Graph-BAS^{3}Net: Boundary-Aware Semi-Supervised Segmentation Network with Bilateral Graph Convolution

Huimin Huang[1], *Lanfen Lin[1], Yue Zhang[1], Yingying Xu[2,1], Jing Zheng[3], XiongWei Mao[4], Xiaohan Qian[3], Zhiyi Peng[3], *Jianying Zhou[3], *Yen-Wei Chen[5,2,1], Ruofeng Tong[1,2]
[1]Zhejiang University, [2]Zhejiang Lab, [3]The First Affiliated Hospital,
[4]Zhejiang University Hospital, [5]Ritsumeikan University

Abstract

Semi-supervised learning (SSL) algorithms have attracted much attentions in medical image segmentation by leveraging unlabeled data, which challenge in acquiring massive pixel-wise annotated samples. However, most of the existing SSLs neglected the geometric shape constraint in object, leading to unsatisfactory boundary and non-smooth of object. In this paper, we propose a novel boundary-aware semi-supervised medical image segmentation network, named Graph-BAS^{3}Net, which incorporates the boundary information and learns duality constraints between semantics and geometrics in the graph domain. Specifically, the proposed method consists of two components: a multi-task learning framework BAS^{3}Net and a graph-based cross-task module BGCM. The BAS^{3}Net improves the existing GAN-based SSL by adding a boundary detection task, which encodes richer features of object shape and surface. Moreover, the BGCM further explores the co-occurrence relations between the semantics segmentation and boundary detection task, so that the network learns stronger semantic and geometric correspondences from both labeled and unlabeled data. Experimental results on the LiTS dataset and COVID-19 dataset confirm that our proposed Graph-BAS^{3}Net outperforms the state-of-the-art methods in semi-supervised segmentation task.

1. Introduction

Accurate medical image segmentation is an essential prerequisite for many clinical applications [19]. Recently, a variety of convolutional neural networks (CNNs) have been developed for segmentation tasks. Though these methods achieved satisfactory results, they needed massive pixel-wise annotated samples and to be trained in fully supervision. In the medical field, however, sufficient labeled data is unavailable as the manual annotation is costly and time-consuming. To address this issue, semi-supervised learning (SSL) has been introduced, which uses both labeled data and arbitrary amounts of unlabeled data in training.

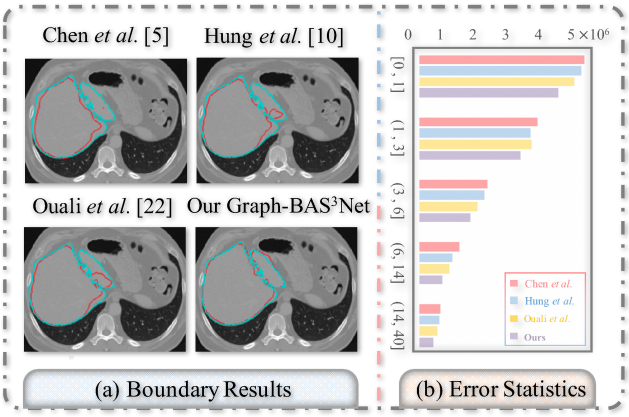

Figure 1. (a) shows the boundary results of four methods on LiTS dataset with 10% labeled data, where **cyan edges** are ground truth boundaries; while **red edges** are predictions. (b) shows the number of error pixels (horizontal axis) *vs.* their Euclidean distances (vertical axis) to the boundaries on four methods. We can see that pixels with larger distance tend to be well-classified, while pixels with smaller distance (boundary pixels) have larger errors.

Recent efforts in SSL have been focused on incorporating unlabeled data into training, which can be categorized into following groups: self-training [2, 5], co-training [22,29,35], GAN-based methods [10,14,21,33,34] and self-ensembling (Π model [12, 16] and Mean-Teacher model [7, 25, 31]). For example, Chen *et al.* [5] proposed a self-training-based SSL that alternately updated the segmentation results of unlabeled data; while Ouali *et al.* [22] achieved co-training by exploiting cross-consistency, which learned the generalized feature from the unlabeled data. Hung *et al.* [10] designed a GAN-based SSL that enforced the segmentation of unlabeled data to be similar to the labeled ones. Tarvainen *et al.* [25] proposed a Mean-Teacher model to guide the student network learning. However, they often ignored the geometric information and/or the inher-

*Corresponding Authors: Lanfen Lin (llf@zju.edu.cn), Jianying Zhou (zjyhz@zju.edu.cn), Yen-Wei Chen (chen@is.ritsumei.ac.jp)

ent semantic and geometric correspondences, which lead to unsatisfactory boundary and non-smooth of object since the ambiguity of structure boundary and heterogeneous texture (see Fig.1(a)). As shown in Fig.1(b), the number of error pixels significantly decrease with larger distances to the boundary. In other words, the boundary accuracy is crucial to the final semantic segmentation, yet its importance is often overlooked in previous methods.

Therefore, in this work, we propose a novel Graph-based boundary-aware semi-supervised segmentation network (Graph-BAS^{3}Net) to address the aforementioned limitations. Our main idea is to incorporate the boundary representation in the network, and learn the duality constraints between semantics and boundaries in the graph domain. The Graph-BAS^{3}Net comprises of two components: (**i**) a **B**oundary-**A**ware **S**emi-**S**upervised **S**egmentation **Net**work (BAS^{3}Net) that mitigates the blurry boundary problem by incorporating a boundary detection task into the GAN-based segmentation framework. (**ii**) a **B**ilateral **G**raph **C**onvolution **M**odule (BGCM) that models the duality constraints between tasks and captures long-range dependencies over non-local regions. The design rationale of the above two components is elaborated as below.

Firstly, considering that the boundary surrounded the mask encodes richer features of object shape and surface, our generator of BAS^{3}Net jointly predicts semantic segmentation and object boundary with a shared encoder. The shared encoder encourages the network to extract common features for different tasks, thus making the network more compact. To utilize the unlabeled data and learn more detailed edge information, we then introduce a discriminator to distinguish the predicted semantic segmentation map and boundary detection results ('fake') from ground truth labels ('real') for semi-supervised learning. In this multi-task learning way, the semantic segmentation provides the smoothness and continuity constraints; while the boundary detection enforces a global shape and geometric constraints.

Secondly, as there exists duality constraints between two tasks, it is obvious and remarkable that semantic segmentation and boundary detection can benefit each other by mutual interaction and promotion to boost the overall performance of semi-supervised segmentation. Based on this, we design the BGCM to explore co-occurrence relations and diffuse information between the semantics segmentation and boundary detection task. To establish the relationships effectively, we utilize graph convolution [6, 8, 11, 13, 15, 17, 26, 28, 32] to mine the intra-task and inter-task relations within and between two tasks. Specifically, the intra-task reasoning can capture long-range dependencies over non-local regions and refine the visual features in separate tasks; while the inter-task reasoning can model the similar latent representations between tasks and enable information propagation in a bidirectional way. In this way, our Graph-BAS^{3}Net, that consists of the backbone BAS^{3}Net and the cross-task module BGCM, can be aware of the reciprocal relations between semantics segmentation and boundary detection and exhibits superior performance.

