Cardiac Adipose Tissue Segmentation via Image-Level Annotations

Ziyi Huang, Yu Gan, Theresa Lye, Yanchen Liu, Haofeng Zhang, Andrew Laine, Elsa Angelini, and Christine Hendon

Abstract— **Automatically identifying the structural substrates underlying cardiac abnormalities can potentially provide real-time guidance for interventional procedures. With the knowledge of cardiac tissue substrates, the treatment of complex arrhythmias such as atrial fibrillation and ventricular tachycardia can be further optimized by detecting arrhythmia substrates to target for treatment (i.e., adipose) and identifying critical structures to avoid. Optical coherence tomography (OCT) is a real-time imaging modality that aids in addressing this need. Existing approaches for cardiac image analysis mainly rely on fully supervised learning techniques, which suffer from the drawback of workload on labor-intensive annotation process of pixel-wise labeling. To lessen the need for pixel-wise labeling, we develop a two-stage deep learning framework for cardiac adipose tissue segmentation using image-level annotations on OCT images of human cardiac substrates. In particular, we integrate class activation mapping with superpixel segmentation to solve the sparse tissue seed challenge raised in cardiac tissue segmentation. Our study bridges the gap between the demand on automatic tissue analysis and the lack of high-quality pixel-wise annotations. To the best of our knowledge, this is the first study that attempts to address cardiac tissue segmentation on OCT images via weakly supervised learning techniques. Within an in-vitro human cardiac OCT dataset, we demonstrate that our weakly supervised approach on image-level annotations achieves comparable performance as fully supervised methods trained on pixel-wise annotations.**

Index Terms— **Optical coherence tomography, cardiac tissue analysis, deep learning, image segmentation, weakly supervised learning.**

I. INTRODUCTION

Ziyi Huang, Theresa Lye, Yanchen Liu, and Christine Hendon, are with the Department of Electrical Engineering, Columbia University, New York, NY, 10027, USA

Yu Gan is with the Department of Biomedical Engineering, Stevens Institute of Technology, Hoboken, NJ, 07030, USA

Haofeng Zhang is with the Department of Industrial Engineering and Operations Research, Columbia University, New York, NY, 10027, USA

Andrew Laine is with the Department of Biomedical Engineering, Columbia University, New York, NY, 10027, USA

Elsa Angelini is with the Department of Biomedical Engineering, Columbia University, New York, NY, 10027, USA; the NIHR Imperial Biomedical Research Centre and ITMAT Data Science Group, Imperial College London, London, UK; and the Telecom Paris, LTCI, Institut Polytechnique de Paris, France

C ARDIOVASCULAR disease is the leading cause of death in the United States, with atrial fibrillation alone affecting at least 2.3 million people [29]. Treatment of complex arrhythmias such as atrial fibrillation and ventricular tachycardia is through catheter ablation, which directly destroys the cardiac substrates that cause irregular impulse propagation. However, this treatment is sub-optimal, due to the lack of capability to accurately identify optimal ablation targets. With the knowledge of patients' heart structure, the ablation strategy can be further optimized by avoiding critical structures and identifying arrhythmia substrates, such as areas with increased amounts of adipose tissues. Recent work has shown that an increased amount of adipose tissues within the myocardium is a substrate for cardiac arrhythmias [7], [8], [12], [46].

Optical coherence tomography (OCT) is a non-destructive optical imaging modality that could provide an ideal balance between a penetration depth of 2 mm with a resolution of 4-10 μm [40]. Recent advances have demonstrated the capability of OCT on capturing myocardial structures such as Purkinje network [55], atrial ventricular nodes [20], sinoatrial nodes [5], and myofiber organization [18]. In addition, it can be used to resolve critical tissue substrates of arrhythmias, such as fibrosis and adipose tissues [38]. With the development of OCT-integrated catheters [14], OCT can image the heart wall in real time through percutaneous access [51], which holds promise to aid catheter ablation.

To benefit from the real-time capacity of OCT imaging, analysis of OCT images is expected to be automated for timely decision making. Evaluation of adipose tissue distribution within a human atrial sample requires pixel-wise analysis of large volumetric datasets [15]. Manually annotating adipose tissues within a single OCT volumes can take a well-trained annotator over 10 hours. Therefore, automated identification of cardiac tissues, especially adipose tissue, in OCT images is greatly needed.

