Examinee-Examiner Network: Weakly Supervised Accurate Coronary Lumen Segmentation Using Centerline Constraint

Yaolei Qi, Han Xu, Yuting He, Guanyu Li, Zehang Li, Youyong Kong, *Member, IEEE*, Jean-Louis Coatrieux, *Life Fellow, IEEE*, Huazhong Shu, *Senior Member, IEEE*, Guanyu Yang, *Senior Member, IEEE*, and Shengxian Tu

***Abstract*— Accurate coronary lumen segmentation on coronary-computed tomography angiography (CCTA) images is crucial for quantification of coronary stenosis and the subsequent computation of fractional flow reserve. Many factors including difficulty in labeling coronary lumens, various morphologies in stenotic lesions, thin structures and small volume ratio with respect to the imaging field complicate the task. In this work, we fused the continuity topological information of centerlines which are easily accessible, and proposed a novel weakly supervised model, Examinee-Examiner Network (EE-Net), to overcome the challenges in automatic coronary lumen segmentation. First, the EE-Net was proposed to address the fracture in segmentation caused by stenoses by combining the semantic features of lumens and the geometric constraints of continuous topology obtained from the centerlines. Then, a Centerline Gaussian Mask Module was proposed to deal with the insensitiveness of the network to the centerlines. Subsequently, a weakly supervised learning strategy, Examinee-Examiner Learning, was proposed to handle the weakly supervised situation with few lumen labels by using our EE-Net to guide and constrain the segmentation with customized prior conditions. Finally, a general network layer, Drop Output Layer, was proposed to adapt to the class imbalance by dropping well-segmented regions and weights the classes dynamically. Extensive experiments on two different data sets demonstrated that our EE-Net has good continuity and generalization ability on coronary lumen segmentation task compared with several widely used CNNs such as 3D-UNet. The results revealed our EE-Net with great potential for achieving accurate coronary lumen segmentation in patients with coronary artery disease. Code at http://github.com/qiyaolei/Examinee-Examiner-Network.**

***Index Terms*— 3D accurate coronary lumen segmentation, weakly supervised learning, examinee-examiner network, centerline Gaussian mask module, examinee-examiner learning, drop output layer, CT angiography image.**

Manuscript received November 24, 2020; revised July 16, 2021 and October 19, 2021; accepted October 21, 2021. Date of publication November 10, 2021; date of current version November 17, 2021. This work was supported in part by the National Natural Science Foundation under Grant 31800825, Grant 62171125, and Grant 61828101; and in part by the Excellence Project Funds of Southeast University. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiangqian Wu. *(Yaolei Qi and Han Xu contributed equally to this work.) (Corresponding authors: Guanyu Yang; Shengxian Tu.)*

Yaolei Qi, Han Xu, and Yuting He are with LIST, Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 210096, China.

Guanyu Li, Zehang Li, and Shengxian Tu are with the School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200241, China (e-mail: sxtu@sjtu.edu.cn).

Youyong Kong, Huazhong Shu, and Guanyu Yang are with LIST, Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 210096, China, also with the Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Southeast University, Nanjing 210096, China, and also with the Centre de Recherche en Information Biomédicale Sino-Français (CRIBs), 35000 Rennes, France (e-mail: yang.list@seu.edu.cn).

Jean-Louis Coatrieux is with the Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Southeast University, Nanjing 210096, China, and also with the Centre de Recherche en Information Biomédicale Sino-Français (CRIBs), 35000 Rennes, France.

Digital Object Identifier 10.1109/TIP.2021.3125490

I. Introduction

ACCURATE coronary lumen segmentation with stenoses on coronary-computed tomography angiography (CCTA) images, which obtains accurate segmentation quality in regions with stenoses, plays a pivotal role in Coronary Artery Disease (CAD) diagnosis and treatment. Clinically, functional coronary artery stenosis can be detected effectively and non-invasively using the FFR-CT [1] technique which depends on the computational fluid dynamics (CFD) simulation with the well-delineated coronary lumen in CCTA images [2]. Therefore, the accurate coronary lumen segmentation with stenoses obtains the well-delineated coronary lumen automatically and fleetly, thus becoming a key prerequisite for the accurate estimation of FFR-CT and assisting the radiologists to locate and diagnose the stenosis, to plan the further treatment method.

Deep neural networks (DNNs) have been widely used in tissue and organ segmentation [3], but it is still a challenging task for accurate coronary lumen segmentation with stenoses owing to the following difficulties: **Challenge 1: Difficult labeling.** The coronary lumen on the cross-sectional planes presents a morphology with small areas and fuzzy boundaries (Fig. 1(b)) and manual annotation needs to be drawn slice by slice on the planes along the vessels. It is a time-consuming labor work to delineate the contours, so that the amount of labeled data used for training is limited, leading to the low generalization on the network. **Challenge 2: Stenotic lesions.** Due to the compression of the plaques, the coronary lumen is abnormally narrowed and forms the stenotic lesions (Fig. 1(a)(b)). Stenosis makes the coronary lumen have different morphological features from those in the healthy

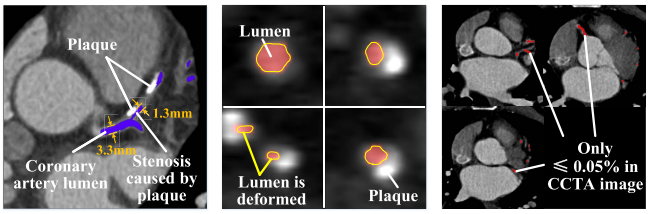

Fig. 1. The challenges of accurate coronary lumen segmentation with stenoses. a) shows the sharp narrowing caused by plaque. The normal diameter of the artery is about 3.3mm, while the diameter is only 1.3mm in the stenosis regions. b) shows the stenosis on the cross-sectional plane. Lumen is deformed by plaque compression. c) shows the small volume ratio. Lumen accounts for less than 0.05%.

regions, which will lead to the underfitting of the network, thus causing the fracture of the segmentation in regions with stenoses. **Challenge 3: Thin structures.** The normal diameter of the coronary lumen is between 2 mm and 5 mm, while it is only about 1 mm in stenosis regions(Fig. 1(a)). Such thin structures have numerous hard-to-segment regions and will make the network more prone to over-segmentation or under-segmentation. **Challenge 4: Small volume ratio.** The coronary lumen only accounts for less than 0.05% in CCTA images (Fig. 1(c)) which makes the target and the background have large-scale differences. The differences cause the class imbalance so that the network will show weak segmentation on the minority classes.

To accomplish the accurate coronary lumen segmentation with stenoses, we proposed a novel end-to-end model, called Examinee-Examiner Network (EE-Net). It simulates the mode that the Examinee Network (E_1-Net) implements the segmentation task while the Examiner Network (E_2-Net) facilitates the supervision from the weakly supervised label and evaluates the segmentation results of E_1-Net, and they cooperate to acquire preferable segmentation quality. In detail, our EE-Net has completed three assignments. 1) E_1-Net realizes the end-to-end coronary lumen segmentation using lumen labels to learn semantic features from pixel-to-pixel. 2) E_2-Net customizes the prior condition of the geometric constraints by extracting the continuous topological structure information of the centerline. 3) E_2-Net applies the prior condition to evaluate the segmentation result of E_1-Net and gives the feedback to optimize the E_1-Net. Based on EE-Net, we made detailed designs for each challenge mentioned above:

A. Addressing Challenge 1 and 2

Weakly supervised learning strategy [4] is suitable for dealing with the lack of annotations caused by the difficult labeling. However, some previous weakly supervised strategies [3], [5], [6] cannot be applied to our task due to the inadaptability to thin and changeable segmentation and the loss of continuity. Fortunately, the centerlines have more concentrated continuity topology information which can be used as the weakly supervised label. And the automated centerline extraction is convenient [7]–[9], which inspires us to propose a novel weakly supervised learning strategy, Examiner-Examinee Learning (EE-Learning), in our EE-Net. Different from the Student-Teacher in the combination mode and the task of each network, we call our strategy EE-Learning instead of the student-teacher strategy. The E_2-Net learns the relationship between the lumen labels and the continuous topological features of the centerline labels as a prior condition, to constrain the E_1-Net to satisfy the segmentation results with a continuous topological structure. Therefore, when the data only with the centerline labels are available, the E_2-Net evaluates the segmentation results from the E_1-Net directly using the prior conditions, thus achieving good results on the weakly labeled dataset.

As a weakly supervised label, the centerlines have more concentrated continuity topology information which is suitable for dealing with the segmentation fracture caused by the stenotic lesions. However, the centerlines are composed of a series of voxels and accounts for a small proportion, which makes it difficult for the network to pay attention. We proposed the Centerline Gaussian Mask (CGM) Module to enhance the extracted features of the centerlines, which makes it easier for the network to focus on the continuous structure information.

B. Addressing Challenge 3 and 4

The majority classes are easy to converge, while the minority classes are difficult thus forming the hard-to-segment regions. Therefore, we proposed a general network layer, DropOutput Layer (DO-Layer), which will dynamically adapt to the learning process to alleviate the class imbalance problem. Our Drop Output Layer drops well-segmented regions in the output map and weights the classes dynamically. In this way, the network will give hard-to-segment regions more training opportunities and show superior performance on the minority classes.