The major contributions of this work are four-fold: (**i**) We propose a Graph-BAS^{3}Net to enforce semantic and geometric constraints in semi-supervised medical image segmentation. It combines a multi-task learning framework BAS^{3}Net and a graph-based cross-task module BGCM reasoning between tasks. (**ii**) We devise a BAS^{3}Net that jointly predict the semantic segmentation and object boundary, which improves the segmentation performance of the generator and further introduces boundary information to the discriminator. (**iii**) We propose a BGCM to enforce duality constraints between semantics and boundaries by using bilateral graph convolution, which globally mines the intra-task and inter-task relations.(**iv**) We conduct extensive experiments on a typical liver datasets and a more challenging COVID-19 dataset, where the proposed Graph-BAS^{3}Net outperforms the state-of-the-art methods.

2. Related Work

Semantic Segmentation. Current state-of-the-art methods for semantic segmentation are based on the rapid development of CNNs, e.g., FCNs [18], SegNet [1], and a series of UNet variations [9, 23, 36] designed for medical image segmentation. However, to achieve high robustness and segmentation accuracy, fully supervised segmentation approaches require massive pixel-level annotated data, which is often expensive and complex to collect.

GAN-based SSL. SSL approaches are developed to reduce the workload of ground truth labeling, among which generative adversarial networks (GAN) based SSLs yield progressive performance. Hung *et al*. [10] proposed a pioneering adversarial learning segmentation network (ALS-Net). It regarded the segmentation network as the generator, while the objective of the discriminator is to differentiate ground truths from segmentation probability maps to obtain a confidence map. Nie *et al*. [21] extended ALS-Net to segment the pelvic images with a focal loss based attention mechanism. Zheng *et al*. [34] proposed a deep atlas prior and incorporated it into ALS-Net to further improve the performance of liver segmentation. However, these GAN-based SSLs fail to produce trustworthy pseudo label in the organ boundary, especially when lacking sufficient labeled data.

Visual Reasoning via Graph Convolutional Network. Recently, graph convolution [11] has been incorporated into computer vision tasks to capture long-range dependencies, which projects the feature into a non-coordinate space. For instance, Graph Convolutional Unit (GCU) [17] assigns pixels with similar features to the same vertex via nonlinear feature encoding method; Global Reasoning (GloRe) unit

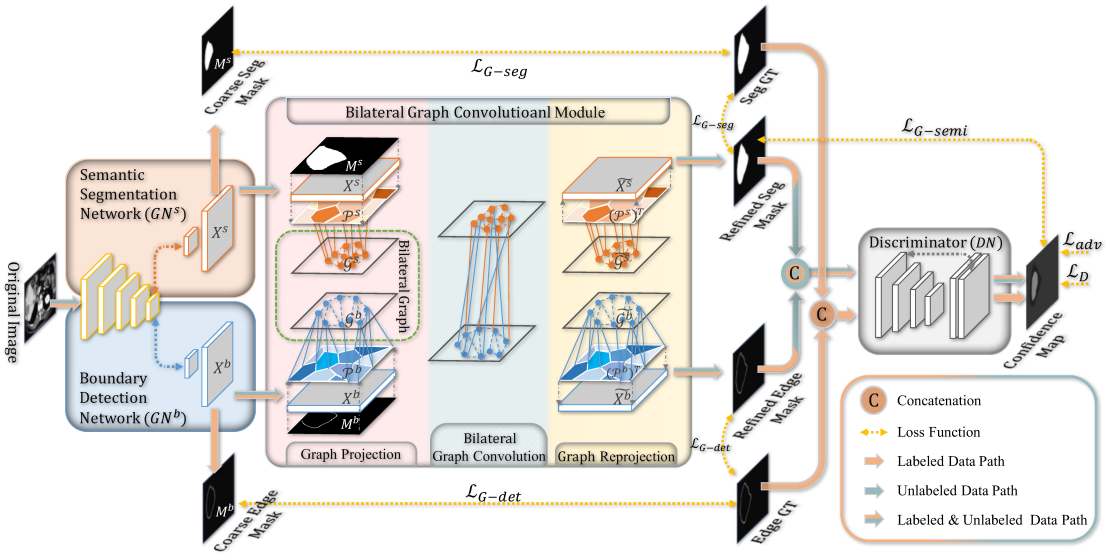

Figure 2. An overview of our Graph-BAS^{3}Net. First, the input image passes two networks, Semantic Segmentation Networks and Boundary Detection Networks with a shared encoder. This generates coarse results supervised by the ground truth for labeled data. BGCM then takes semantic segmentation and boundary detection features as the inputs and outputs the enhanced features after Graph Projection, Bilateral Graph Convolution and Graph Reprojection. This results in the refined semantic segmentation map and boundary detection result. The refined results are then concatenated and fed into the discriminator to obtain a confidence map.

[6] constructed a fully connected graph via channel-wise similarity. Te *et al*. [26] further introduced the edge attention into the feature projection, which emphasizes features of edge pixels. Considering the graph interaction, Wu *et al*. [28] then mined the relations within and between the foreground objects and background stuff classes for Panoptic Segmentation. Different from these approaches, we introduce the graph structure into a multi-task framework for semi-supervised medical image segmentation to globally model the mutual relations between multi-tasks.

3. Graph-BAS^{3}Net

In this section, we first provide an overview of our method and then present each component in detail. As seen in Fig.2, Graph-BAS^{3}Net consists of two parts: (**i**) BAS^{3}Net that improves the existing GAN-based SSL by adding a boundary-aware task and works as a backbone network; (**ii**) BGCM that interacts between tasks to further improve the segmentation accuracy.

The BAS^{3}Net is composed of three networks: The semantic segmentation network (GN^s), boundary detection network (GN^b), and discriminator network (DN). During training, GN^s and GN^b learn the feature representations X^s and X^b by focusing on the semantics and boundaries, respectively. To explore the mutual information between two tasks, our BGCM firstly projects X^s and X^b in the coordinate domain into the fully-connected graphs $\mathcal{G}^s$ and $\mathcal{G}^b$ in the graph domain, where relational reasoning can be efficiently computed. After reasoning, relation-aware features are reversed back to the coordinate domain for further prediction. Then the refined results are concatenated and delivered to DN, which differentiates the prediction from the ground truth. The networks GN^s, GN^b, and DN are designed to work in an adversarial fashion, tackling the problems of insufficient labeled data and blurry boundaries.

3.1. BAS^{3}Net

Formally, let $D_L = \{(I_l, Y_l)\}_{l=1}^{N}$ denotes the labeled set and $D_U = \{I_u\}_{u=N+1}^{N+M}$ denotes the unlabeled set, where $Y_l = \{Y_l^s, Y_l^b\}$ is the segmentation ground truth Y_l^s and boundary ground truth Y_l^b extracted from the semantic ground truth using the Roberts operator [20].

Multi-task Generator Network. As seen in Fig.2, the generator of BAS^{3}Net contains two networks, i.e., GN^s for semantic segmentation and GN^b for boundary detection, which share the same encoder but have task-specific decoders. GN^s is trained with the pixel-wise semantic annotations and yields coarse segmentation mask $GN^s(I)$, while GN^b is optimized to predict object edges $GN^b(I)$.