Current automated analysis on cardiac OCT images is mostly based on fully supervised learning models [22], [23], [34]. These models were limited and suffered from the drawback of manual workload in the labeling process. To avoid overfitting, a large amount of data is required to support the model training. For segmentation tasks, the labeling process is extremely time-consuming and has limited accuracy. Moreover, OCT images are volumetric, adding an additional challenge to labeling. Thus, automatic analysis with weakly supervised learning models is of great interest.
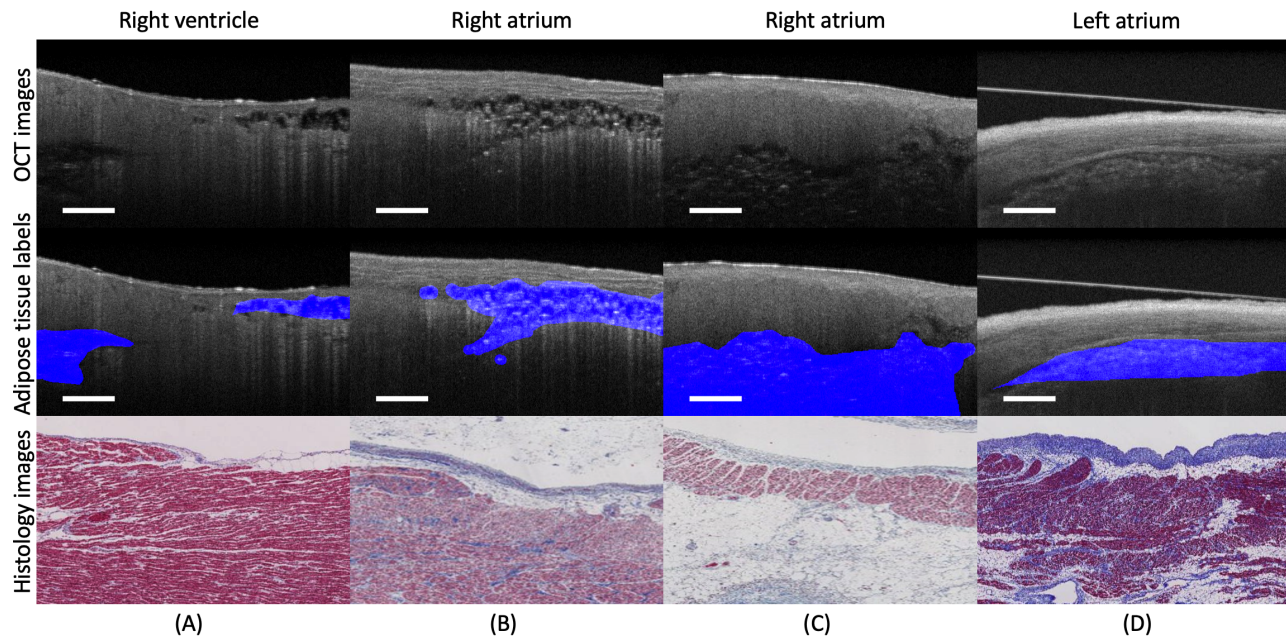
Fig. 1. Representative OCT images from cardiac dataset. Sample (A) is obtained from right ventricle. Sample (B) and sample (C) are obtained from right atrium. Sample (D) is obtained from left atrium submerged in PBS solution. The features of adipose tissue present great variations among different locations and imaging conditions. The unclear boundary and irregular shape of adipose tissues add unique challenges for automated segmentation. Scale bar: $500 \ \mu m$.

Although recent studies have conducted retinal OCT analysis [52], [56], transferring OCT retinal segmentation to cardiac solution for weakly supervised cardiac OCT segmentation is still elusive for three reasons. First, cardiac adipose and fibrosis tissues can appear in multiple sub-regions with irregular shapes and infiltrating patterns. Thus, cardiac OCT images are more complicated than retinal substrates with rather regular layered structures. Second, boundaries between cardiac substrates are more blurry than between the retinal layers. Third, cardiac substrates have a larger variance among patients than retinal tissues.

