualGCN* [43], and *GloRe* module [44] 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 build two separate graphs for *PM* segmentation and *mSDF* regression individually, where there are no associations or geometric associations between the dual graph. Tab. IV shows that our model achieves more accurate and reliable results than [50] and outperforms the *SGR* [42], *DualGCN* [43], and *GloRe* [44] by 1.1 %, 0.9 % and 0.9 % mean *Dice* on the *SEG* test datasets.

Weakly/Semi-supervision We perform experiments to evaluate the effectiveness of the proposed dual consistency regularisation paradigm in semi-supervised learning and the proposed differentiable *vCDR* estimation module in a weakly-supervised manner. The results are shown in TBALE V.

TABLE II

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 [50]	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* [42]	87.2	83.6	96.5	87.7	0.105	0.430
w/ *DualGCN* [43]	87.5	83.7	96.6	88.1	0.104	0.427
w/ *GloRe* [44]	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 III

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**

Specifically, we evaluate the region-wise consistency loss, the boundary-wise consistency loss, and the *vCDR* estimation loss, respectively. We represent our model that is trained with only 5 % *SEG* training data as *Ours (Label-only)*. Firstly, we retain the same model structure and eliminate the *vCDR* estimation loss to focus on the dual consistency regularisation losses (*w/ Both*). Following that, we remove the region-wise unsupervised loss (*w/o* L_{R^u}), boundary-wise unsupervised loss (*w/o* L_{B^u}) respectively. Secondly, we remove both of the consistency losses and only apply the weakly-supervised *vCDR* estimation loss (*w/* L_{vCDR}). Then we add 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 are boosted. Tab. V demonstrates that the proposed unsupervised dual consistency losses and weakly supervised loss can 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.

Data Utilisation Efficiency In this section, we show more ablation study results on the data utilisation efficiency. In detail, we examine the performance of cutting-edge semi-supervised methods *UAMT* [24], *DTCNet* [20] and *Ours (Semi)* with different ratio of labeled and unlabeled images. We evaluate 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 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* [24], *DTCNet* [20] on both tasks under different labeled data utilisations. Primarily when less labeled data is used, *Ours (Semi)* suppresses the other two methods by a large margin. On the other hand, for unlabeled images, we vary 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)* achieves consistent superior performance over the *UAMT* [24], *DTCNet* [20] on both tasks under different unlabeled data utilisations, which indicates that our method effectively utilises the unlabeled data. When more unlabeled data is used, *Ours (Semi)* significantly outperforms the other two methods by a large margin.

B. Limitation and Future Work

Limitations. We design a weakly/semi-supervised learning paradigm specifically for the task of *OD* & *OC* segmentation and achieve promising results. The designed dual consistency regularisation mechanism can be widely applied to other semi-supervised medical image segmentation tasks, such as endoscopy polyps, ultrasound fetal head segmentation, *etc.*. However, it may not work as well for highly complex objects, such as curvilinear structures like vessels [?], [63], [64]. The primary reason for this is that vessels' region and boundary areas can be challenging to distinguish due to their complex topology and tortuosity. Thus, an inevitable perturbation will be included in the marginal and regional consistency regularisation, thus harming the semi-supervised performance.

Future works. Our proposed dual consistency regularisation mechanism could also be extended to tackle 3D image-based segmentation tasks. In 3D settings, we can regard the *B-ROI* as a surface region of interest and the *PM* as voxels prediction maps. Thus, the regional and marginal consistency will benefit the model from many unlabeled 3D images.

C. Conclusion

We propose a novel graph-based weakly/semi-supervised segmentation framework. The geometric associations between