Concretely, we adopt DeepLabV2 [4] as the encoder shared by GN^s and GN^b. We also remove the last classification layer and modify the stride of the last two convolution layers from 2 to 1. This reduces the resolution of the output feature maps to $1/8$ of the size of the input image size. GN^s and GN^b adopt the same decoder architecture but do not share the parameters. To enlarge the receptive fields, we apply the dilated convolution [30] in $conv4$ and $conv5$ layers with strides of 2 and 4, respectively. We employ the Atrous Spatial Pyramid Pooling (ASPP) [4] to

fuse the feature having different receptive scale. Finally, we apply an up-sampling layer to transfer feature map from $H/8 \times W/8 \times 64$ to $H \times W \times 64$ and apply a convolution with 1×1 kernel size as the classifier.

Discriminator Network. The discriminator of BAS^{3}Net is employed to distinguish predicted segmentations from manually-annotated labels. We further introduce the boundary information into the discriminator network (DN) via combining the boundary-aware detection result $GN^b(I)$ with the semantic-aware segmentation map $GN^s(I)$.

Specifically, it consists of four convolution layers with kernel of 3×3, channel numbers of $\{16, 32, 64, 128\}$, and stride of 2. A deconvolution layer is further added to the last layer to rescale the output to the size of the input map. To maintain more detailed information, we concatenate the result with the first encoder layer which has the size of the input map. Then, a convolution with 1×1 kernel size is applied as the final classifier.

3.2. Bilateral Graph Convolutional Module

Formally, we define a graph as $\mathcal{G} = (\mathcal{N}, \mathcal{A}, \mathcal{H})$, where $\mathcal{N}$ is a set of nodes, and $|\mathcal{N}|$ denotes the number of nodes. The adjacent matrix $\mathcal{A} \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{N}|}$ describes the edge weights and $\mathcal{H} \in \mathbb{R}^{|\mathcal{N}| \times K}$ is the feature matrix of the graph. Our BGCM consists of three operations: Graph Projection, Bilateral Graph Reasoning and Graph Reprojection. Specifically, Graph Projection is the first step that maps the feature map X in the coordinate domain onto a set of node features $\mathcal{H}$ in the graph domain; while Graph Reprojection is the final step that finally reverse the updated graph features $\widetilde{\mathcal{H}}$ back to $\widetilde{X}$. Bilateral Graph Convolution is the critical step that models the intra-task and inter-task relations and diffuses information between tasks.

3.2.1 Graph Projection and Reprojection

We adopt the same strategies to project and reproject semantic-aware graph $\mathcal{G}^s$ and boundary-aware graph $\mathcal{G}^b$. For simplicity, we take $\mathcal{G}^s$ as an example. As seen in Fig.3, an attention mechanism is applied into the projection, which monitors the object parts through performing a dot multiplication $\odot$ between coarse segmentation mask M^s $(GN^s(I))$ and X^s. The dot multiplication assigns a higher weight to the features of pixels which belong to the object, and suppresses the non-object regions. In practice, we use a convolution, $\phi^s(\cdot)$, with a kernel size of 1×1 to reduce the dimension of X^s from C to L, which enhances the capacity of the projection process. The next step is to perform an average pooling, $AvgPool(\cdot)$,with stride s to obtain the anchors of the vertices. These anchors represent the centers of each region of pixels. We adopt the multiplication, $\otimes$, of $\phi^s(X^s)$ and anchors to capture the similarity between anchors and each pixel. The range of the projection matrix, p^s, is constrained to $(0, 1)$ by applying a softmax function:

$$p^s = softmax(AvgPool(\phi^s(X^s) \odot M^s) \otimes \phi^s(X^s)^T) \quad (1)$$

Based on the projection matrix p^s, the feature map X^s is then mapped into the graph domain as the following: $\mathcal{H}^s = p^s \otimes \theta^s(X^s)$, where $\theta^s(\cdot)$ is a convolution operation with 1×1 kernel to obtain the features of dimension reduction, leading to $\theta^s(X^s) \in \mathbb{R}^{HW \times K}$. The projection process is formulated as a linear combination, which aggregates the pixels with similar features as an anchor to one node. This results in a semantic-aware graph feature, $\mathcal{H}^s \in \mathbb{R}^{HW/s^2 \times K}$. Similarly, we can obtain a boundary-aware graph feature $\mathcal{H}^b$.Note that the down-sampling rate, δ, in $AvgPool(\cdot)$ can be different from the one (stride s) in constructing $\mathcal{G}^s$, resulting in $|\mathcal{N}^s| \neq |\mathcal{N}^b|$.

After the reasoning, we adopt linear reprojection given by $\widehat{X^s} = (p^s)^T \widetilde{\mathcal{H}^s}$. But the small-sized $\widehat{X^s} \in \mathbb{R}^{HW \times K}$ is inconsistent with the original feature map $X^s \in \mathbb{R}^{HW \times C}$. Thus, we attach a 1×1 convolution layer $\psi^s(\cdot)$ for dimension expansion, so that the output can seamlessly match the input to form a residual path:

$$\widetilde{X^s} = X^s + \psi^s((p^s)^T \widetilde{\mathcal{H}^s}) \quad (2)$$

3.2.2 Bilateral Graph Reasoning

Given $\mathcal{G}^s$ and $\mathcal{G}^b$, we adopt the graph convolution to diffuse information on the graphs. In this work, we use a similar approach as in [11] to define graph convolution. We first define the augmentation form of a bilateral graph as:

$$\mathcal{H} = \left[(\mathcal{H}^s)^T, (\mathcal{H}^b)^T\right]^T, \mathcal{W} = \left[(\mathcal{W}^s)^T, (\mathcal{W}^b)^T\right]^T \quad (3)$$

where $\mathcal{H} \in \mathbb{R}^{(|\mathcal{N}^s| + |\mathcal{N}^b|) \times K}$, $\mathcal{W} \in \mathbb{R}^{2K \times K'}$ is the augmented form of bilateral node feature and weight matrix. $\mathcal{W}^s$ and $\mathcal{W}^b \in \mathbb{R}^{K \times K'}$ are two trainable weight matrixes that alter the node dimension of $\mathcal{H}^s$ and $\mathcal{H}^b$, respectively.