In this study, we present a weakly supervised learning framework for cardiac tissue segmentation using image-level labels. Our training approach has two stages, namely pseudo label generation and segmentation network training. We first use the class activation map (CAM) results obtained from a binary classification network to generate adipose location seeds. Then, we develop a superpixel-based segmentation algorithm to generate pseudo labels followed by segmentation training. Our contributions are as follows:
(1) We propose a weakly supervised learning framework for cardiac tissue segmentation. Our model is trained without the need of pre-training or domain adaptive learning.
(2) We combine CAM with superpixel segmentation to effectively address the tissue segmentation challenges caused by irregular shape and blurry boundary in cardiac OCT images.
(3) We evaluate our approach on a human cardiac dataset and demonstrate that our weakly supervised model achieves comparable performance with fully supervised algorithms.

II. RELATED WORK

Regarding tissue analysis on cardiac OCT images, [16] imaged and analyzed features on dense collagen, loose collagen, fibrotic myocardium, normal myocardium, and adipose tissue for automatic classification. In [42], segmentation was obtained from the variance map through compressive sensing reconstruction. In [38], the distributions of adipose tissues and fiber orientations were retracted and mapped throughout human left atrium, while in [6], the visualization of cardiac fibers in the atrium, ventricle, atrioventricular node, and sinoatrial node were presented. Overall, conventional cardiac OCT image analysis relies on handcrafted features for tissue characterization or fiber orientation-based methods to focus on myofibers.

Superpixel is a classic unsupervised segmentation method that has been widely used in biomedical data analysis. Without requirement on pre-annotated training sets, the superpixel based methods group the pixels into homogeneous clusters according to the similarity among pixels. [54] combined superpixel with LogitBoost adaptive boosting to detect glaucomatous damage in 3D OCT images. In [44], the superpixel technique was applied to generate the flexible kernels of local statistics on the Jones matrix-based polarization sensitive OCT.

Deep learning approaches have achieved great success in OCT image segmentation tasks [19], [25], [32], [36], [41], [43], [48]. [47] developed a fully convolutional network with Gaussian process based post processing for retinal OCT segmentation. [13] proposed a novel framework that combined a hybrid convolutional neural network and graph search method for retinal layer boundary detection. [4] developed a fully convolutional network-based AV-Net for artery-vein classification. Their model contained a multi-modal training process that involved both en-face OCT and optical coherence tomography angiography (OCTA) to provide the intensity and geometric profiles. [23] trained a fully supervised segmentation network for cardiac tissue segmentation and used model uncertainty to estimate tissue heterogeneity. Existing work mainly relies upon fully supervised learning techniques.
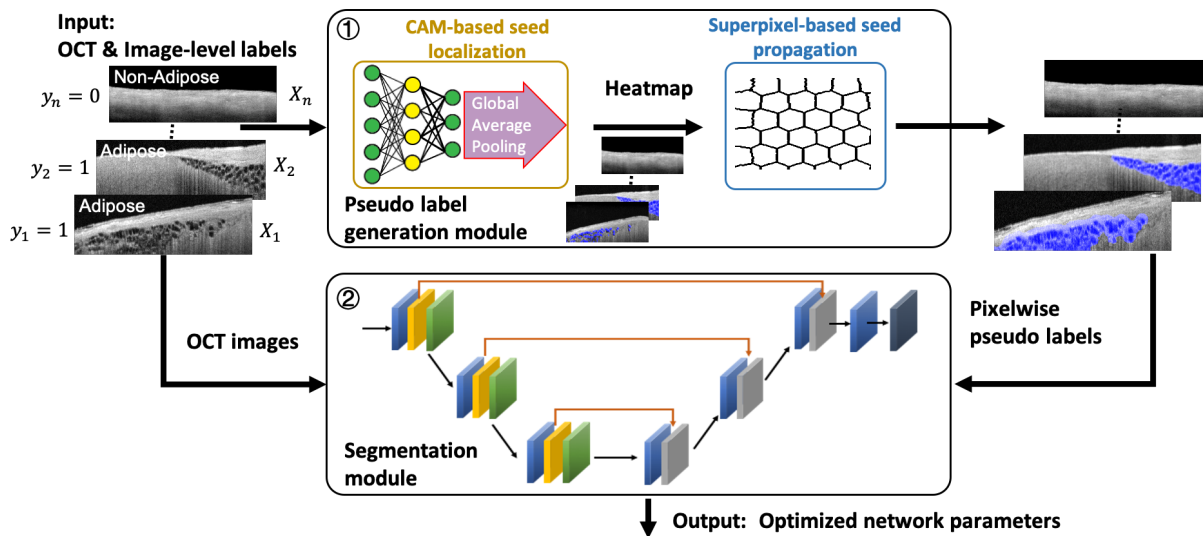
Fig. 2. Algorithm training flow of the proposed weakly supervised segmentation approach. The framework consists of two separate modules, namely pseudo label generation and segmentation network training. In pseudo label generation module, pixel-wise pseudo annotations are generated by the integration of CAM and superpixel methods. In segmentation network training module, a segmentation network is trained on the pseudo labels with a novel loss function.