To summarize, our work realizes the accurate coronary lumen segmentation with stenoses and the specific contributions are three-fold:

- We proposed a novel end-to-end model, EE-Net, to serve the weakly supervised labels and realize the segmentation. Our EE-Net customizes the prior condition of the geometric constraints by extracting the continuous topological features of the centerlines and embedded the customized prior conditions into the network to constrain and optimize the segmentation results. To improve the sensitiveness of the network to the continuity topology features, CGM Module is proposed to enhance the centerlines. In the analysis, the internal and external test data are used simultaneously, which fully verifies the effectiveness of our model.
- We proposed a novel weakly supervised learning strategy, EE-Learning, to handle the weakly supervised situation. When the weakly supervised data only with the centerline labels are available, the E_2-Net evaluates the segmentation results from the E_1-Net directly using the prior customized conditions generated before, thus achieving good results on the dataset with limited lumen labels.

- We proposed a general network layer, DO-Layer, to adapt to the situation of classes imbalance. Our DO-Layer drops well-segmented regions in the output map and weights the classes dynamically, which will adapt to the training process to balance the categories. The network will give hard-to-segment regions more training opportunities and show superior performance on the minority classes.

For the rest of the paper, we list the related work in Sec. II. Next, we demonstrate the details of our proposed Examinee-Examiner Network (EE-Net) in Sec. III, including Centerline Gaussian Mask Module (CGM Module), Examinee-Examiner Learning (EE-Learning), and Drop Output Layer (DO-Layer). Then the dataset description, experiment settings, and evaluation measures are illustrated in Sec. IV. To verify the superiority of our method, Sec. V shows the results and analysis of comparison and ablation experiments. Finally, Sec. VI summarizes and discusses the whole paper.

II. Related Work

Recently, traditional methods and deep learning methods have made progress in coronary lumen segmentation. However, accurate coronary lumen segmentation with stenoses is still an open problem.

A. Vessel Segmentation

Numerous works have made efforts on vessel segmentation due to the significance for CAD diagnosis and preoperative planning. Some traditional methods [10], [11], such as region-growing approaches [12], level-set methods [13], centerline-based methods [8] and other traditional methods [14], [15], have obtained good results in vessel segmentation scenarios. But, these methods rely on hand-crafted features and the generalization ability is limited in the scene with changeable morphological structures. With the development of deep learning, vessel segmentation based on deep learning [16]–[23] is used owing to their efficiency, high accuracy and powerful generalization ability. Chen *et al.* [16] proposed a fully automatic framework that adopts a paired multiscale 3D deep convolutional neural networks to segment the coronary artery. Huang *et al.* [18] presented approaches with 3D U-Net for both CACT data with and without centerline. Wolterink *et al.* [24] proposed to use graph convolutional networks(GCNs) to predict the spatial location of vertices in a tubular surface mesh that segments the coronary artery. Kong *et al.* [19] proposed a tree-structure convolutional gated recurrent unit model to learn the anatomical structure of the coronary artery and conduct voxel-wise segmentation. Gu *et al.* [20] proposed a global feature embedded network, which contains semantic information and detailed features, for coronary arteries segmentation. Jun *et al.* [21] proposed a nested encoder-decoder architecture for the main vessel segmentation in coronary angiography. Mou *et al.* [22] proposed a generic and unified convolution neural network for the segmentation of curvilinear structures and illustrate in several 2D/3D medical imaging modalities. Some work [25], [26] uses the auto-context strategy, which will help the network get better segmentation results through iteration.

However, these methods cannot meet the requirements of our task because of some limitations: 1) Lack of comprehensive assessment. Their works are devoted to coronary artery segmentation ignoring the effect of stenosis. Therefore, the segmentation performance is unknown for data with stenotic lesions. 2) Dependence on accurately labeled datasets. These methods are fully supervised approaches. These supervised models are trained based on accurately labeled datasets that need labor-intensive annotations, and if the labeled data are limited, their generalization ability will be weak.

B. Weakly Supervised Learning

Weakly supervised learning has great potential in improving the generalization ability of neural networks [4] via utilizing partial labels [27], [28], pseudo labels [29], [30], image-level labels [31], regional bounding box [32] and consistency constraint in co-segmentation [33], [34]. 1) Partial labels. Due to the difficulty in obtaining the complete annotations, some methods using partial labels have been proposed. Kervadec *et al.* [27] proposed a differentiable penalty, which enforces inequality constraints directly in the loss function, for weakly supervised segmentation. Peng *et al.* [28] proposed a method, based on the alternating direction method of multipliers (ADMM) algorithm, to train a CNN with discrete constraints and regularization priors. 2) Pseudo labels. Some methods have been proposed to better used the pseudo labels. Zhang *et al.* [29] proposed a weakly supervised training framework that learns from noisy pseudo labels generated from automatic vessel enhancement for weakly supervised vessel segmentation in X-ray Angiograms. Ma *et al.* [30] proposed a weakly supervised model with a multi-scale class activation map for GA segmentation in Spectral-Domain Optical Coherence Tomography images. 3) Image-level labels. Weakly supervised semantic segmentation uses image-level labels to identify the object regions. Meng *et al.* [31] proposed a new strategy employing a class-level multiple group cosegmentation and fusion method to deal with the image-level labels. 4) Regional bounding box. Bounding box annotations are widely used due to their simplicity and low-annotation cost. Kervadec *et al.* [32] proposed a weakly supervised learning segmentation based on several global constraints derived from regional bounding box. 5) Consistency constraint in co-segmentation. Chen *et al.* [33] proposed a method for jointly matching and segmenting object instances of the same category within a collection of images. Souly *et al.* [34] proposed a semi-supervised framework based on GANs to force the real samples to be close in the feature space by adding the large fake visual data, thus improving multiclass pixel classification.

However, these weak-supervised methods cannot be applied to our task directly due to some limitations: 1) Small and changeable vascular segmentation scenarios. It is difficult for these methods to construct accurate regularization constraints such as sizes and boundaries according to the small and changeable vessels. Therefore, the generalization results are weak in our task of accurate coronary segmentation. 2) Loss of continuity. The weakly-labeled data used to obtain pseudo labels will lose the continuous information easily. Therefore,

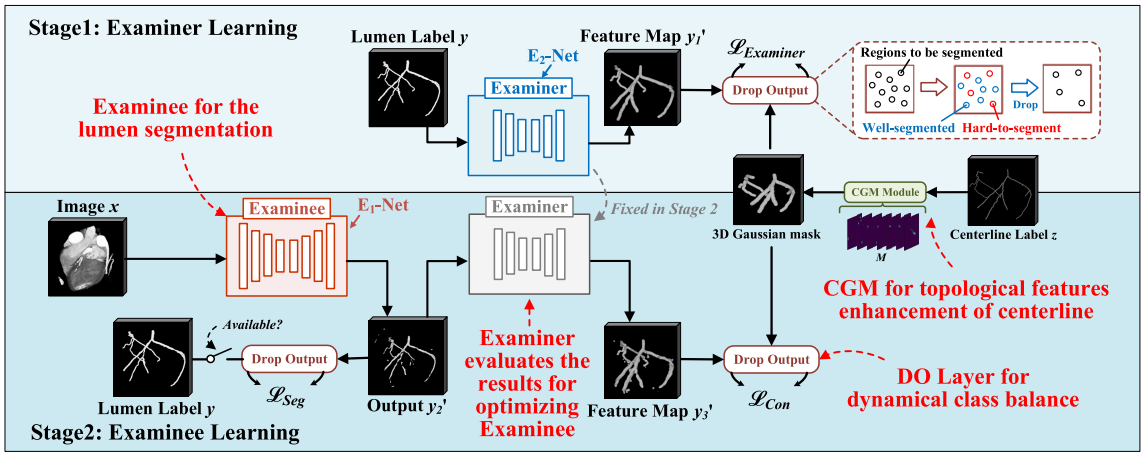

Fig. 2. Our EE-Net for accurate coronary lumen segmentation: **Stage 1: Examiner Learning.** E_2-Net learns the mapping relationship from the lumen labels to the weakly supervised labels, which is used in the next stage to facilitate the supervision from the specific label. **Stage 2: Examinee Learning.** E_2-Net is fixed as an examiner module. At this stage, E_1-Net will send the segmentation result as an input into the fixed E_2-Net, and take the output of the E_2-Net as the supervision information. The loss calculated from the output will be fed back to E_1-Net as the evaluation standard, to optimize the final segmentation of the E_1-Net.

the limited labels that bring inaccurate optimization objectives will interfere with the training process and deteriorate the performance of the model in our task.

C. Class Imbalance

Class imbalance means the distribution of examples across the known classes is biased or skewed in the dataset, a class imbalance will cause an inefficient training process and the degenerate models. The weighted cross-entropy loss [35] makes the network pay more attention to the minority class by giving higher weight, but the weight is a hyper-parameter and is difficult to determine. Based on Dice Loss, the Generalized Dice Loss (GDL) [36] weight contribution of each label by the inverse of its volume, thus reducing the well-known correlation between region size and Dice score. The Focal Loss [37] down-weights the loss assigned to well-classified examples and focuses training on a sparse set of hard examples. The hard region adaptation (HRA) loss [38] samples the loss function according to the segmentation quality so that the network will attach significance to the hard-to-segment region dynamically, thus keeping the class balanced.

However, the GDL loss does not distinguish the difficulty of hard regions in the segmentation task which leads to local defects easily and the Focal Loss adopts two hyper-parameters which should be tuned with a lot of effort and is a static loss which is not adaptive for the changing of data distribution, which varies along with the training process.