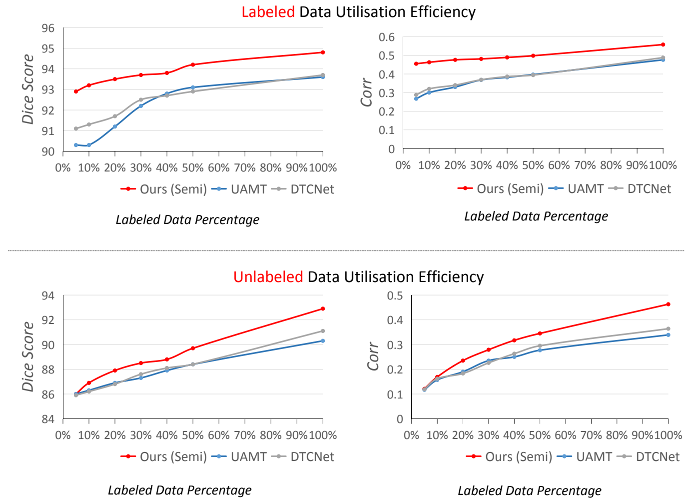

Fig. 5. The mean *OD* & *OC* segmentation performance of our semi-supervised approach with different ratio of labeled data. The performance is reported with *Dice* and *Corr*.

TABLE IV
ABLATION STUDY ON DIFFERENT GRAPH CONVOLUTIONS. THE PERFORMANCE IS REPORTED AS *Dice* (%), *BIoU* (%), *MAE* AND *Corr* ON THE 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 [50]	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* [42]	87.2	83.6	96.5	87.7	0.105	0.430
w/ *DualGCN* [43]	87.5	83.7	96.6	88.1	0.104	0.427
w/ *GloRe* [44]	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 THE 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**

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 regularisations, and weakly-supervised guidance. 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.

REFERENCES

[1] J. I. Orlando, H. Fu, J. B. Breda, K. van Keer, D. R. Bathula, A. Diaz-Pinto, R. Fang, P.-A. Heng, J. Kim, J. Lee *et al.*, "REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Medical Image Analysis*, vol. 59, p. 101570, 2020.

[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 Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1597–1605, 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 Journal of Biomedical and Health Informatics*, vol. 24, no. 4, pp. 1104–1113, 2019.

[4] J. Wu, K. Wang, Z. Shang, J. Xu, D. Ding, X. Li, and G. Yang, "Oval shape constraint based optic disc and cup segmentation in fundus photographs." in *BMVC*, 2019, p. 265.

[5] S. Wang, L. Yu, X. Yang, C.-W. Fu, and P.-A. Heng, "Patch-based output space adversarial learning for joint optic disc and cup segmentation," *IEEE transactions on medical imaging*, vol. 38, no. 11, pp. 2485–2495, 2019.

[6] P. Yin, Q. Wu, Y. Xu, H. Min, M. Yang, Y. Zhang, and M. Tan, "Pm-net: Pyramid multi-label network for joint optic disc and cup segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 129–137.

[7] Y. Meng, W. Meng, D. Gao, Y. Zhao, X. Yang, X. Huang, and Y. Zheng, "Regression of instance boundary by aggregated cnn and gcn," in *European Conference on Computer Vision*. Springer, 2020, pp. 190–207.

[8] Y. Meng, M. Wei, D. Gao, Y. Zhao, X. Yang, X. Huang, and Y. Zheng, "CNN-GCN aggregation enabled boundary regression for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020.

[9] Y. Meng, H. Zhang, Y. Zhao, X. Yang, Y. Qiao, I. J. MacCormick, X. Huang, and Y. Zheng, "Graph-based region and boundary aggregation for biomedical image segmentation," *IEEE Transactions on Medical Imaging*, 2021.

[10] Y. Meng, H. Zhang, D. Gao, Y. Zhao, X. Yang, X. Qian, X. Huang, and Y. Zheng, "Bi-gcn: Boundary-aware input-dependent graph convolution network for biomedical image segmentation," in *32nd British Machine Vision Conference: BMVC 2021*. British Machine Vision Association, 2021.

[11] B. Cheng, R. Girshick, P. Dollár, A. C. Berg, and A. Kirillov, "Boundary IoU: Improving object-centric image segmentation evaluation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2015, pp. 234–241.

[13] H. Fu, J. Cheng, Y. Xu, C. Zhang, D. W. K. Wong, J. Liu, and X. Cao, "Disc-aware ensemble network for glaucoma screening from fundus image," *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2493–2501, 2018.

[14] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: Context encoder network for 2d medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[15] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *ICLR*, 2016.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[17] Y. Chen, D. Zhao, L. Lv, and Q. Zhang, "Multi-task learning for dangerous object detection in autonomous driving," *Information Sciences*, vol. 432, pp. 559–571, 2018.