In contrast to fully supervised methods, weakly supervised approaches use higher level labels including scribbles [26], [35], bounding boxes [28], [53], and image-level labels [52] to guide the pixel-level segmentation training process. [52] successfully segmented lesions by calculating the differences between the input abnormal images and normal-like retinal OCT images from a CycleGAN model. [56] employed a few shot learning technique for retinal disease classification and applied a GAN to enrich normal OCT images with OCT images of rare diseases. [27] proposed a Noise2Noise [33] based weakly supervised learning model for OCTA image reconstruction task.

The image-level label is the most convenient and easiest supervision among all weak annotations, as it does not require any detailed annotations in the input images. As a result, it cannot be directly used for segmentation guidance, due to the lack of location clues for the target tissues. By generating the initial location seeds, CAM [58] provides a practical solution to solve this issue, and thus has been widely adopted as the first step of weakly supervised learning frameworks. [10], [21], [57]. [57] developed an end-to-end approach named reliable region mining for weakly supervised semantic segmentation. Combined with CAM, their model applied additional conditional random field operation to get reliable object regions. In [10], a boundary exploration based segmentation approach was proposed to explore object boundaries in the segmentation training process. Researchers in [21] deployed an iterative learning framework to gradually expand the seeded regions.

III. PROBLEM ANALYSIS

Our study is conducted on a cardiac dataset that was acquired from 44 human hearts with a median age of 62 years. The dataset contains both healthy hearts, end-stage heart failure, atrial fibrillation, coronary heart disease, cardiomyopathy, and myocardial infarction. A detailed clinical characteristic is presented in Section V-A. These various disease conditions might alter the visual features of cardiac substrates, raising the following unique challenges on the algorithm design:

In-depth image and focal plane. OCT provides in-depth cross-sectional images in which the x-axis and y-axis are not interchangeable. The signal intensity decreases with increasing axial (depth) distance due to light attenuation as it travels in the tissue. In addition to the depth, the signal intensity is also affected by the system optics and configuration. The optical focus, corresponding to one depth (horizontal row of pixels), will have a higher intensity than regions deeper within the image. As a result, noise distribution within a 2D image is not uniformly distributed.

Various features and irregular shapes. As shown in Fig. 1, adipose regions present great variations among cardiac OCT images from human donors with cardiovascular disease due to heterogeneous heart remodeling. In Fig. 1 (A), intra-scan inconsistency can be clearly observed in the two sub-regions. Meanwhile, in comparison with Fig. 1 (B), the size of fat cells in Fig. 1 (A) is much smaller and the number of fat cells is larger, as indicated in the histology images. In addition, the distance to the endocardium tissue can also affect tissue appearance. Adipose tissues in Fig. 1 (C) are deeper in the myocardium and appear darker and blurrier than in Fig. 1 (B). Finally, the features and shapes can be further impacted by experimental conditions. In Fig. 1 (D), the OCT image was obtained from tissues submerged in phosphate buffered saline (PBS). In this sample, the adipose tissues have very low contrast with the surrounding normal tissues. Thus, algorithms relying on shape constraints or boundary exploration cannot be easily extended to this task.

Similar pattern among adipose tissue and noise. Image noise and artifacts are inevitable during the acquisition process. Features of adipose tissue are very similar to those of speckle noise and artifacts. To avoid over-smoothing the adipose regions, we do not deploy any image registration/denoising algorithm.

Data imbalance and limited training data. In human cardiac samples, the majority of regions are normal tissues, such as myocardium and endocardium, rather than targeted adipose tissues. In our dataset, only around 11% OCT images show visible clusters of adipose tissues. At the pixel level, pixels belonging to adipose tissues only account for 2.6% of the total pixels to label. Even for images that contain adipose tissues, the ratio of number of adipose-related pixels over the total number of pixels is very small. Hence, the samples that are informative for model training are very limited.