Instead of performing on a single graph as in [11], our bilateral graph convolution captures the co-occurrence relations over two graphs via the intra-graph and inter-graph reasoning. Specifically, the intra-graph reasoning models the non-local dependencies in each graph. This is performed on the semantic-to-semantic edges ($\mathcal{A}^{s \to s}$) and boundary-to-boundary edges ($\mathcal{A}^{b \to b}$). The inter-graph reasoning explores the mutual relations between the graphs hence it is applied to the semantic-to-boundary edges ($\mathcal{A}^{s \to b}$) and boundary-to-semantic edges ($\mathcal{A}^{b \to s}$). Based on the above, the adjacent matrix $\mathcal{A}$ in this work is a combination of intra-graph matrix ($\mathcal{A}^{intra}$) and inter-graph matrix ($\mathcal{A}^{inter}$) which is formulated as:

$$\mathcal{A} = \mathcal{A}^{intra} + \mathcal{A}^{inter}, \quad (4)$$

$$\mathcal{A}^{intra} = \begin{pmatrix} \mathcal{A}^{s \to s} & 0 \\ 0 & \mathcal{A}^{b \to b} \end{pmatrix} \mathcal{A}^{inter} = \begin{pmatrix} 0 & \mathcal{A}^{b \to s} \\ \mathcal{A}^{s \to b} & 0 \end{pmatrix} \quad (5)$$

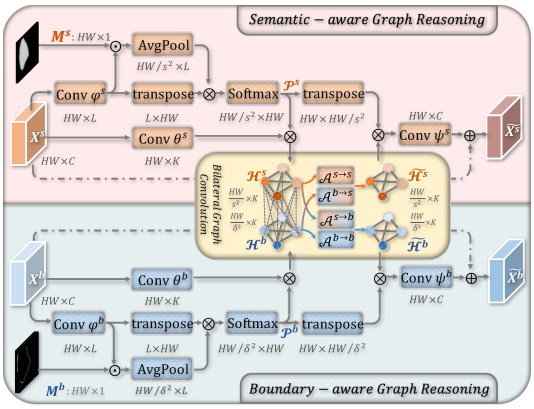

Figure 3. Architecture of the Bilateral Graph Convolution Module.

where $\mathcal{A}^{s\to b} = \left\{a_{ij}^{s\to b}\right\} \in \mathbb{R}^{|\mathcal{N}^s|\times|\mathcal{N}^b|}$ assembles the correlation weight from j-th node of $\mathcal{G}^s$ to the i-th node of $\mathcal{G}^b$, and $\mathcal{A}^{s\to s}$,$\mathcal{A}^{b\to b}$,$\mathcal{A}^{b\to s}$are explained similarly. The coefficients, a_{ij}, which indicate the importance of node j for node i, are obtained for every neighboring node pair using an attention mechanism [27]:

$$a_{ij} = \frac{exp\left(\delta\left(W\left[h_i \,||\, h_j\right]\right)\right)}{\sum_{z\in\mathcal{N}_i} exp\left(\delta\left(W\left[h_i \,||\, h_z\right]\right)\right)} \tag{6}$$

where the attention function is a single-layer neural network parameterized by a weight vector $W \in \mathbb{R}^{2K}$. $||$ is the concatenation and δ is LeakyReLU nonlinear. $\mathcal{N}_i$ is the neighborhood of node i, which contains all nodes in our fully-connected graph. Note that the graphs constructed here is a directional graph, as weight vector W is different in learning a_{ij} and a_{ji}. With the normalized adjacent matrix $\mathcal{A}$, augmented bilateral node feature $\mathcal{H}$ and weight matrix $\mathcal{W}$, a single graph convolution layer is formulated as:

$$\widetilde{\mathcal{H}} = \mathcal{F}\left(\mathcal{H} \,||\, \sigma\left(\mathcal{A}\left(\mathcal{H}\otimes\mathcal{W}\right)\right)\right) \tag{7}$$

$$\mathcal{H}\otimes\mathcal{W} = \left[\left(\mathcal{H}^s\mathcal{W}^s\right)^T, \left(\mathcal{H}^b\mathcal{W}^b\right)^T\right]^T \tag{8}$$

where $\mathcal{F}(\cdot)$ fuses the original features and updated features, which is realized by a convolution with kernel 1×1 .

3.3. Loss Functions

In our approach, we optimize the Graph-BAS^{3}Net with five losses:$\mathcal{L}_D$, $\mathcal{L}_{G-adv}$, $\mathcal{L}_{G-seg}$, $\mathcal{L}_{G-det}$, $\mathcal{L}_{G-semi}$.

$\mathcal{L}_D$ is the binary cross-entropy loss of the discriminator network, which is utilized to distinguish the ground truths and the segmentation maps:

$$\begin{aligned}\mathcal{L}_D = -\sum_{H,W}(1-y_l)\,log\left(1-DN\left(\widetilde{GN^s}(I_l)\,||\,\widetilde{GN^b}(I_l)\right)\right)\\ +y_l log\left(DN\left(Y_l^s\,||\,Y_l^b\right)\right)\end{aligned} \tag{9}$$

where $||$ is the concatenation operation. $y_l = 0$ if the input is drawn from the refined semantic segmentation map $\widetilde{GN^s}(I_l)$ and the refined boundary detection result $\widetilde{GN^b}(I_l)$, and $y_l = 1$ if the input is combined with their corresponding ground truth labels Y_l^s and Y_l^b. Note that the unlabeled data is not included in the calculation of $\mathcal{L}_D$.

Furthermore, $\mathcal{L}_{G-adv}$ is the adversarial loss term, which improves the generator and fools the discriminator via maximizing the probability maps from the generator being considered as the ground truth labels. Hence, it enforces higher-order consistency between the automatic segmentation and the ground-truths and is defined as:

$$\mathcal{L}_{G-adv} = -\sum_{H,W} log\left(DN\left(\widetilde{GN^s}(I_l)\,||\,\widetilde{GN^b}(I_l)\right)\right) \tag{10}$$

In our approach, $\mathcal{L}_{G-seg}$ is the segmentation loss of the labeled data; while $\mathcal{L}_{G-det}$ is defined as the detection loss. Given an input image I_l, one-hot encoded ground truth Y_l^s and Y_l^b, the binary cross-entropy loss is calculated by:

$$\mathcal{L}_{G-seg} = -\sum_{H,W}\left(Y_l^s log\left(GN^s(I_l)\right)\right) + Y_l^s log\left(\widetilde{GN^s}(I_l)\right) \tag{11}$$

$$\mathcal{L}_{G-det} = -\sum_{H,W}\left(Y_l^b log\left(GN^b(I_l)\right)\right) + Y_l^b log\left(\widetilde{GN^b}(I_l)\right) \tag{12}$$

Where $GN^s(I_l)$ and $GN^b(I_l)$ are the coarse segmentation mask and coarse edge mask, respectively.

Moreover, $\mathcal{L}_{G-semi}$ is the semi-supervised loss of the unlabeled data I_u. Benefiting from the confidence map generated from the discriminator, we select the partial high confidence pixels from the masked segmentation prediction, which can be considered as ground truth for unlabeled data. This "self-taught" process is formulated as:

$$\begin{aligned}\mathcal{L}_{G-semi} = -\sum_{H,W}\zeta\left(DN\left(\widetilde{GN^s}(I_u)\,||\,\widetilde{GN^b}(I_u)\right) > T_{semi}\right)\\ \cdot\widehat{Y_u^s} log\left(\widetilde{GN^s}(I_u)\right)\end{aligned} \tag{13}$$

where $\zeta(\cdot)$ is an indicator function, and T_{semi} is the threshold that controls the sensitivity of the self-taught process. $\widehat{Y_u^s} = argmax(\widetilde{GN^s}(I_u))$ is a binarized segmentation prediction. Combined with the self-taught target $\widehat{Y_u^s}$, $\mathcal{L}_{G-semi}$ can be viewed as a masked binary entropy loss.