III. Methodology

Our method includes four aspects (Fig. 2): 1) Our proposed Examinee-Examiner Network (EE-Net, Sec. III-A) adopts the mode of cooperation between the Examinee Network (E_1-Net) and the Examiner Network (E_2-Net). The E_2-Net uses the relationship between the lumen labels and the continuous topological features extracted from the centerlines as a prior condition, to constrain the E_1-Net to satisfy the segmentation results with a continuous topological structure. 2) Our proposed Centerline Gaussian Mask Module (CGM, Sec. III-B) generates a 3D Gaussian mask according to the morphological features of the centerline to enhance the continuous topological structure features, thus helping the E_2-Net to be more sensitive to the centerline composed of voxels to generate geometric constraints. 3) Our proposed Examinee-Examiner Learning (EE-Learning, Sec. III-C) is based on our EE-Net. When the weakly labeled data with the centerline labels only is available, the E_2-Net evaluates the segmentation results from the E_1-Net directly using the prior conditions. Therefore, the E_2-Net will guide and constrain the segmentation results from the E_1-Net to achieve good results on the weakly labeled dataset.

4) Our proposed Drop Output Layer (DO-Layer, Sec. III-D) drops well-segmented regions in the output map and weights the classes dynamically to give the hard-to-segment regions more training opportunities thus balancing classes so that our EE-Net will adapt to the situation of class imbalance. All the symbols mentioned in this paper are defined in Tab. I.

TABLE I
The Detailed Definitions of Symbols in Our Paper

Symbol table		
Input	Input image	x
	Lumen label	y
	Centerline label	z
Output	Enhanced centerline label	z_M
	Output feature map	y_1', y_2', y_3'
	Weighted output after DO-Layer	$\hat{y_1}'$, $\hat{y_2}'$, $\hat{y_3}'$
	Weighted enhanced centerline label	$z_{\hat{M}}$
	Weighted lumen label	$\hat{y}$
Loss	Loss calculated from $\hat{y_1}'$ and $z_{\hat{M}}$	$\mathcal{L}_{Examiner}$
	Loss calculated from $\hat{y_2}'$ and $\hat{y}$	$\mathcal{L}_{Seg}$
	Loss calculated from $\hat{y_3}'$ and $z_{\hat{M}}$	$\mathcal{L}_{Con}$

A. EE-Net for Accurate Coronary Lumen Segmentation

Our proposed Examinee-Examiner Network (EE-Net) adopts the mode of cooperation between the Examinee

Network (E_1-Net) and the Examiner Network (E_2-Net). The E_1-Net realizes the task of coronary lumen segmentation using lumen labels to learn semantic features from pixel-to-pixel. While the E_2-Net learns the relationship between the lumen labels and the continuous topological features extracted from the centerlines as a prior condition. Based on the prior condition, E_2-Net generates geometric constraints for the feedback to evaluate and constrain the E_1-Net to satisfy the segmentation results with a continuous topological structure. Therefore, our EE-Net combines the semantic features and the geometric constraints, thus well coping with the fracture in segmentation caused by stenoses and improving the segmentation quality of the results.

1) Learning Process of Our EE-Net: As shown in Fig. 2, our EE-Net has two stages: 1) In the Examiner learning stage, to help the E_2-Net focus on the continuous topological structure features of the centerline label, the CGM is used to enhance the centerline label z to get the enhanced centerline label z_M. Then the lumen label y is sent to the E_2-Net as an input, and the output feature map y_1' along with the enhanced centerline z_M are sent to the DO-Layer, and the weighted output $\hat{y_1'}$ and $\hat{z_M}$ are obtained, where $\hat{y_1'}$ and $\hat{z_M}$ respectively represent the prediction result and the label after the class balance. Afterward, the loss $\mathcal{L}_{Examiner}$ is calculated, so that the E_2-Net learns the continuous topological structure features contained in the lumen. 2) In the Examinee learning stage, E_2-Net is fixed in this stage. Image x is sent to the E_1-Net as an input, and the obtained output y_2' is sent to the DO-Layer together with the lumen label y. Then the weighted results $\hat{y_2'}$ and $\hat{y}$ are obtained to calculate the segmentation loss $\mathcal{L}_{Seg}$. In this process, E_1-Net focuses on recognizing lumen-like regions and learning to discard most of the negative samples. Afterwards, the output y_2' is sent to the E_2-Net E_2 as an input, and the obtained output feature map y_3' along with the enhanced centerline label z_M are sent to the DO-Layer to get the weighted results $\hat{y_3'}$ and $\hat{z_M}$ and calculate the constraint loss $\mathcal{L}_{Con}$. In this process, Examiner extracts the continuous structural features from the lumen prediction using the knowledge previously learned, to evaluate and optimize the output results of Examinee, and the learning process is summarized as (Eq. 1, 2):

In the Examiner learning stage:

$$\begin{cases} z_M = CGM(z), \\ y_1' = E_2(y), \\ \hat{y_1'}, \hat{z_M} = DO(y_1', z_M), \\ \mathcal{L}_{Examiner} = CrossEntropy(\hat{y_1'}, \hat{z_M}), \end{cases} \tag{1}$$

In the Examinee learning stage:

$$\begin{cases} y_2' = E_1(x), \\ \hat{y_2'}, \hat{y} = DO(y_2', y), \\ \mathcal{L}_{Seg} = CrossEntropy(\hat{y_2'}, \hat{y}), \\ y_3' = E_2(y_2'), \\ \hat{y_3'}, \hat{z_M} = DO(y_3', z_M), \\ \mathcal{L}_{Con} = CrossEntropy(\hat{y_3'}, \hat{z_M}) \\ \mathcal{L}_{total} = \mathcal{L}_{Seg} + \lambda \mathcal{L}_{Con}, \end{cases} \tag{2}$$

2) Architecture of Our EE-Net: E_1-Net and E_2-Net follow the 3D-Unet framework, where E_1-Net uses 4 max-pooling layers while E_2-Net uses 2 max-pooling layers. Each convolution layer contains a group normalization (GN) and a rectified linear unit (ReLU). The skip connection is used throughout the network.

3) Advantages of Our EE-Net: **1) Customized prior conditions:** Our E_2-Net customizes the prior conditions according to the specific tasks. In our task, to improve the continuity of the segmentation, we use the continuous topological structure features of the centerline as the prior conditions. **2) Segmentation continuity improvement:** Our E_2-Net uses the prior conditions to generate the geometric constraints for the feedback to evaluate and optimize the E_1-Net through the backpropagation. Therefore, the segmentation presents a continuous structure and produces better segmentation quality in regions with stenoses. **3) A plug-and-play module:** Our E_2-Net is a plug-and-play module that will not participate in the test phase. Hence, there are no additional parameters, and the performance is improved under the same computing resources.

B. CGM Module for Features Enhancement

Our proposed Centerline Gaussian Mask (CGM) Module is used in our E_2-Net, which generates a 3D Gaussian mask z_M according to the morphological features of the centerlines, to enhance the continuous topological structure features learning, thus helping the E_2-Net to be more sensitive to the centerline to generate geometric constraints.

1) Implement of Our CGM Module: There are only independent and scattered voxels in one cross-section of the centerline label, and the network is insensitive to a very small number of voxels, which will make the network lose the continuity information of the centerline in the training process.

Formally, it is formulated as (Eq. 3, 4):

$$\mathcal{M}_{ijk} = \max_{n=1,2,\cdots,N} G(i, j, k; x_n, y_n, z_n, \sigma_{w_n}, \sigma_{h_n}, \sigma_{d_n}) \tag{3}$$

$$G(i, j, k; x, y, z, \sigma_w, \sigma_h, \sigma_d) = e^{-(\frac{(i-x)^2}{2\sigma_w^2} + \frac{(j-y)^2}{2\sigma_h^2} + \frac{(k-z)^2}{2\sigma_d^2})} \tag{4}$$

where N is the number of voxels in an image, (i, j, k) represents the coordinates of each voxel, $(x_n; y_n; z_n; w_n; h_n; d_n)$ is the center coordinates, width, height and depth of the nth object, and the variances $(\sigma_{w_n}, \sigma_{h_n}, \sigma_{d_n})$ of the Gaussian mask are proportional to the width, height and depth of individual objects. If these masks have overlaps, we choose the maximum values.

2) Advantages of Our CGM Module: The centerline labels have only one voxel in each cross-section, so the network is easy to ignore such feature information. Therefore, CGM Module generates the 3D Gaussian mask which enhances the features of the centerline in each cross-section and enables the network to be more sensitive to the centerlines and obtain the continuous topological structure features better, thus ensuring the continuity of segmentation.

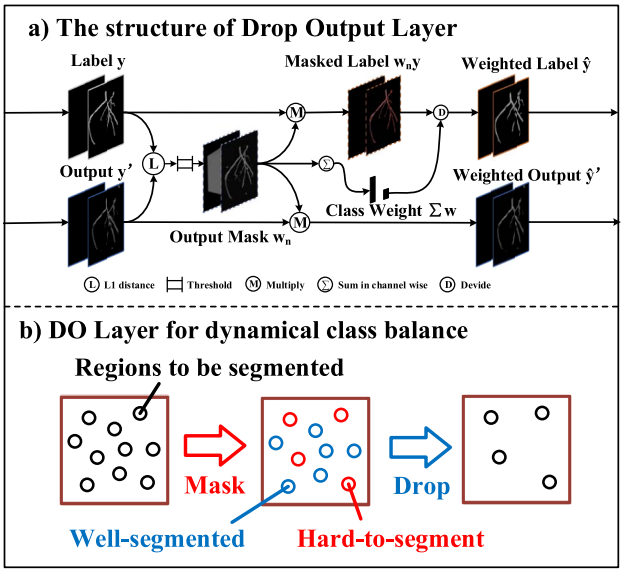

Fig. 3. The DO-Layer. The label y and output y' are the inputs of the DO-Layer. After the processing, weighted label $\hat{y}$ and weighted output $\hat{y}'$ are the outputs. Weighted by the masked label, hard-to-segment regions will have more opportunities to be trained.