Mathematical modeling. Let $\mathcal{X} = \{X_i;\ i = 1, \ldots, N\}$ denote the set of images. For each image $X_i \in \mathcal{X}$ $(i = 1, \ldots, N)$, the image-level annotation $y_i \in \{0, 1\}$ indicates whether or not X_i contains the adipose tissues. The training data is denoted by $\mathcal{D}_{tr} = \{(X_i, y_i);\ i = 1, \ldots, N'\}$, which consists of N' images $X_1, \ldots, X_{N'}$ with corresponding image-level annotations $y_1, \ldots, y_{N'}$. Let $\mathcal{X}' = \{X_i;\ i = 1, \ldots, N'\}$. The goal of our approach is to build a segmentation network $f_{seg}(X; \theta)$ with network parameters θ, so that it can generate the pixel-wise segmentation masks $Pred$, which has binary label $Pred(\omega)$ at each pixel position ω in X.

IV. METHODOLOGY

In this paper, we propose a weakly supervised learning framework for cardiac tissue segmentation tasks. Fig. 2 shows the pipeline of our proposed framework. As shown, our training approach consists of two major stages: pseudo label generation and segmentation network training. The pseudo label generation module is formed by two components: we first apply the CAM approach to generate initial adipose seeds and then we use superpixel-base segmentation method to propagate the adipose seeds into pseudo pixel-wise labels. A detailed pseudo algorithm for the pseudo label generation module is listed in Algorithm 1. In the segmentation module, we introduce a novel loss function with a special focus on the adipose seed regions to increase the detection performance of our segmentation network. The proposed method addresses the issue of great variation among adipose tissues for two reasons. First, the proposed method relaxes a predicting precise irregular boundary issue to a rough localization of adipose seed issue using CAM. Without the need of prior knowledge, the CAM can be used to indicate the location of potential adipose tissues. Second, we determine the actual boundary of detected adipose using an unsupervised way, superpixel segmentation, thus bypassing the challenge of learning from large amounts of adipose regions with great boundary variation.

A. Pseudo Label Generation

1) CAM-based seed localization: The first stage of our model is to find reliable adipose seeds to indicate the location of adipose tissues. We follow the learning steps in [58] to get the initial CAM to indicate the location of adipose regions. To identify the extent of target tissue regions, we employ the global average pooling (GAP) layer to the last convolution layer of the classification network. The final prediction of the network is classified by a fully connected layer. After model

---

Algorithm 1: Algorithm Framework for Pseudo Label Generation Module

Input: Training dataset $\mathcal{X} = \{X_1, X_2, ..., X_n\}$ with image-level labels $\mathcal{Y} = \{y_1, y_2, ..., y_n\}$;
Output: Pixel-wises pseudo labels.
Procedure:
Step 1: Train localization network from $\mathcal{X}$ and $\mathcal{Y}$.
Step 2: Apply the CAM method to generate the initial tissue seed results $\mathcal{C} = \{c_1, c_2, ..., c_n\}$.
Step 3: Apply the boundary masking on $\mathcal{C}$ and get updated tissue seeds $\hat{\mathcal{C}}$.
Step 4: Apply superpixel-based propagation method on $\hat{\mathcal{C}}$ to generate the initial pseudo segmentation labels $\mathcal{S}$.
Step 5: Update $\mathcal{S}$ with the spatial regularisation strategy and get the final pseudo segmentation labels $\hat{\mathcal{S}}$.

---

training, the CAM results for the adipose tissues are obtained as follows:

$$M_{adipose}(\omega) = \sum_j \mu_j f_j(\omega)$$

where $f_j(\omega)$ is the activation of unit j in the last convolutional layer at spatial location ω and μ_j is the weight corresponding to the adipose tissue for unit j.

The class activation maps have strong responses on regions with artifacts and high-intensity noise. To increase the reliability of pseudo label generation, we apply a boundary masking algorithm on the class activation maps to filter out adipose seeds that are located in the background regions (false positive caused by noise) and regions close to the tissue-background boundary (false positive caused by artifacts). We adopt the cardiac layer segmentation algorithm from [16] and use the boundary of the top generated layer as the tissue-background interface. After getting the tissue surface, we remove adipose seeds above or close to the tissue-background boundary.