[18] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, and L. Shao, "ET-Net: A generic edge-attention guidance network for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 442–450.

[19] S. Wang, L. Yu, K. Li, X. Yang, C.-W. Fu, and P.-A. Heng, "Boundary and entropy-driven adversarial learning for fundus image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 102–110.

[20] X. Luo, J. Chen, T. Song, and G. Wang, "Semi-supervised medical image segmentation through dual-task consistency," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 10, 2021.

[21] Y. Xue, H. Tang, Z. Qiao, G. Gong, Y. Yin, Z. Qian, C. Huang, W. Fan, and X. Huang, "Shape-aware organ segmentation by predicting signed distance maps," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12 565–12 572.

[22] X. Luo, W. Liao, J. Chen, T. Song, Y. Chen, S. Zhang, N. Chen, G. Wang, and S. Zhang, "Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021.

[23] 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 *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021.

[24] 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 *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 605–613.

[25] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 674–12 684.

[26] S. Li, C. Zhang, and X. He, "Shape-aware semi-supervised 3d semantic segmentation for medical images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 552–561.

[27] H. Seo, L. Yu, H. Ren, X. Li, L. Shen, and L. Xing, "Deep neural network with consistency regularization of multi-output channels for improved tumor detection and delineation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3369–3378, 2021.

[28] J. Chen, H. Zhang, R. Mohiaddin, T. Wong, D. Firmin, J. Keegan, and G. Yang, "Adaptive hierarchical dual consistency for semi-supervised left atrium segmentation on cross-domain data," *IEEE Transactions on Medical Imaging*, 2021.

[29] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Advances in Neural Information Processing Systems*, 2017, pp. 1195–1204.

[30] Y. Meng, H. Zhang, Y. Zhao, X. Yang, X. Qian, X. Huang, and Y. Zheng, "Spatial uncertainty-aware semi-supervised crowd counting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 549–15 559.

[31] Y. Lu, S. Pirk, J. Dlabal, A. Brohan, A. Pasad, Z. Chen, V. Casser, A. Angelova, and A. Gordon, "Taskology: Utilizing task relations at scale," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8700–8709.

[32] A. R. Zamir, A. Sax, N. Cheerla, R. Suri, Z. Cao, J. Malik, and L. J. Guibas, "Robust learning through cross-task consistency," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 197–11 206.

[33] J. Lee, E. Kim, S. Lee, J. Lee, and S. Yoon, "Ficklenet: Weakly and semi-supervised semantic image segmentation using stochastic inference," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5267–5276.

[34] J. Lee, J. Yi, C. Shin, and S. Yoon, "Bbam: Bounding box attribution map for weakly supervised semantic and instance segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2643–2652.

[35] X. Liu, Q. Yuan, Y. Gao, K. He, S. Wang, X. Tang, J. Tang, and D. Shen, "Weakly supervised segmentation of covid19 infection with scribble annotation on ct images," *Pattern recognition*, vol. 122, p. 108341, 2022.

[36] K. Wu, B. Du, M. Luo, H. Wen, Y. Shen, and J. Feng, "Weakly supervised brain lesion segmentation via attentional representation learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 211–219.

[37] M. Rajchl, M. C. Lee, O. Oktay, K. Kamnitsas, J. Passerat-Palmbach, W. Bai, M. Damodaram, M. A. Rutherford, J. V. Hajnal, B. Kainz *et al.*, "Deepcut: Object segmentation from bounding box annotations using convolutional neural networks," *IEEE Transactions on Medical Imaging*, vol. 36, no. 2, pp. 674–683, 2016.

[38] I. Laradji, P. Rodriguez, O. Manas, K. Lensink, M. Law, L. Kurzman, W. Parker, D. Vazquez, and D. Nowrouzezahrai, "A weakly supervised consistency-based learning method for covid-19 segmentation in ct images," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 2453–2462.

[39] H. Lee and W.-K. Jeong, "Scribble2label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 14–23.

[40] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2921–2929.