C. EE-Learning for Weak Supervision

Our proposed Examinee-Examiner Learning (EE-Learning) is based on our EE-Net. When the weakly labeled data with the centerline labels only is available, the E_2-Net evaluates the segmentation results from the E_1-Net directly using the prior conditions. Therefore, the E_2-Net will guide and constrain the segmentation results from the E_1-Net, thus achieving good results on the dataset with few lumen labels.

1) Learning Process of Our EE-Learning: As shown in Fig. 2, our EE-Net has two stages, but with the addition of a large number of training data without lumen labels, the second stage of EE-Net, Examinee Learning, has been adjusted. For training data without lumen labels, image x is sent to the E_1-Net as an input. However, there is no accurate lumen label to evaluate the predicted result, so the obtained output y_2' is directly sent to the fixed E_2-Net. E_2-Net evaluates the predicted result through the knowledge obtained from previous training. The output feature map y_3' from the E_2-Net along with the enhanced 3D Gaussian mask z_M are sent to the DO-Layer to get the weighted results $\hat{y_3'}$ and $\hat{z_M}$ and calculate the constraint loss $\mathcal{L}_{Con}$.

2) Advantages of Our EE-Learning: Our EE-Learning adapts to datasets with few labels, which has good generalization ability and superior segmentation results on weakly labeled data.

D. DO-Layer for Dynamical Class Balance

1) Structure and Operation of DO-Layer: As shown in Fig. 3, the $L1$ distance of the label y_n and the output y_n' (where n represents the different categories) is calculated, and the binary threshold is used and obtain the output masks w_n which prints out the hard-to-segment regions. Then, the label y_n and predicted y_n' will be masked by the output mask w_n for the masked label $w_n y_n$ and weighted output $w_n y_n'$. Meanwhile, the volume of the remaining regions from the output mask w_n are calculated as a class weight vector $w = \sum_n^N w_n$ to balance the classes (where N represents the total number of categories). Then, the masked label is divided by the class weight vector w for the weighted label $\hat{y_n}$. The output masks w_n can be written as: Eq. 5, and the weighted label $\hat{y_n} = \frac{w_n y_n}{w}$ and the weighted output $\hat{y_n'} = w_n y_n'$ after weighting. Finally, $\hat{y_n}$ and $\hat{y_n'}$ are used to calculate the cross-entropy loss for model optimization.

$$w_n = \begin{cases} |y_n - y_n'|, & |y_n - y_n'| > 0.1 \\ 0, & |y_n - y_n'| \leq 0.1 \end{cases} \tag{5}$$

2) Advantages of DO-Layer: **1) Adapted training process.** DO-Layer dynamically calculates the output mask w according to the optimization in the training process, thus adapting to the whole training. DO-Layer optimizes each region in the early stage of training and gives higher weights to the hard-to-segment regions in the later stage to get more training opportunities. **2) Dynamical class balance.** DO-Layer uses the dynamical weight, which represents the difference between the labels and the outputs. During the training process, when the dominant category changes, the dynamical weight will change according to the change of the dominant category, thus giving the inferior category more training opportunities.

IV. Experiments Configurations

Our experiments involve two datasets, one is **Cardiac CCTA Data** and the other is **ASOCA Data**. **Cardiac CCTA Data** is used to train our EE-Net and show the performance of our model. **ASOCA Data** is used to verify the generalization ability of our method on the cross-dataset tasks.

A. Cardiac CCTA Data

1) Source and Acquisition Protocols: Cardiac CCTA Data is multi-center and multi-device data. CCTA images from 132 patients whose stenoses were found in the major epicardial coronary arteries by CTA were selected. Two hospitals in China (Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai and Beijing Anzhen Hospital, Capital Medical University, Beijing) provided the imaging data. For the data from the first hospital, a 128-slice multidetector CT (Definition AS, Siemens Medical Solutions, Forchheim, Germany) was employed for scanning. For the data from the second hospital, CCTA was performed on three different CT scanners, 256-row detector CT scanner (Revolution CT, GE Healthcare, Milwaukee, USA), 320–detector row CT scanner (Aquilion One; Toshiba, Otawara, Japan), dual-source (Somatom Definition; Siemens, Forchheim, Germany) or dual-source (Somatom Definition Flash; Siemens, Forchheim, Germany) CT scanner.

2) Detailed Information: According to the data of 132 cases used in the experiment, the age of patients ranged from 38 to 83 years old, with an average age of 62.2 years. Among them, there were 99 males and 33 females. In the data, there are 147 lesions of clinical concern, the average quantitative diameter stenosis rate is 45.2%, and the minimum lumen area is 2.13 mm^2. Among them, 60 patients had stenosis with a quantitative diameter stenosis rate of more than 50%.

3) DATA Processing: CCTA data were annotated in an image core lab (CardHemo, Med-X Research Institute, Shanghai Jiao Tong University) by an experienced analyst using a commercialized CCTA analysis software (CtaPlus, version 1.0, Pulse medical imaging technology, Shanghai, China). First, the lumen contours in eight longitudinal cuts through a straightened MPR image along the coronary artery were delineated semi-automatically. Then, a senior imaging expert with 10 years of experience in cardiac imaging performed quality control on the annotated dataset. The original image was 512×512 per slice, with 200 to 500 slices per image. Automated centerline extraction techniques were used in our experiments to extract the centerlines of all data in the preprocessing firstly. The extracted centerlines were checked and adjusted by experts to ensure the accuracy of the centerline if necessary. The coronary lumen boundary was delineated manually by a trained researcher and verified by a clinician. Three-fold cross-validation was adopted. 44 of them were used as a training set and the other 22 as the test set. The remaining 66 images only with the centerlines were used for the weakly supervised training. The region of interest(ROI) was extracted and these heart ROIs were cropped into small 3D patches of $288 \times 288 \times 224$ online before been fed into the network due to the limitation of GPU memory.

B. ASOCA Data

ASOCA Data comes from the 2020 MICCAI Challenge "Automated Segmentation of Coronal Arteries." They provided a training data set of 40 CCTA images with contrast agent showing the coronary arteries, comprising 20 healthy patients and 20 patients with confirmed coronary artery disease. The training data set is used in our paper as the test data to verify the generalization ability of our method on the cross-dataset tasks. Annotations produced by three expert annotators are provided for this data set. Data was collected using retrospective ECG-gated acquisition (GE LightSpeed 64 slice CT Scanner, USA. The time point used for the challenge is late diastole (75% cardiac cycle).

C. Training Strategies and Implementation

Other networks and Our EE-Net were all optimized by Adam with the learning rate of 1×10^{-4}. The training batch size was 1 and the networks were all trained iteratively with 400 epochs. Our implementation used Pytorch.

D. Evaluation Measures

To demonstrate the advantages of our proposed EE-Net, we performed comparative experiments and ablation studies. The comparison methods include the classical segmentation methods such as V-Net [39], 3 D U-Net [40], 3D Attention U-Net [41] and DenseBiasNet [38], the-state-of-art methods for vessel segmentation such as CS^2-Net [22] and weakly-supervised method such as Partial cross-entropy loss [27], [42], Discretely-constrained deep network [28] and AR-SPL [29]. These models were trained on the same dataset with the same implementation and evaluated by the metrics as follows. All the metrics were calculated for each image and the mean values were obtained in the end.

1) Mean Dice Coefficient (Dice): The Dice coefficient was used to evaluate the similarity of the foreground regions in the two images according to Eq. 6.

$$Dice(P, G) = \frac{2\,|P \cap G|}{|P| + |G|} \tag{6}$$

where P represents the segmentation in the predicted result and G represents the ground truth of the coronary lumen.

2) Relavant Dice Coefficient (RDice): Considering some unlabeled side-branches or distal parts of the coronary arteries will be segmented by our method, we propose a relative-dice coefficient (RDice) to evaluate the results accurately. Endings and branches not related to the FFR calculation are not annotated. However, these regions can be segmented by networks. Inspired by [43], we propose RDice (Eq. 7):

$$RDice = \frac{2\,|(P_M) \cap G|}{|P_M| + |G|} \tag{7}$$

where $P_M = P \cap M$. The ground truth G was expanded to a mask M by a morphological expansion operator with a 3D spherical kernel of 8mm diameter. Then the mask M intersects the prediction P to exclude the regions that are not defined in the ground truth G_M.

3) Stenoses Dice Coefficient (SDice): To objectively verify the effectiveness of our results in regions with stenoses, we propose a Stenoses dice coefficient(SDice). The SDice can be written as Eq. 8:

$$SDice = \frac{2\,|(P_S) \cap S|}{|P_S| + |S|} \tag{8}$$

where $P_S = P \cap S$. We use pre-annotated stenoses label S to intersect the predicted result P to include the regions that indicate the regions with stenoses.

4) Overlap Until First Error (OF) and the Overlap (OV): To objectively verify the continuity of the segmentation, the overlap until first error(OF) [43] is used to evaluate the first fracture position of the extracted vessel centerline. Meanwhile, the overlap(OV) [44] is used to evaluate the completeness of the extracted vessel centerline.

5) Hausdorff Distance: Hausdorff Distance [45] is also widely used to describe the similarity between two sets of point sets. For thin and small structures, it is recommended to use the distance-based metric.