2) Superpixel-based seed propagation: Superpixels are generated as in [2] with an approximate number of superpixels as 2000. An entire superpixel is labeled as adipose tissue if one of its inner pixels is labeled as adipose tissue. Upon the generation of superpixels, the initial segmentation pseudo labels can be further improved to eliminate the following two misclassifications: 1) The adipose seeds may omit some adipose regions, and 2) the adipose seeds may incorrectly mark the normal regions as the adipose regions, due to the artifacts and intensity noise. To further remove the noisy annotations, we apply the Markov spatial regularization strategy [19] to add the ignored regions and remove the noisy adipose superpixels which only contain the normal tissues. Since the adipose cells are clustered in the cardiac tissue, the neighbors of an adipose region are more likely to belong to the adipose tissue clan, while small isolated adipose superpixels are more likely to be the normal regions corrupted by noise. Based on these criteria, we develop a simple yet effective spatial regularisation strategy: the label of a superpixel will be updated if most of its neighbors ($\geq 80\%$) belong to another class.

B. Segmentation Network Training

Adipose tissues are sparse compared with normal tissues, such as myocardium and endocardium. Without special consideration, the segmentation performance might be severely limited due to data imbalance. To overcome this challenge, we use an adipose seed loss, inspired by [31], to optimize our segmentation network. In comparison with the original form of seed loss in [31], our adipose seed loss function only focuses on the regions of target tissues while the background regions are omitted. The final outputs of our segmentation network are pixel-wise segmented adipose location results.

We first introduce the loss function. Let $p_k(\omega)$ denote the network output probability for class k at the pixel position $\omega \in \Omega$ with $\Omega \subset \mathbb{R}^2$ and $s_k(\omega)$ denote the pseudo-pixel-wise label where $k \in \{0, 1\}$. The cross-entropy loss (CEL) and adipose seed loss (ASL) is defined as follows:

$$CEL = -\frac{1}{|\Omega|} \sum_{\omega \in \Omega} \sum_{k=0}^{1} s_k(\omega) \log(p_k(\omega)) \tag{1}$$

$$ASL = -\frac{1}{|\Omega_1|} \sum_{\omega \in \Omega_1} s_1(\omega) \log(p_1(\omega)) \tag{2}$$

where $\Omega_1 = \{\omega \in \Omega : s_1(\omega) = 1\}$ is the set of spatial locations that are pseudo-labeled with class 1 (i.e., the adipose class). Compared with the CEL (Equation 1), the ASL (Equation 2) only focuses on the regions of adipose tissue, and thus, it helps to reduce the impact of the false negatives in the pseudo segmentation labels.

We also use Dice loss (DL) in our loss function to learn the context information. The DL is defined as:

$$DL = 1 - \frac{1}{2} \sum_{k=0}^{1} \frac{2 \sum_{\omega \in \Omega} (p_k(\omega) s_k(\omega))}{\sum_{\omega \in \Omega} (p_k(\omega))^2 + \sum_{\omega \in \Omega} (s_k(\omega))^2} \tag{3}$$

Finally, our segmentation network $f_{seg}(X; \theta)$ is jointly optimized by the combination of CEL, ASL, and DL:

$$Loss = \alpha_1 \cdot CEL + (1 - \alpha_1) \cdot ASL + \alpha_2 DL \tag{4}$$

where α_1 and α_2 are weight hyper-parameters. The sum of CEL and ASL is equivalent to a weighted loss. For better understanding, we use the form in Equation 4 to separate the weight of CEL and ASL.

V. EXPERIMENT EVALUATION

A. Dataset

We evaluate the performance of our proposed model on the human cardiac dataset previously used in [15]. It consists of *in-vitro* cohort of 385 images taken from 44 human atria and ventricles using the Thorlabs OCT system. The samples were acquired through a National Disease Research Interchange approved protocol from Columbia University. All specimens were de-identified and considered not human subjects research, according to the Columbia University Institutional Review Board under 45 CFR 46. Table I presents the clinical characteristic of the human donor hearts. A detailed feature analysis study among different group of people was presented in [39]. The source of the Thorlabs OCT system was centered at