The final loss of the generator $\mathcal{L}_G$. is the combination of $\mathcal{L}_{G-adv}$, $\mathcal{L}_{G-seg}$, $\mathcal{L}_{G-det}$, and $\mathcal{L}_{G-semi}$:

$$\mathcal{L}_G = \mathcal{L}_{G-seg} + \lambda_{det}\mathcal{L}_{G-det} + \lambda_{semi}\mathcal{L}_{G-semi} + \lambda_{adv}\mathcal{L}_{G-adv} \tag{14}$$

where the λ_{det}, λ_{semi} and λ_{adv} are the constraints for balancing the multi-task training.

4. Experiments and Results

4.1. Datasets

We conducted experiments to verify our proposed method on a typical liver segmentation and a challenging COVID-19 infection segmentation: (i) **LiTS dataset** [3]: ISBI LiTS 2017 Challenge dataset contains 131 contrast-enhanced abdominal scans. This dataset was acquired by different scanners from six different clinical sites, with a

Models(SSLs)	0.1:0.9		0.3:0.7		0.5:0.5		0.7:0.3		1.0:0.0	
	Dice [%]	VOE [%]	Dice [%]	VOE [%]	Dice [%]	VOE [%]	Dice [%]	VOE [%]	Dice [%]	VOE [%]
LiTS Dataset										
Fully-supervised	$83.87^{\pm1.71}$	$27.21^{\pm2.21}$	$88.34^{\pm2.04}$	$20.51^{\pm3.18}$	$91.02^{\pm2.03}$	$16.32^{\pm3.38}$	$92.35^{\pm1.58}$	$14.09^{\pm2.66}$	$93.54^{\pm0.95}$	$12.10^{\pm1.68}$
Sedai *et al.* [24]	$86.54^{\pm1.16}$	$23.55^{\pm1.81}$	$89.02^{\pm1.55}$	$19.64^{\pm2.51}$	$91.50^{\pm1.30}$	$15.51^{\pm2.19}$	$92.79^{\pm1.49}$	$13.35^{\pm2.55}$	$93.35^{\pm0.98}$	$12.33^{\pm1.61}$
Ouali *et al.* [22]	$89.46^{\pm1.13}$	$18.96^{\pm1.87}$	$91.31^{\pm1.32}$	$15.84^{\pm2.19}$	$92.81^{\pm1.04}$	$13.30^{\pm1.73}$	$93.04^{\pm1.10}$	$12.93^{\pm1.89}$	$93.56^{\pm0.97}$	$12.08^{\pm1.70}$
Chen *et al.* [5]	$87.82^{\pm1.34}$	$21.53^{\pm2.12}$	$89.28^{\pm1.16}$	$19.22^{\pm1.85}$	$91.88^{\pm0.66}$	$14.93^{\pm1.10}$	$93.17^{\pm0.95}$	$12.61^{\pm1.75}$	$93.80^{\pm1.04}$	$11.61^{\pm1.81}$
Hung *et al.* [10]	$88.86^{\pm0.92}$	$19.90^{\pm1.48}$	$90.77^{\pm1.07}$	$16.79^{\pm1.77}$	$92.16^{\pm0.82}$	$14.35^{\pm1.48}$	$93.46^{\pm0.61}$	$12.18^{\pm0.95}$	$93.51^{\pm1.33}$	$12.07^{\pm2.28}$
Nie *et al.* [21]	$89.04^{\pm1.72}$	$19.54^{\pm2.79}$	$91.01^{\pm1.22}$	$16.35^{\pm2.06}$	$91.96^{\pm0.93}$	$14.66^{\pm1.50}$	$93.06^{\pm1.22}$	$12.91^{\pm2.11}$	$93.67^{\pm0.85}$	$11.78^{\pm1.41}$
Zheng *et al.* [34]	$90.18^{\pm0.98}$	$17.73^{\pm1.48}$	$91.71^{\pm1.02}$	$15.21^{\pm1.73}$	$93.27^{\pm0.78}$	$12.47^{\pm0.61}$	$93.89^{\pm0.81}$	$11.39^{\pm1.34}$	$94.49^{\pm0.56}$	$10.42^{\pm1.01}$
BAS^{3}Net	$91.11^{\pm0.91}$	$16.19^{\pm1.51}$	$92.65^{\pm0.88}$	$13.56^{\pm1.47}$	$94.01^{\pm0.75}$	$11.27^{\pm1.33}$	$94.81^{\pm0.67}$	$10.10^{\pm0.95}$	$95.23^{\pm0.54}$	$9.38^{\pm0.71}$
Graph-BAS^{3}Net	$93.19^{\pm0.94}$	$12.69^{\pm1.61}$	$94.56^{\pm0.77}$	$10.27^{\pm1.36}$	$94.97^{\pm0.72}$	$9.83^{\pm1.03}$	$95.25^{\pm0.48}$	$9.33^{\pm0.60}$	$95.58^{\pm0.44}$	$8.76^{\pm0.49}$
COVID-19 Dataset										
Fully-supervised	$65.87^{\pm4.56}$	$50.25^{\pm5.61}$	$69.55^{\pm3.97}$	$45.67^{\pm4.89}$	$74.88^{\pm3.35}$	$39.09^{\pm4.10}$	$77.11^{\pm3.22}$	$36.33^{\pm4.15}$	$79.33^{\pm2.90}$	$33.51^{\pm3.61}$
Sedai *et al.* [24]	$67.09^{\pm3.21}$	$48.72^{\pm3.95}$	$71.65^{\pm3.96}$	$43.03^{\pm4.91}$	$75.74^{\pm3.86}$	$37.96^{\pm4.79}$	$78.55^{\pm2.54}$	$34.47^{\pm3.14}$	$79.40^{\pm2.78}$	$33.54^{\pm3.58}$
Ouali *et al.* [22]	$69.79^{\pm3.00}$	$45.39^{\pm3.67}$	$73.05^{\pm3.05}$	$41.01^{\pm3.49}$	$76.05^{\pm2.84}$	$37.72^{\pm3.54}$	$78.97^{\pm2.73}$	$33.95^{\pm3.39}$	$79.33^{\pm2.90}$	$33.51^{\pm3.61}$
Chen *et al.* [5]	$67.98^{\pm3.54}$	$47.63^{\pm4.34}$	$72.31^{\pm2.99}$	$42.21^{\pm3.71}$	$76.44^{\pm3.07}$	$37.11^{\pm3.81}$	$77.36^{\pm2.81}$	$36.05^{\pm3.59}$	$79.74^{\pm2.81}$	$32.84^{\pm3.82}$
Hung *et al.* [10]	$68.55^{\pm2.83}$	$46.93^{\pm3.46}$	$72.33^{\pm2.67}$	$42.19^{\pm3.31}$	$76.24^{\pm2.94}$	$37.29^{\pm3.61}$	$79.23^{\pm2.95}$	$33.63^{\pm3.66}$	$79.60^{\pm2.73}$	$33.30^{\pm3.51}$
Nie *et al.* [21]	$70.33^{\pm2.88}$	$44.76^{\pm3.50}$	$73.36^{\pm2.69}$	$40.66^{\pm3.09}$	$76.79^{\pm2.91}$	$36.76^{\pm3.71}$	$79.43^{\pm2.68}$	$33.38^{\pm3.32}$	$79.98^{\pm2.60}$	$32.74^{\pm3.26}$
BAS^{3}Net	$72.98^{\pm2.11}$	$41.38^{\pm2.62}$	$74.85^{\pm2.86}$	$38.82^{\pm3.23}$	$77.98^{\pm2.14}$	$35.18^{\pm2.65}$	$80.28^{\pm2.11}$	$32.40^{\pm2.70}$	$80.91^{\pm2.10}$	$31.65^{\pm2.70}$
Graph-BAS^{3}Net	$74.22^{\pm2.65}$	$39.97^{\pm3.16}$	$77.35^{\pm2.04}$	$36.00^{\pm2.49}$	$80.23^{\pm1.51}$	$32.42^{\pm1.89}$	$81.48^{\pm1.89}$	$30.84^{\pm2.39}$	$82.09^{\pm1.76}$	$29.70^{\pm2.48}$