V. Results and Discussion

Our proposed Examiner-Examinee Network adapts to the cases with stenoses, datasets with few labels, and ensures the class balance. In this paragraph, we will evaluate and analyze the effectiveness of our proposed EE-Net in four-folds. Firstly, the performance of our proposed method and the classic methods on the 3D coronary lumen segmentation task with stenosis are verified and compared by the following metrics. The visual effects of different methods are simultaneously shown (Sec. V-A). Then, we analyzed the effectiveness of our proposed EE-Net to improve the continuity of the segmentation, the generalization ability of the network, and the

TABLE II

OUR EE-NET HAS ACHIEVED THE MOST COMPETITIVE RESULTS WHERE L INDICATES THE USE OF THE LUMEN LABELS AND C INDICATES THE USE OF THE CENTERLINES

Network	Labeled	Unlabeled	Dice(%)	RDice(%)	SDice(%)	OV(%)	OF			Hausdorff (voxels)
							LAD(%)	LCX(%)	RCA(%)	
3D U-Net [40]	44 L	—	72.7±4.7	76.0±4.2	74.3±7.6	83.5±6.4	75.1±28.9	76.5±24.7	81.6±28.5	3.65±0.99
Coarse-to-fine	44 L	—	73.0±4.2	76.2±3.7	74.8±6.9	84.2±5.5	75.4±27.4	77.6±25.4	83.8±24.6	3.55±0.95
VoxResNet [25]	44 L	—	72.9±4.7	76.1±4.4	73.7±9.6	83.9±6.8	65.7±33.3	73.8±27.1	82.0±25.9	3.80±1.59
V-Net [39]	44 L	—	73.1±4.8	75.9±3.9	75.3±7.0	83.0±6.6	69.7±31.6	76.9±21.4	78.3±31.0	**3.51±1.01**
3D Attention U-Net [41]	44 L	—	73.6±4.5	77.0±3.9	75.7±7.1	85.1±6.3	67.7±29.3	74.6±24.8	77.8±30.3	3.62±1.05
DenseBiasNet [38]	44 L	—	73.1±4.9	76.8±3.9	75.5±7.1	85.6±6.4	71.6±30.8	75.4±26.6	83.7±25.8	3.75±1.12
CS^2-Net [22]	44 L	—	69.6±5.9	72.9±5.4	71.1±7.5	78.2±8.0	55.9±30.4	65.2±24.4	76.6±31.4	4.22±1.46
EE-Net (Ours)	44 (L + C)	—	**75.0±4.1**	**78.3±3.7**	**76.8±6.7**	**85.9±5.7**	**78.1±25.7**	**78.8±23.5**	**83.8±26.9**	3.60±0.99
3D U-Net (Partial CE [27])	44 (L + C)	66 C	74.4±4.0	77.9±4.1	76.6±6.5	84.7±5.3	71.3±31.4	80.0±21.2	81.2±26.4	3.47±1.08
V-Net (Partial CE)	44 (L + C)	66 C	72.7±4.8	76.3±4.3	74.6±6.8	85.5±6.2	72.2±29.2	75.4±26.2	81.9±27.2	3.55±1.03
3D Attention U-Net (Partial CE)	44 (L + C)	66 C	74.1±4.0	77.4±3.4	76.2±6.5	86.6±5.0	72.9±28.8	81.5±21.1	80.1±29.4	3.39±0.90
DenseBiasNet (Partial CE)	44 (L + C)	66 C	74.4±4.3	78.2±3.3	76.4±7.2	86.6±5.3	74.4±28.1	81.7±20.7	81.5±27.5	3.51±1.09
CS^2-Net (Partial CE)	44 (L + C)	66 C	72.1±4.7	76.0±4.2	74.3±7.2	85.7±6.1	67.8±28.5	76.0±24.8	79.3±29.4	3.69±1.23
Discretely-constrained [28]	44 (L + C)	66 C	73.8±3.5	76.7±4.1	75.7±7.7	84.8±6.4	76.0±26.0	78.3±22.8	81.0±27.6	3.35±0.92
AR-SPL [29]	44 (L + C)	66 C	73.5±4.0	76.5±3.6	75.0±7.4	83.4±6.3	73.6±26.0	73.6±27.3	79.6±28.2	3.51±0.94
GAN [34]	44 (L + C)	66 C	72.3±4.6	75.9±4.3	75.0±6.8	83.1±7.1	66.7±31.7	75.2±23.4	83.4±24.8	3.79±0.96
EE-Net (Ours)	44 (L + C)	66 C	**76.2±4.0**	**79.1±3.7**	**77.8±7.0**	**86.9±5.6**	**78.4±28.2**	**81.9±21.1**	**84.2±27.6**	**3.35±0.90**

TABLE III

EACH MODULE IN OUR EE-NET PLAYS A VITAL ROLE. CGM REPRESENTS THE CENTERLINE GAUSSIAN MASK, DO REPRESENTS THE DROP OUTPUT LAYER AND WS REPRESENTS THE WEAK SUPERVISION

Network	CGM	WS	DO	Dice(%)	RDice(%)	SDice(%)
3D U-Net	×	×	×	72.7±4.7	76.0±4.0	74.8±7.2
3D U-Net	×	×	√	74.1±4.5	77.2±4.1	76.2±7.0
EE-Net	√	×	×	73.2±4.0	76.3±3.7	75.4±7.5
EE-Net	√	√	×	74.6±4.0	77.7±3.2	76.3±6.3
EE-Net	√	×	√	75.0±4.1	78.3±3.7	76.8±6.7
EE-Net	√	√	√	**76.2±4.0**	**79.1±3.7**	**77.8±7.0**

help of the Drop Output Layer to deal with the class imbalance (Sec. V-B). Next, we analyzed the performance of our EE-Net on the **ASOCA Data**[1] [46], which shows that our network has strong generalization ability (Sec. V-C). Finally, the parameters related to the framework are also discussed (Sec. V-D).

A. Evaluation Metrics Advantages

Our proposed EE-Net has a strong performance both in visual effects and various indicators.

1) Quantitative Evaluation: The advantages of our method on each metric are demonstrated in Tab. II and the ablation experiments of our method are shown in Tab. III. Our proposed EE-Net achieves the best segmentation results compared with other methods with Dice of 76.2%, RDice of 79.1%, and SDice of 77.8%, and their corresponding standard deviations are 4.0%, 3.7% and 7.0%. Compared with the results of the classical segmentation network with Partial CE, our method achieves at most 3.5% Dice, 2.8% RDice, 3.2% SDice, 2.2% OV, 7.1% OF of LAD, 6.5% OF of LCX and 4.1% OF of RCA improvements and at least 1.8% Dice, 0.9% RDice, 1.2% SDice, 0.3% OV, 4.0% OF of LAD, 0.2% OF of LCX and 2.3% OF of RCA improvements.

[1] https://asoca.grand-challenge.org/Home/

2) Qualitative Evaluation: Our proposed EE-Net has powerful visual superiority in any aspect (Fig. 4). The case1, case2, case3, and case4 respectively illustrate the competitive performance of our EE-Net from four angles: overall segmentation effect, the case with stenoses, hard-to-segment regions, and false segmentation. Compared with the ground truth, our EE-Net achieves the best visual superiority in these cases. (1) Case1 shows the overall segmentation effect. The yellow box shows the fracture in segmentation, especially at the bifurcation and the end of vessels. (2) Case2 shows the segmentation effect on cases with stenosis. The yellow box indicates that our EE-Net segments the complete trunk in the regions with stenoses, while other methods have different degrees of fracture in the trunk. (3) Case 3 shows the segmentation effect in hard-to-segment regions. The yellow box shows the continuity of our EE-Net in segmentation of hard-to-segment regions, such as at the bifurcation and the end of vessels. (4) Case 4 shows the situation of false segmentation. Due to the influence of other vessel-like structures, other methods are insensitive to distinguish these structures, which will lead to serious false segmentation. However, our EE-Net has the ability to segment the coronary artery lumen accurately.

B. Ablation Experiment Analysis

As shown in Tab. III, the ablation experiments prove the importance of our EE-Net, EE-Learning strategy for weakly supervised images and our DO Layer. (1) To prove the effectiveness of the DO layer. The first and second lines of the table and the fourth and sixth lines of the table show the performance of the DO layer in our model. With the addition of the DO Layer, Dice increased by 1.4% and 1.6%, RDice increased by 1.2% and 2.0%, and SDice increased by 1.4% and 1.4%, respectively. The results show that DO Layer plays a key role in the model, helping the network to distinguish hard-to-segment regions and keep class balance. (2) To prove the effectiveness of EE-Net in full supervision. The first and third lines of the table show the performance of the EE-Net in full supervision. With the addition of the