TABLE I
CLINICAL CHARACTERISTICS OF HEART DONORS

Characteristic	Value
N	44
Demographic profile	
Age in years, median (average)	62 (62.2)
Female, n (%)	20 (45.5)
Medical history, n (%)	
Heart failure	10 (22.7)
Cardiomyopathy	8 (18.2)
Coronary artery disease	11 (25.0)
Myocardial infarction	10 (22.7)
Atrial fibrillation	3 (6.8)
Chronic obstructive pulmonary disease	16 (36.4)
Diabetes	17 (38.6)
Hypertension	27 (61.4)
Cause of death, n (%)	
Cardiac arrest	18 (40.9)
Cardiopulmonary arrest	2 (4.5)
Respiratory failure	5 (11.4)
Chronic obstructive pulmonary disease	1 (2.27)
Congestive heart failure	1 (2.27)
Others, cardiac related	11 (25.0)
Others, not cardiac related	6 (13.6)

1325 nm with a bandwidth of 150 nm. The axial and lateral resolutions were 6.5 and 15 μm in air, respectively. All datasets were acquired at 28 kHz. Each OCT image is of size 512 $\times$ 800 pixels with a field of view of 2.51 mm $\times$ 4 mm. We crop the images into small overlapping patches with size of 512 $\times$ 128 pixels. Three experts, blinded to the algorithm design, annotate the OCT images under the guidance from a pathologist. All images are carefully annotated at pixel level with visual cron-check on corresponding histology images. Our evaluation is conducted on a five-fold cross validation strategy with validation sets randomly divided over human subjects.

B. Implementation Details

Seed localization network. To avoid overfitting, we only train a localization network with three hidden layers for adipose tissue seed generation. We use the ReLU function as the activation function. The number of channels in each hidden layer is 32, 32, and 64. We applied the GAP layer to the last convolutional layer to learn the cluster pattern of adipose tissues. The output of the GAP layer is classified with a fully connected layer. The digital resolution of output CAM drops down to $\frac{1}{4}$ of original digital resolution after passing through the localization network. The networks are optimized on cross entropy loss via Adam optimizer [30] with random Glorot uniform initialization [17]. The batch size used for training is 16. Over the cross validation sets, all networks converged within 300 epochs with a learning rate of $1e^{-3}$. After normalizing, the background threshold in CAM generation is 0.15.

Segmentation network. We employ the classic medical segmentation network UNet [49] as the baseline of our learning framework. In the loss function (Equation 4), the hyper-parameter α_2 is 0.5 while $\alpha_1 \in [0, 1]$ is determined by the proportion of adipose tissues in the training set: $\alpha_1 = 1$ if the ratio of adipose images is less than 5%. Otherwise, $\alpha_1 = 0.5$. All segmentation networks were randomly initialized
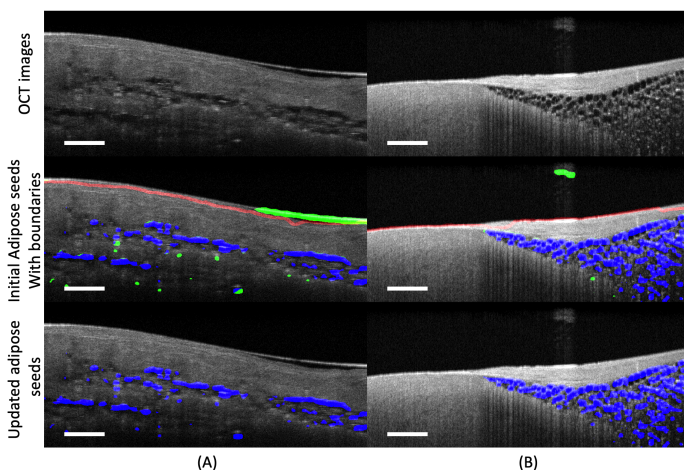
Fig. 3.   Comparison of tissue seeds before and after the boundary masking algorithm. Red: the detected tissue-background boundaries; blue: accurately annotated adipose seeds; green: false positives. As shown, the boundary masking algorithm can effectively remove the false positive adipose seeds caused by the artifacts and noise. Benefiting from it, the adipose seeds are more precise to be propagated for segmentation guidance. Scale bar: $500\ \mu m$.