Table 1. Comparison of our methods (**in orange**)with the state-of-the-art semi-supervised methods on two datasets.

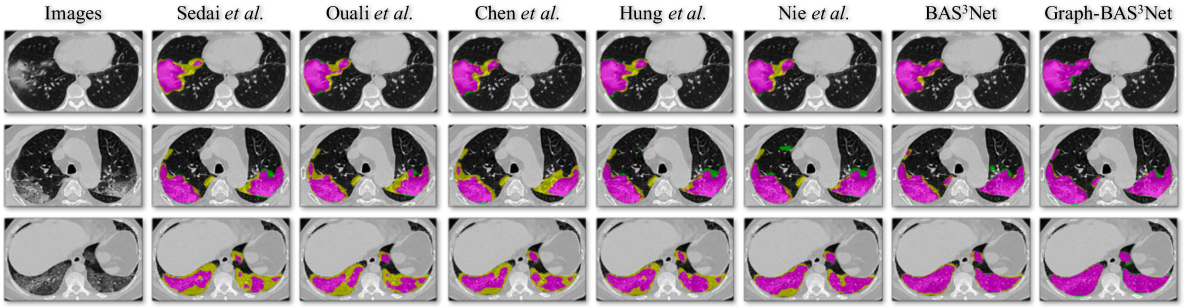

Figure 4. Qualitative comparisons of three typical examples with the state-of-the-art methods on COVID-19 dataset with 10% labeled data. The **purple areas** are the true positives (TP); the **yellow areas** are false negatives (FN), and the **green areas** are false positives (FP).

largely varying in-plane resolution from 0.55 to 1.0 mm and slice spacing from 0.45 to 6.0 mm. The image resolution is a relatively high 512×512. The dataset also includes 103 and 28 volumes for training and testing, respectively. We further randomly divided the 103 training cases into a training set and a validation set at the ratio of 3:1. To eliminate the effect of randomness, we conducted the partition operation twice. The hyper-parameters optimization and network development were conducted on validation set. (**ii**) **COVID-19 dataset**: we collected 102 COVID-19 CT scans from the First Affiliated Hospital. The left, right lung and infections were annotated by two radiologists with 5-year experience in chest radiology. Each case had a slice-plane resolution of 512×512 and was resampled with the same spacing of 1.0×1.0×1.0 mm^3. To reduce the randomness, the dataset was randomly split into a training set, a validation set, and a testing set at the ratio of 3:1:1 for twice.

4.2. Implementation Details

To update the parameters of semantic segmentation and boundary detection network, we adopted Stochastic Gradient Descent. Here the momentum was set to 0.9 and the weight decay to 1e-4. The initial learning rate was 1e-3, which was decreased following the polynomial decay with a power of 0.9. As for the discriminator, we performed Adam optimizer with the learning rate as 1e-4 and the same polynomial decay. The betas were set as 0.9 and 0.999.

We trained the models in 150k iterations on LiTS dataset and COVID-19 dataset with a batch size of 3. To capture space context along z-axis, the input image consisted of three slices: the slice to be segmented and the upper and lower slices, which was resized to 320×320×3. Our results showed that BAS^{3}Net was more stable with pre-trained semantic segmentation and boundary detection networks. Therefore, we first pre-trained two networks in a fully supervised mode for 10k iterations with the labeled data. Then, the discriminator network joined the optimization, which was also updated with the labeled data. To eliminate the noisy predictions, semi-supervised learning began after training for 20k iterations. To ensure the evaluation robustness, we used two random seeds to sample the labeled and unlabeled data and obtained the average value of these

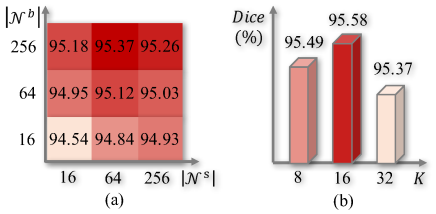

Figure 5. Hyper-parameter of graph node number (a) and feature dimension (b) on LiTS dataset in fully-supervised mode.

4.3. Comparison with the State-of-the-Art Methods

In Table 1, our methods were compared to other semi-supervised state-of-the-art methods. Different from GAN-based SSLs [10, 21, 34], Chen *et al.* [5] trained an auto-encoder to reconstruct synthetic segmentation labels created by attention mechanism. Sedai *et al.* [24] proposed an uncertainty guided SSL for medical image segmentation by using the Monte Carlo (MC) dropout; while Ouali *et al.* [22] presented a cross-consistency based SSL for semantic segmentation. Note that the atlas-prior proposed by Zheng *et al.* [34] was difficult to achieve in practice for COVID-19 infections with large variation in pose and shape. For fair comparison, we used the same DeepLabV2 [4] backbone in these methods. We randomly sampled 10%, 30%, 50%, 70%, 100% images as the labeled data, and used the rest of the training images as unlabeled data. Compared with the fully-supervised mode that only trained with the labeled data (shown in the first row), incremental improvements were coming from the use of unlabeled data. Moreover, our Graph-BAS^{3}Net outperformed other methods on both two datasets, especially with few labeled data.

4.4. Hyper-Parameter Analysis

Hyper-parameter of the graph node number and feature dimension. Here we first investigated how the node numbers, $|\mathcal{N}^s|$ and $|\mathcal{N}^b|$ affect the performance. The experiments were performed with 100% labeled data on the LiTS dataset. Note that we set the node feature dimension to 32, i.e., $K = 32$, in this analysis. As seen in Fig.6(a), the accuracy was improved when increasing $|\mathcal{N}^b|$. This is because the boundary is changeable and requires more anchors. By fixing $|\mathcal{N}^b|$ to 256, it is seen that the accuracy was the highest for $|\mathcal{N}^s| = 64$. However, increasing $|\mathcal{N}^s|$ may break the holistic semantic representation and increase the computational complexity. Therefore, we chose the $|\mathcal{N}^s| = 64$ and $|\mathcal{N}^b| = 256$, which provided the best results within a reasonable computational cost.