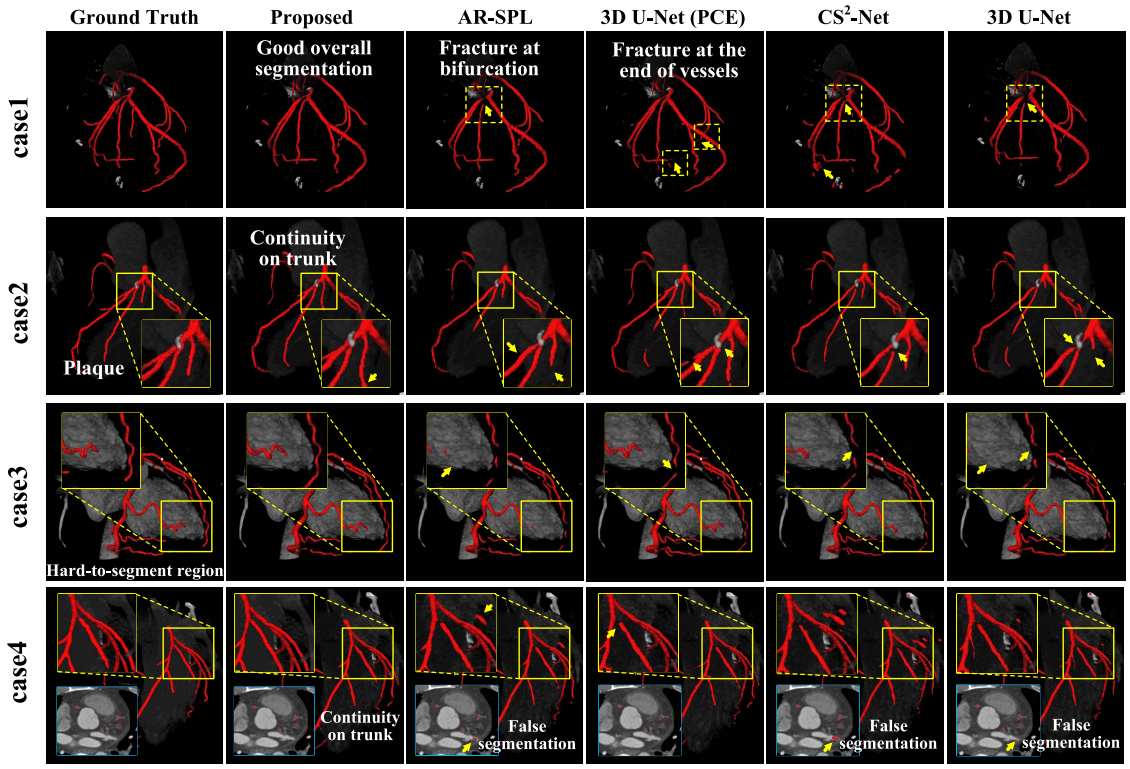

Fig. 4. Our proposed framework realizes accurate coronary lumen segmentation and has powerful visual superiority, especially in any aspect. The case1, case2, case3, and case4 respectively illustrate the competitive performance of our EE-Net from four angles: overall segmentation effect, the case with stenoses, hard-to-segment regions, and false segmentation.

EE-Net, Dice increased by 0.4%, RDice increased by 0.3% and SDice increased by 0.6%, respectively. The results show that EE-Net provides additional continuous topological structure information from the centerline for the data with lumen labels in full supervision. (3) To prove the effectiveness of EE-Net for weak supervision. The third and fourth lines of the table and the fifth and sixth lines of table show the performance of the EE-Net for weakly supervised labels. With the addition of the EE-Net, Dice increased by 1.4% and 1.2%, RDice increased by 1.4% and 0.8% and SDice increased by 0.9% and 1.0%, respectively. The results show that EE-Net and EE-Learning make better use of the weakly supervised images only with centerline labels.

1) EE-Net for Cases With Stenoses: To demonstrate the segmentation performance of our EE-Net in regions with stenoses more clearly, Fig. 6 shows the MPR view of a segmentation result for an entire artery. Compared with the ground truth, our proposed EE-Net has a good continuous segmentation effect in regions with stenoses, while other methods have different degrees of fracture problems.

2) EE-Net for Partial Annotations: Our proposed EE-Learning shows good generalization performance when the annotations are incomplete. To simulate the partially annotated scenario, random destruction was performed on the lumen labels of the labeled data. We destroyed 20%, 40%, and 60% of the original lumen labels respectively. For comparison, we used the backbone network 3D U-Net and the network with the Partial CE as the weakly supervised model. Three-fold cross-validation was used to verify the results of our EE-Net with destroyed labeled data. Our proposed EE-Learning shows better results than other methods when the completeness of the lumen label is decreased (Fig. 7). As the completeness of the lumen label decreases, the comparison of Dice between our EE-Net and other methods shows the good generalization ability of our method. When the completeness of the lumen label is decreased to 60% of the original lumen label, the Dice of our EE-Net is 5% higher than 3D U-Net and 3.4% higher than 3D U-Net with Partial CE. With the help of EE-Learning, the network obtains weak supervision information from the centerline, thus obtaining relatively better results on incomplete lumen labels.

3) EE-Learning for Dataset With Fewer Labels: Our proposed EE-Learning shows good stability when the number of labeled training data decreases (Fig. 8). The Dice is compared when the amount of labeled training data decreases gradually. As the number of labeled data decreases, the comparison of Dice between our EE-Net and U-Net shows the good generalization ability of our method. When the amount of labeled data is decreased to 5 images, the Dice of our EE-net is 15.4% higher than U-Net. EE-Net is more stable than U-Net in the process of data reduction. U-Net declined by 18.8%, while our EE-Net only declined by 6.9%. With the help of EE-Learning, the network obtains weak supervision information from the centerline, thus reducing the dependence on the amount of labeled data.

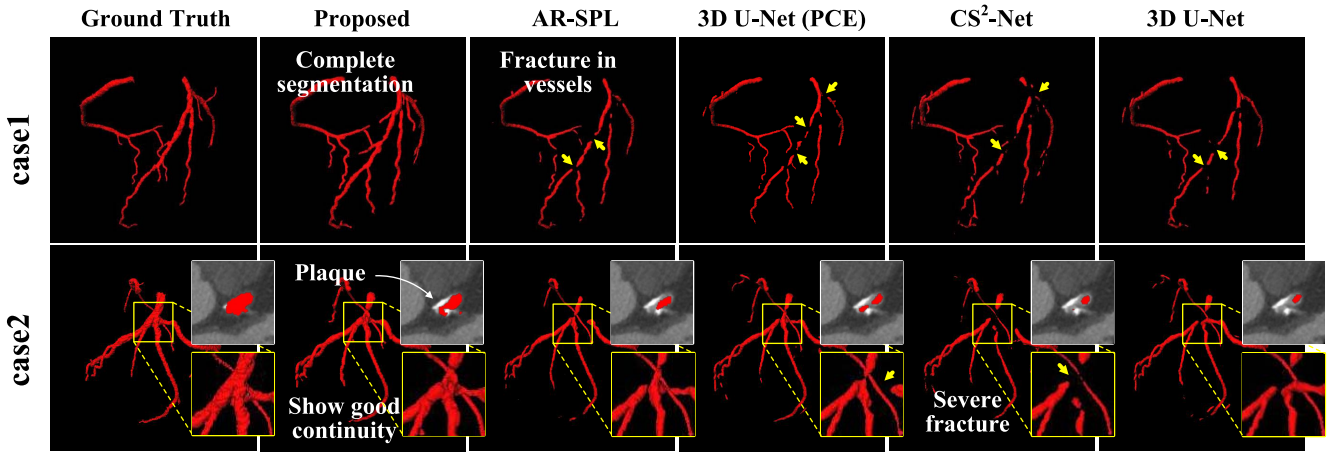

Fig. 5. Our EE-Net shows a good segmentation performance on the external testing dataset **ASOCA data**. The results demonstrate that our EE-Net has a strong generalization ability. The case1 and case2 respectively illustrate the competitive performance of our EE-Net from two angles: overall segmentation effect and case with stenoses.

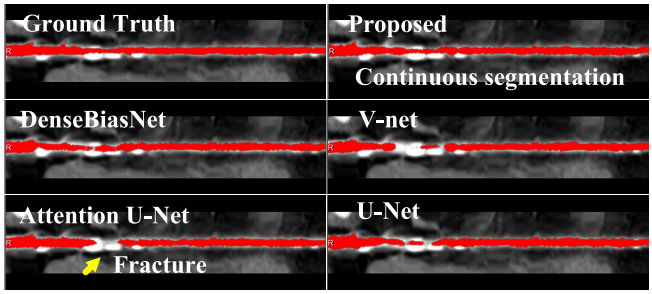

Fig. 6. MPR view of a segmentation result for an entire artery. Compared with the ground truth, our proposed EE-Net has a good continuous segmentation effect in regions with stenoses, while other methods have different degrees of fracture problems.

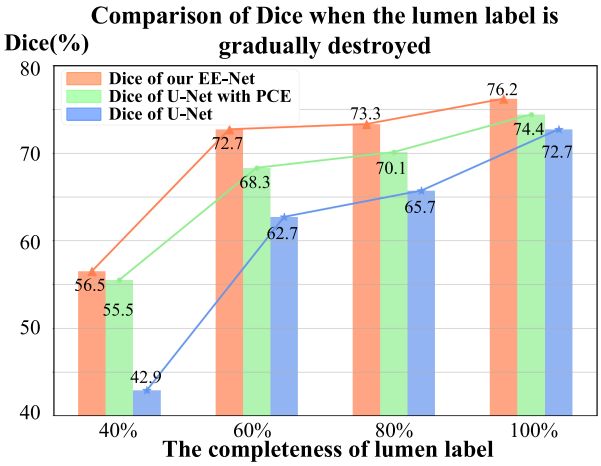

Fig. 7. When the lumen label is gradually destroyed, the Dice comparison between our EE-Net and U-Net shows that our method achieves relatively better results on incomplete lumen labels. When the completeness of the lumen label is decreased to 60% of the original lumen label, the Dice of our EE-Net is 10.0% higher than U-Net. When the completeness of the lumen label is decreased to 40% of the original lumen label, the neural network has no complete generalization ability to obtain better segmentation results.