[41] H. Huang, L. Lin, Y. Zhang, Y. Xu, J. Zheng, X. Mao, X. Qian, Z. Peng, J. Zhou, Y.-W. Chen *et al.*, "Graph-bas3net: Boundary-aware semi-supervised segmentation network with bilateral graph convolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 7386–7395.

[42] X. Liang, Z. Hu, H. Zhang, L. Lin, and E. P. Xing, "Symbolic graph reasoning meets convolutions," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018.

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

[44] Y. Chen, M. Rohrbach, Z. Yan, Y. Shuicheng, J. Feng, and Y. Kalantidis, "Graph-based global reasoning networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 433–442.

[45] J. Yao, J. Cai, D. Yang, D. Xu, and J. Huang, "Integrating 3D geometry of organ for improving medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 318–326.

[46] U. Wickramasinghe, E. Remelli, G. Knott, and P. Fua, "Voxel2mesh: 3D mesh model generation from volumetric data," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 299–308.

[47] T. Zhao, K. Cao, J. Yao, I. Nogues, L. Lu, L. Huang, J. Xiao, Z. Yin, and L. Zhang, "3d graph anatomy geometry-integrated network for pancreatic mass segmentation, diagnosis, and quantitative patient management," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 13 743–13 752.

[48] F. Kong, N. Wilson, and S. Shadden, "A deep-learning approach for direct whole-heart mesh reconstruction," *Medical Image Analysis*, vol. 74, p. 102222, 2021.

[49] X. Chen, N. Ravikumar, Y. Xia, R. Attar, A. Diaz-Pinto, S. K. Piechnik, S. Neubauer, S. E. Petersen, and A. F. Frangi, "Shape registration with learned deformations for 3d shape reconstruction from sparse and incomplete point clouds," *Medical Image Analysis*, vol. 74, p. 102228, 2021.

[50] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *ICLR*, 2017.

[51] X. Li, Y. Yang, Q. Zhao, T. Shen, Z. Lin, and H. Liu, "Spatial pyramid based graph reasoning for semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8950–8959.

[52] Y. Li and A. Gupta, "Beyond grids: Learning graph representations for visual recognition," in *Advances in Neural Information Processing Systems*, 2018, pp. 9225–9235.

[53] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Advances in Neural Information Processing Systems*, 2016, pp. 3844–3852.

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

[55] X. Luo, "Ssl4mis. https://github.com/hilab-git/ssl4mis," 2020.

[56] J. Sivaswamy, S. Krishnadas, G. D. Joshi, M. Jain, and A. U. S. Tabish, "Drishti-GS: Retinal image dataset for optic nerve head (ONH) segmentation," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2014, pp. 53–56.

[57] Z. Zhang, F. S. Yin, J. Liu, W. K. Wong, N. M. Tan, B. H. Lee, J. Cheng, and T. Y. Wong, "ORIGA-light: An online retinal fundus image database for glaucoma analysis and research," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 2010, pp. 3065–3068.

[58] A. Almazroa, S. Alodhayb, E. Osman, E. Ramadan, M. Hummadi, M. Dlaim, M. Alkatee, K. Raahemifar, and V. Lakshminarayanan, "Retinal fundus images for glaucoma analysis: the RIGA dataset," in *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*, vol. 10579. International Society for Optics and Photonics, 2018, p. 105790B.

[59] F. Fumero, S. Alayón, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "RIM-ONE: An open retinal image database for optic nerve evaluation," in *24th International Symposium on Computer-based Medical Systems (CBMS)*. IEEE, 2011, pp. 1–6.

[60] S. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. H. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[61] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," in *5th International Conference on Learning Representations*, 2017.

[62] M. M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi Medical Journal*, vol. 24, no. 3, 2012.

[63] Y. Zhao, L. Rada, K. Chen, S. P. Harding, and Y. Zheng, "Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images," *IEEE Transactions on Medical Imaging*, vol. 34, no. 9, pp. 1797–1807, 2015.

[64] Y. Ma, H. Hao, J. Xie, H. Fu, J. Zhang, J. Yang, Z. Wang, J. Liu, Y. Zheng, and Y. Zhao, "ROSE: A retinal OCT-angiography vessel segmentation dataset and new model," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 928–939, 2020.