Following the experiment of the number of the graph nodes, node feature dimension K was also varied in a similar experiment setting. We set $|\mathcal{N}^s|$ to 64 with $|\mathcal{N}^b| = 256$ to assess the impact of the K which varies from 8 to 32. As seen in Fig.6(b), the accuracy was improved by decreasing the dimension of the feature and was at its best for $K = 16$. Therefore, we chose $K = 16$ in our experiments.

labeled	λ_{det}	λ_{adv}	λ_{semi}	T_{semi}	Dice [%]
100%	0	0	0	N/A	93.54
	0.5	0	0	N/A	94.06
	1.0	0	0	N/A	94.73
100%	1.0	0.001	0	N/A	95.47
	1.0	0.005	0	N/A	95.58
	1.0	0.05	0	N/A	95.03
10%	1.0	0.005	0	N/A	89.54
	1.0	0.005	0.005	0.2	92.47
	1.0	0.005	0.01	0.2	92.99
	1.0	0.005	0.02	0.2	92.60
10%	1.0	0.005	0.01	0	92.24
	1.0	0.005	0.01	0.2	92.99
	1.0	**0.005**	**0.01**	**0.3**	**93.19**
	1.0	0.005	0.01	0.5	92.95
	1.0	0.005	0.01	1.0	91.81

Table 2. Hyper-parameter of λ_{det}, λ_{adv}, λ_{semi} and T_{semi} in Graph-BAS^{3}Net architecture. Experiments are performed on LiTS dataset under the fully/semi-supervised settings.

#	GN^s	GN^b	DN		Dice [%]
			$GN^s(I_l)$	$GN^s(I_l) \,\|\|\, GN^b(I_l)$	
1	√				83.87
2	√	√			86.43
3	√	√	√		89.56
4	√	√		√	**91.11**

Table 3. Ablation of BAS^{3}Net (w/o BGCM) on LiTS dataset with 10% labeled data.

Hyper-parameter of λ_{det}, λ_{adv}, λ_{semi} and T_{semi}. The experiments were conducted on the LiTS dataset in the fully supervised mode (100% labeled data) and semi-supervised mode (10% labeled data). We first evaluated the effect on λ_{det} in the fully supervised mode, which achieved its best at $\lambda_{det} = 1.0$. It indicates that the boundary detection task is equally important as the semantic segmentation. Second, we showed comparisons of different values of λ_{adv} under the fully supervised setting. Overall, λ_{adv} with the medium value of 0.005 achieved the best performance. We further examined different values of λ_{semi} in the semi-supervised mode and set T_{semi} as 0.2 for comparisons. As shown in Table 2, the method performed the best for $\lambda_{semi} = 0.01$. Based on the above analysis, we also investigated how the choice of T_{semi} affected the performance and observed that the best performance was achieved for $T_{semi} = 0.3$.

4.5. Ablation Study

Ablation of BAS^{3}Net. Table 3 presented the segmentation accuracy on LiTS dataset with 10% labeled data, where the components were gradually added to the semantic segmen-

#	BAS^{3}Net	$\mathcal{G}^s$ construction		$\mathcal{G}^b$ construction		Reasoning direction		Dice [%]
		w/o seg map	seg map	w/o edge map	edge map	$\mathcal{G}^s \rightarrow \mathcal{G}^b$	$\mathcal{G}^b \rightarrow \mathcal{G}^s$	
1	√							91.11
2	√	√						91.78
3	√		√					92.56
4	√			√				91.51
5	√				√			91.82
6	√		√		√	√		92.27
7	√		√		√		√	92.82
8	√		√		√	√	√	**93.19**

Table 4. Ablation of BGCM on LiTS dataset with 10% labeled data.

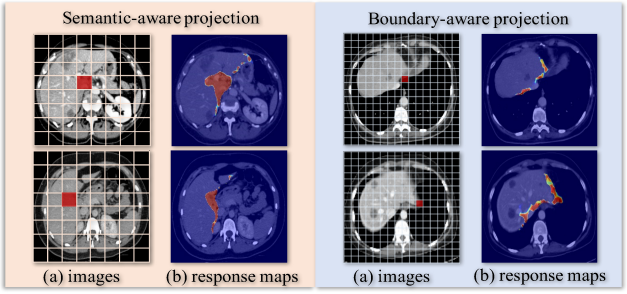

Figure 6. Interpretation of two graph projections on LiTS dataset. The left row in each block indicated the input image with an anchor marked in the red rectangle, and the right visualized response maps to the anchor. Darker color indicated a higher response.

tation network (GN^s). First, we examined the impact of introducing a boundary detection network (GN^b), and the accuracy was improved from 83.87% to 86.43%. This indicates the essential role of edge information in the segmentation. A significant increment from 89.56% to 91.11% was achieved by combining boundary detection result $GN^b(I)$, which helped the discriminator focus on the edge part.

Ablation of BGCM. To validate the efficiency of the proposed BGCM, we considered different graph construction and reasoning directions. Regarding the single $\mathcal{G}^s$, the dice accuracy was improved because of the semantic-wise reasoning that considered the correlations among proposals. Using the segmentation attention map M^s, $\mathcal{G}^s$ achieved a performance of 92.56% with 1.45% gain. This further indicates the effectiveness of incorporating anatomic knowledge for better localization. As for the single $\mathcal{G}^b$, the performance was increased from 91.11% to 91.82%, where 0.31% was incrementally obtained by using the edge attention map M^b. This validates the necessity of enhancing the feature map with edge via attention mechanism.

With applying the intra-task reasoning, an increment from 91.82% to 93.19% was achieved, where 0.45% and 1.00% improvements were produced by using a single direction of $\mathcal{G}^s \rightarrow \mathcal{G}^b$ (from $\mathcal{G}^s$ to $\mathcal{G}^b$) and $\mathcal{G}^b \rightarrow \mathcal{G}^s$ (from $\mathcal{G}^b$ to $\mathcal{G}^s$). The best performance was achieved at 93.19% accuracy by using bilateral reasoning, which learned the mutual relations from tasks and yielded clear boundaries.

4.6. Interpretation of Graph Projections

We further visualized the semantic-aware and boundary-aware graph projections for interpretation. In the graph projection, we defined a set of anchors (e.g. 8×8 anchors in $\mathcal{G}^s$ and 16×16 anchors in $\mathcal{G}^b$) to bridge the connection between the pixels and the corresponding nodes. As seen in Fig.7, we visualized the weight of each pixel that contributed to an anchor marked in a red rectangle, where darker color indicated a higher response. It is seen that two graph projections aggregate pixels with similar appearance to the same anchor. This is because the response areas are consistent with the anchor. As expected, the semantic-aware projection modeled long-range dependencies while the boundary-aware projection focused on edge discriminative patterns.