4) CGM Module for Centerline Features Enhancement: The centerline features enhancement (Tab. III) of our CGM module plays a significant role when the network uses the features of the centerline. The introduction of CGM has brought about an increase of 0.6% in the SDice index which reflects the segmentation quality in regions with stenoses.

5) Drop Output Layer for Class Balance: As revealed in Fig. 9, our proposed Drop Output Layer drops well-segmented regions in the output map and weights the classes dynamically to give the hard-to-segment regions more training opportunities thus balancing classes. At the initial stage of training, the large specific gravity means that the background and the lumen to be segmented are in a state of class imbalance. With the help of the DO-Layer, in the first 100 rounds, the proportion dropped rapidly approaching 1. In the later training process, the proportion gradually stabilized.

C. Evaluation on ASOCA Data

Quantitative Evaluation: As demonstrated in Tab. IV, our EE-Net has a good generalization ability in **ASOCA Data**. It is worth noting that, our proposed EE-Net shows good generalization performance when tested on the training dataset of **ASOCA data** compared with other methods with Dice of 72.7%, RDice of 76.3%, OV of 87.6%, LAD OF of 91.1%, LCX OF of 82.2%, RCA OF of 91.4% and Hausdorff Distance of 3.71 voxels.

To evaluate our EE-Net objectively, the evaluation on different cases proves that our EE-Net achieves competitive results not only in diseased cases but also in healthy cases.

1) Results in Healthy Cases: Dice and RDice are calculated on healthy data (Tab. V). Our EE-Net shows good generalization ability with the Dice of 75.3% which increased by 2.4%, RDice coefficient of 79.0% which increased by 3.4% compared with 3D U-Net [40]. This result revealed our EE-Net achieves great competitive results on healthy cases.

2) Results in Diseased Cases: Dice and RDice are also calculated on the diseased data (Tab. V). Our EE-Net shows good segmentation continuity and generalization ability with

TABLE IV

OUR EE-NET SHOWS GOOD GENERALIZATION PERFORMANCE WHEN TESTED ON THE TRAINING DATASET OF **ASOCA DATA**

Network	Labeled	Unlabeled	Dice(%)	RDice(%)	OV(%)	OF			Hausdorff (voxels)
						LAD(%)	LCX(%)	RCA(%)	
3D U-Net [40]	44	—	69.7±6.1	72.1±7.0	81.1±10.8	77.2±31.6	75.7±29.2	88.7±22.9	4.07±1.07
V-Net [39]	44	—	66.5±8.3	69.4±9.1	77.5±10.3	69.1±33.1	63.3±31.9	81.3±23.2	4.27±1.20
3D Attention U-Net [41]	44	—	69.5±6.3	72.8±7.0	82.9±9.6	83.5±32.4	77.5±31.9	90.4±16.5	3.93±1.32
DenseBiasNet [38]	44	—	68.7±6.9	71.4±8.8	81.7±10.9	72.5±34.0	71.3±30.5	85.5±21.7	4.14±1.22
CS^2-Net [22]	44	—	65.4±7.9	68.8±8.4	77.2±11.0	61.1±35.2	56.3±33.2	84.6±22.3	4.60±1.47
3D U-Net (Partial CE [27])	44	66	70.5±5.6	74.2±6.6	86.1±9.5	81.5±30.2	83.2±25.6	87.2±21.9	3.85±1.04
V-Net (Partial CE)	44	66	67.9±5.4	72.3±6.1	85.9±7.5	80.3±30.0	76.1±30.3	90.1±17.9	4.33±1.27
3D Attention U-Net (Partial CE)	44	66	70.9±4.9	74.2±6.1	86.2±8.1	83.1±28.6	80.7±25.3	90.7±16.9	3.78±1.04
DenseBiasNet (Partial CE)	44	66	71.5±4.6	75.5±5.6	87.1±8.3	81.2±30.7	82.4±25.0	90.2±20.2	3.98±1.10
CS^2-Net (Partial CE)	44	66	66.9±6.6	71.6±7.3	85.1±8.4	70.1±33.9	71.9±32.6	86.1±22.7	4.41±1.30
Discretely-constrained [28]	44	66	70.4±5.0	73.6±6.6	85.4±8.1	80.0±31.7	80.8±27.2	90.1±17.8	3.70±1.04
AR-SPL [29]	44	66	70.7±5.7	73.8±6.6	82.2±10.1	80.6±28.4	71.7±27.9	85.5±21.1	3.99±1.08
EE-Net (Ours)	44	66	**72.7±5.0**	**76.3±5.6**	**87.6±7.2**	**91.1±19.3**	**82.2±24.3**	**91.4±20.1**	**3.71±1.03**

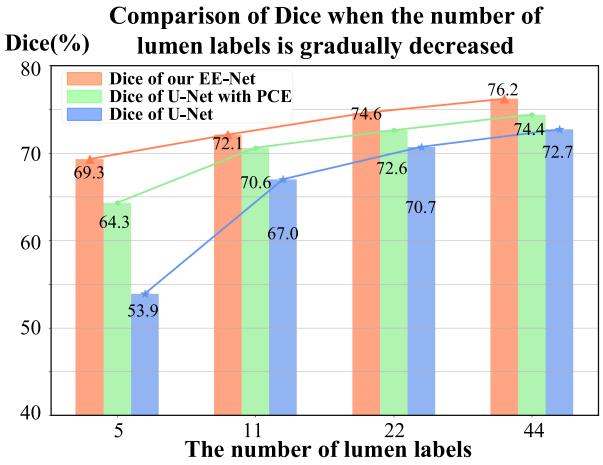

Fig. 8. As the number of labeled data decreases, the comparison of Dice between our EE-Net and U-Net shows the generalization ability of our method. When the amount of labeled data is decreased to 5 images, the Dice of our EE-Net is 15.4% higher than U-Net. EE-Net is more stable than U-Net in the process of data reduction. U-Net declined by 18.8%, while our EE-Net only declined by 6.9%.

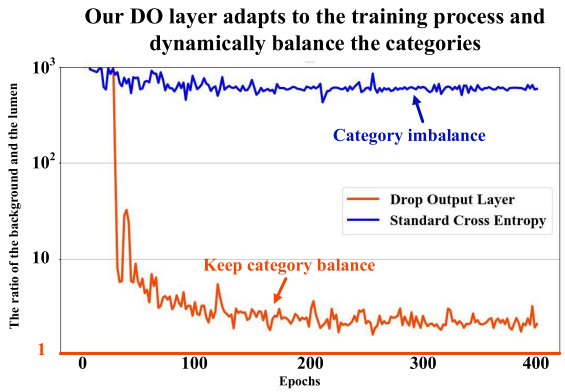

Fig. 9. The ratio of the background and the lumen in regions concerned by the network during the training process shows the importance of our DO Layer for dynamically balancing classes. When the DO layer is not used, the ratio fluctuates around 1000. After the DO layer is used, the ratio is steadily lower than 10, and gradually approaches 1 in the later stage of training, which means class balance is realized.

the Dice of 70.0% which increased by 4.2%, RDice coefficient of 73.5% which increased by 4.9% compared with 3D U-Net [40]. Stenosis brings great challenges to our task, but this result reveals our EE-Net has great potential to realize accurate coronary lumen segmentation in cases with stenoses.

Qualitative Evaluation: Our EE-Net shows good generalization performance when tested on the training dataset of ASOCA Data (Fig. 5). The case1 and case2 respectively illustrate the competitive performance of our EE-Net from two angles: overall segmentation effect and case with stenoses. Compared with the ground truth, our proposed EE-Net achieves the best segmentation in these cases. (1) Case1 shows the overall segmentation effect. Compared with other methods, our EE-Net has the best integrity in segmentation results, especially in some hard-to-segment regions, such as bifurcation and regions with stenoses. (2) Case2 shows the segmentation effect on cases with stenosis. The white box shows the stenoses. The yellow box indicates that our EE-Net segments the complete trunk in the regions with stenoses, while other methods have different degrees of fracture in the trunk.

TABLE V

OUR EE-NET HAS A SUPERIOR GENERALIZATION ABILITY ON BOTH HEALTHY DATA AND DISEASED DATA

Cases	Network	Dice(%)	RDice(%)
Healthy	3D U-Net [40]	72.9±3.8	75.6±4.6
	V-Net [39]	68.5±8.6	71.7±9.2
	3D Attention U-Net [41]	72.9±4.4	76.4±4.8
	DenseBiasNet [38]	71.8±4.6	75.7±5.2
	CS^2-Net [22]	67.7±7.3	71.4±7.8
	Discretely-constrained [28]	72.9±4.4	76.1±5.0
	AR-SPL [29]	73.7±3.8	77.1±4.3
	EE-Net (Ours)	**75.3±3.1**	**79.0±3.8**
Diseased	3D U-Net [40]	65.8±6.2	68.6±7.1
	V-Net [39]	66.2±6.2	69.2±7.1
	3D Attention U-Net [41]	64.5±7.6	67.2±8.4
	DenseBiasNet [38]	65.5±7.4	68.9±8.4
	CS^2-Net [22]	63.0±7.5	66.2±8.1
	Discretely-constrained [28]	66.2±6.2	71.1±5.5
	AR-SPL [29]	67.7±6.0	70.5±6.8
	EE-Net (Ours)	**70.0±5.2**	**73.5±5.8**

TABLE VI

HYPER-PARAMETER ANALYSIS. THE PARAMETER OF BACKPROPAGATION LOSS (λ) IS STUDIED TO ANALYZE THE EFFECTS

λ	Dice(%)	RDice(%)	SDice(%)
10	74.1±4.6	78.0±3.6	74.5±7.8
5	75.0±4.1	77.7±3.4	75.8±7.7
1	**76.2±4.0**	**79.1±3.7**	**77.8±7.0**
0.5	75.7±3.9	78.8±3.4	77.2±7.1
0.1	75.7±4.0	79.1±3.2	77.7±6.3

D. Parameters Analysis

As shown in Tab. VI, we performed ablation experiments on the parameter of backpropagation loss(λ) to analyze the effects of the parameter. The results show that our EE-Net achieves the best performance when $\lambda = 1$.

VI. CONCLUSION

In this study, we proposed a novel weakly supervised model, Examinee-Examiner Network, to achieve accurate coronary lumen segmentation with stenoses for the first time. 1) We proposed an EE-Net to cope with the fracture in segmentation caused by stenoses by combining the semantic features of the lumen and the geometric constraints of the continuous topology obtained from the centerline. A CGM Module was proposed to deal with the insensitiveness of the network to the continuity topology information of the centerline by generating the Gaussian mask to enhance the features of the centerline. 2) We proposed a weakly supervised learning strategy, EE Learning, based on our EE-Net to handle the situation where only weakly labeled data is available by using our EE-Net to guide and constrain the segmentation with the customized prior conditions. 3) We proposed a general network layer, DO Layer, to adapt to the class imbalance by dropping well-segmented regions and weights the classes dynamically.

REFERENCES

[1] B. S. Ko *et al.*, "Noninvasive CT-derived FFR based on structural and fluid analysis," *JACC, Cardiovascular Imag.*, vol. 10, no. 6, pp. 663–673, Jun. 2017.

[2] J. K. Min *et al.*, "Diagnostic accuracy of fractional flow reserve from anatomic CT angiography," *J. Amer. Med. Assoc.*, vol. 308, no. 12, pp. 1237–1245, 2012.

[3] S. A. Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: A review," *Artif. Intell. Rev.*, vol. 54, no. 1, pp. 137–178, 2020.

[4] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, 2018.

[5] S. Afshari, A. Bentaieb, G. Hamarneh, and Z. MiriKharaji, "Weakly supervised fully convolutional network for PET lesion segmentation," *Proc. SPIE Med. Imag. Image Process.*, vol. 10949, Mar. 2019, Art. no. 109491K.

[6] Z. Mirikharaji, Y. Yan, and G. Hamarneh, "Learning to segment skin lesions from noisy annotations," in *Domain Adaptation and Representation Transfer and Medical Image Learning With Less Labels and Imperfect Data*, vol. 11795. Cham, Switzerland: Springer, 2019, pp. 207–215.

[7] J. M. Wolterink, R. W. van Hamersvelt, M. A. Viergever, T. Leiner, and I. Išgum, "Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier," *Med. Image Anal.*, vol. 51, pp. 46–60, Jan. 2019.

[8] G. Yang *et al.*, "Automatic centerline extraction of coronary arteries in coronary computed tomographic angiography," *Int. J. Cardiovascular Imag.*, vol. 28, no. 4, pp. 921–933, Apr. 2012.

[9] M. A. Gülsün, G. Funka-Lea, P. Sharma, S. Rapaka, and Y. Zheng, "Coronary centerline extraction via optimal flow paths and CNN path pruning," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 317–325.

[10] D. Lesage, E. D. Angelini, I. Bloch, and G. Funka-Lea, "A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes," *Med. Image Anal.*, vol. 13, no. 6, pp. 819–845, Aug. 2009.

[11] Y. Chen *et al.*, "Curve-like structure extraction using minimal path propagation with backtracking," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 988–1003, Feb. 2016.

[12] H. Sekiguchi, N. Sugimoto, S. Eiho, T. Hanakawa, and S. Urayama, "Blood vessel segmentation for head MRA using branch-based region growing," *Syst. Comput. Jpn.*, vol. 36, no. 5, pp. 80–88, May 2005.

[13] R. Manniesing and W. Niessen, "Local speed functions in level set based vessel segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2004, pp. 475–482.

[14] J. Mille and L. Cohen, "Deformable tree models for 2D and 3D branching structures extraction," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2009, pp. 149–156.

[15] R. Toledo *et al.*, "Tracking elongated structures using statistical snakes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2001, pp. 157–162.

[16] F. Chen, Y. Li, T. Tian, F. Cao, and J. Liang, "Automatic coronary artery lumen segmentation in computed tomography angiography using paired multi-scale 3D CNN," in *Proc. SPIE Med. Imag. Biomed. Appl. Mol., Struct., Funct. Imag.*, vol. 10578, pp. 652–658, Mar. 2018.

[17] P. Moeskops *et al.*, "Deep learning for multi-task medical image segmentation in multiple modalities," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 478–486.

[18] W. Huang *et al.*, "Coronary artery segmentation by deep learning neural networks on computed tomographic coronary angiographic images," in *Proc. 40th Annu. Int. Conf. Eng. Med. Biol. SoC. (EMBC)*, Jul. 2018, pp. 608–611.

[19] B. Kong *et al.*, "Learning tree-structured representation for 3D coronary artery segmentation," *Comput. Med. Imag. Graph.*, vol. 80, Mar. 2020, Art. no. 101688.

[20] J. Gu, Z. Fang, Y. Gao, and F. Tian, "Segmentation of coronary arteries images using global feature embedded network with active contour loss," *Comput. Med. Imag. Graph.*, vol. 86, Dec. 2020, Art. no. 101799.

[21] T. J. Jun., J. Kweon, Y.-H. Kim, and D. Kim, "T-Net: Nested encoder–decoder architecture for the main vessel segmentation in coronary angiography," *Neural Netw.*, vol. 128, pp. 216–233, Aug. 2020.

[22] L. Mou *et al.*, "CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101874.

[23] Y. He *et al.*, "Meta grayscale adaptive network for 3D integrated renal structures segmentation," *Med. Image Anal.*, vol. 71, Jul. 2021, Art. no. 102055.

[24] J. M. Wolterink, T. Leiner, and I. Išgum, "Graph convolutional networks for coronary artery segmentation in cardiac CT angiography," in *Graph Learning in Medical Imaging*. Cham, Switzerland: Springer, 2019, pp. 62–69.

[25] H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, "VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images," *NeuroImage*, vol. 170, pp. 446–455, Apr. 2017.

[26] Z. Tu, "Auto-context and its application to high-level vision tasks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[27] H. Kervadec, J. Dolz, M. Tang, E. Granger, Y. Boykov, and I. B. Ayed, "Constrained-CNN losses for weakly supervised segmentation," *Med. Imag. Analys.*, vol. 54, pp. 88–99, May 2019.

[28] J. Peng, H. Kervadec, J. Dolz, I. Ben Ayed, M. Pedersoli, and C. Desrosiers, "Discretely-constrained deep network for weakly supervised segmentation," *Neural Netw.*, vol. 130, pp. 297–308, Oct. 2020.

[29] J. Zhang, G. Wang, H. Xie, S. Zhang, N. Huang, S. Zhang, and L. Gu, "Weakly supervised vessel segmentation in X-ray angiograms by self-paced learning from noisy labels with suggestive annotation," *Neurocomputing*, vol. 417, pp. 114–127, Dec. 2020.

[30] X. Ma, Z. Ji, S. Niu, T. Leng, D. L. Rubin, and Q. Chen, "MS-CAM: Multi-scale class activation maps for weakly-supervised segmentation of geographic atrophy lesions in SD-OCT images," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 12, pp. 3443–3455, Dec. 2020.

[31] F. Meng, K. Luo, H. Li, Q. Wu, and X. Xu, "Weakly supervised semantic segmentation by a class-level multiple group cosegmentation and foreground fusion strategy," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4823–4836, Dec. 2020.

[32] H. Kervadec, J. Dolz, S. Wang, E. Granger, and I. B. Ayed, "Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision," in *Proc. Med. Imag. Deep Learn.*, 2020, pp. 365–381.

[33] Y.-C. Chen, Y.-Y. Lin, M.-H. Yang, and J.-B. Huang, "Show, match and segment: Joint weakly supervised learning of semantic matching and object co-segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3632–3647, Oct. 2021.

[34] N. Souly, C. Spampinato, and M. Shah, "Semi and weakly supervised semantic segmentation using generative adversarial network," 2017, *arXiv:1703.09695*.

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland: Springer, 2015, pp. 234–241.

[36] C. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2017, pp. 240–248.

[37] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[38] Y. He *et al.*, "DPA-DenseBiasNet: Semi-supervised 3D fine renal artery segmentation with dense biased network and deep priori anatomy," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 139–147.

[39] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[40] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 424–432.

[41] O. Oktay *et al.*, "Attention U-Net: Learning where to look for the pancreas," *CoRR*, vol. abs/1804.03999, pp. 1–10, Apr. 2018.

[42] D. Pathak, P. Krahenbuhl, and T. Darrell, "Constrained convolutional neural networks for weakly supervised segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1796–1804.

[43] M. Schaap *et al.*, "Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms," *Med. Image Anal.*, vol. 13, no. 5, pp. 701–714, 2009.

[44] H. Kirişli *et al.*, "Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography," *Med. Image Anal.*, vol. 17, no. 8, pp. 859–876, 2013.

[45] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, no. 1, pp. 1–28, 2015.

[46] R. Gharleghi, D. G. Samarasinghe, P. A. Sowmya, and D. S. Beier. (Mar. 2020). *Automated Segmentation of Coronary Arteries*. [Online]. Available: https://doi.org/10.5281/zenodo.3819799