5. Conclusion

In this paper, we focus on the blurry boundary issue and propose a Graph-BAS^{3}Net, which consists of two components: the backbone BAS^{3}Net and an interaction module BGCM. BAS^{3}Net adopts a multi-task learning generator to jointly conduct segmentation and boundary detection. Then the combined segmentation and boundary maps are fed into the discriminator for further distinguishing. Under this framework, the BGCM takes graph structures and interacts between multi-tasks to mine intra- and inter-task relations, which enhances both two tasks. Experimental results confirmed that our method surpassed all state-of-the-art approaches for semi-supervised medical image segmentation.

Acknowledgement. This work was supported in part by the Major Scientific Research Project of Zhejiang Lab (No. 2020ND8AD01), in part by the Grant-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) (No. 20KK0234, No. 21H03470 and No. 20K21821), and in part by the China Postdoctoral Science Foundation (No. 2020TQ0293).

References

[1] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017. 2

[2] Wenjia Bai, Ozan Oktay, Matthew Sinclair, Hideaki Suzuki, Martin Rajchl, Giacomo Tarroni, Ben Glocker, Andrew King, Paul M Matthews, and Daniel Rueckert. Semi-supervised learning for network-based cardiac mr image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 253–260. Springer, 2017. 1

[3] Patrick Bilic, Patrick Ferdinand Christ, Eugene Vorontsov, Grzegorz Chlebus, Hao Chen, Qi Dou, Chi-Wing Fu, Xiao Han, Pheng-Ann Heng, Jürgen Hesser, et al. The liver tumor segmentation benchmark (lits). *arXiv preprint arXiv:1901.04056*, 2019. 5

[4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 3, 7

[5] Shuai Chen, Gerda Bortsova, Antonio García-Uceda Juárez, Gijs van Tulder, and Marleen de Bruijne. Multi-task attention-based semi-supervised learning for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 457–465. Springer, 2019. 1, 6, 7

[6] Yunpeng Chen, Marcus Rohrbach, Zhicheng Yan, Yan Shuicheng, Jiashi Feng, and Yannis Kalantidis. Graph-based global reasoning networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 433–442, 2019. 2, 3

[7] Wenhui Cui, Yanlin Liu, Yuxing Li, Menghao Guo, Yiming Li, Xiuli Li, Tianle Wang, Xiangzhu Zeng, and Chuyang Ye. Semi-supervised brain lesion segmentation with an adapted mean teacher model. In *International Conference on Information Processing in Medical Imaging*, pages 554–565. Springer, 2019. 1

[8] Huimin Huang, Ming Cai, Lanfen Lin, Jing Zheng, Xiongwei Mao, Xiaohan Qian, Zhiyi Peng, Jianying Zhou, Yutaro Iwamoto, Xian-Hua Han, et al. Graph-based pyramid global context reasoning with a saliency-aware projection for covid-19 lung infections segmentation. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1050–1054. IEEE, 2021. 2

[9] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059. IEEE, 2020. 2

[10] Wei-Chih Hung, Yi-Hsuan Tsai, Yan-Ting Liou, Yen-Yu Lin, and Ming-Hsuan Yang. Adversarial learning for semi-supervised semantic segmentation. *arXiv preprint arXiv:1802.07934*, 2018. 1, 2, 6, 7

[11] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016. 2, 4

[12] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*, 2016. 1

[13] Qiaozhe Li, Xin Zhao, Ran He, and Kaiqi Huang. Pedestrian attribute recognition by joint visual-semantic reasoning and knowledge distillation. In *IJCAI*, pages 833–839, 2019. 2

[14] Shuailin Li, Chuyu Zhang, and Xuming He. Shape-aware semi-supervised 3d semantic segmentation for medical images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 552–561. Springer, 2020. 1

[15] Xia Li, Yibo Yang, Qijie Zhao, Tiancheng Shen, Zhouchen Lin, and Hong Liu. Spatial pyramid based graph reasoning for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8950–8959, 2020. 2

[16] Xiaomeng Li, Lequan Yu, Hao Chen, Chi-Wing Fu, and Pheng-Ann Heng. Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model. *arXiv preprint arXiv:1808.03887*, 2018. 1

[17] Yin Li and Abhinav Gupta. Beyond grids: Learning graph representations for visual recognition. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 9245–9255, 2018. 2

[18] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 2

[19] Fang Lu, Fa Wu, Peijun Hu, Zhiyi Peng, and Dexing Kong. Automatic 3d liver location and segmentation via convolutional neural network and graph cut. *International journal of computer assisted radiology and surgery*, 12(2):171–182, 2017. 1

[20] R Muthukrishnan and Miyilsamy Radha. Edge detection techniques for image segmentation. *International Journal of Computer Science & Information Technology*, 3(6):259, 2011. 3

[21] Dong Nie, Yaozong Gao, Li Wang, and Dinggang Shen. Asdnet: attention based semi-supervised deep networks for medical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 370–378. Springer, 2018. 1, 2, 6, 7

[22] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12674–12684, 2020. 1, 6, 7

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2

[24] Suman Sedai, Bhavna Antony, Ravneet Rai, Katie Jones, Hiroshi Ishikawa, Joel Schuman, Wollstein Gadi, and Rahil Garnavi. Uncertainty guided semi-supervised segmentation of retinal layers in oct images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 282–290. Springer, 2019. 6, 7

[25] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *arXiv preprint arXiv:1703.01780*, 2017. 1

[26] Gusi Te, Yinglu Liu, Wei Hu, Hailin Shi, and Tao Mei. Edge-aware graph representation learning and reasoning for face parsing. In *European Conference on Computer Vision*, pages 258–274. Springer, 2020. 2, 3

[27] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017. 5

[28] Yangxin Wu, Gengwei Zhang, Yiming Gao, Xiajun Deng, Ke Gong, Xiaodan Liang, and Liang Lin. Bidirectional graph reasoning network for panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9080–9089, 2020. 2, 3

[29] Yingda Xia, Fengze Liu, Dong Yang, Jinzheng Cai, Lequan Yu, Zhuotun Zhu, Daguang Xu, Alan Yuille, and Holger Roth. 3d semi-supervised learning with uncertainty-aware multi-view co-training. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3646–3655, 2020. 1

[30] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015. 3

[31] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi-Wing Fu, and Pheng-Ann Heng. Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 605–613. Springer, 2019. 1

[32] Li Zhang, Xiangtai Li, Anurag Arnab, Kuiyuan Yang, Yunhai Tong, and Philip HS Torr. Dual graph convolutional network for semantic segmentation. *arXiv preprint arXiv:1909.06121*, 2019. 2

[33] Yizhe Zhang, Lin Yang, Jianxu Chen, Maridel Fredericksen, David P Hughes, and Danny Z Chen. Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 408–416. Springer, 2017. 1

[34] Han Zheng, Lanfen Lin, Hongjie Hu, Qiaowei Zhang, Qingqing Chen, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, Ruofeng Tong, and Jian Wu. Semi-supervised segmentation of liver using adversarial learning with deep atlas prior. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 148–156. Springer, 2019. 1, 2, 6, 7

[35] Yuyin Zhou, Yan Wang, Peng Tang, Song Bai, Wei Shen, Elliot Fishman, and Alan Yuille. Semi-supervised 3d abdominal multi-organ segmentation via deep multi-planar co-training. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 121–140. IEEE, 2019. 1

[36] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018. 2