Fig. 5.   Comparison of pseudo labels with and without the spatial regularisation strategy. Blue: accurately annotated adipose pixels; red: false negatives. The spatial regularisation strategy helps to correct the mis-labeled pseudo labels by using the context information from nearby regions. After applying it, the false negatives have been significantly reduced. Scale bar: $500\ \mu m$.
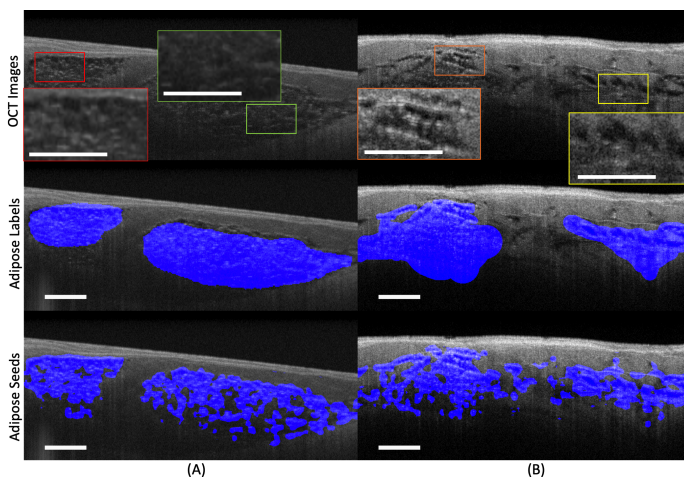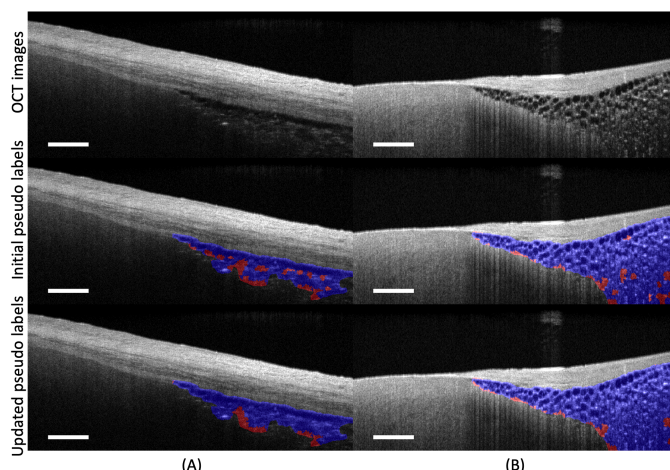

Fig. 4.   The tissue seed results in samples with different adipose features. Sample (A) is obtained from the left atrium. The highlighted regions in sample (A) are two regions with different focal statuses. The red box is within the focal plane while the green box is out of focus. Sample (B) is obtained from the right ventricle. The highlighted regions in sample (B) are with different tissue structures. The orange box is with multiple small fat cells while the yellow box is with a few big fat cells. Scale bar: $500\ \mu m$.

	Before	After
Accuracy	80.79 ± 1.15	83.91 ± 2.45
Precision	56.43 ± 11.95	75.90 ± 8.18

classification network training, generation of pseudo labels, and segmentation network training, is less than 3 hours.

C. Evaluation Metrics

In our experiment, we use accuracy and precision to evaluate the overall accuracy and the detection performance of adipose seed results. For pseudo label generation and segmentation evaluation, we use true positive rate (detection rate), false positive rate, and Dice coefficient (F1 score) to evaluate the tissue segmentation performance.

D. Evaluation of Pseudo Label Generation

Adipose tissue seed localization. The binary accuracy for our proposed localization network achieves very stable performance on all validation sets. Figure 3 presents two representative adipose seed results generated from our localization network. In Fig. 3, the detected tissue-background boundary is delineated in red, and the accurately located adipose seeds are marked in blue, with misclassified adipose seeds marked in green. As seen, the boundary masking algorithm can effectively remove the misclassified edges and background noise from the original adipose seed results. Meanwhile, these results also demonstrate the capability of recognizing various types of adipose features using our proposed network. In the lower-left region of Fig. 3 (A) and lower region of Fig. 3 (B), the adipose region is out-of-focus, and the wall of adipose cells is not distinctive. In the majority of parts of Fig. 3 (B), the adipose regions are in focus and appear in the honeycomb

and converged within 200 epochs with a learning rate of 0.1^{-3}, batch size 32, and weight decay 10^{-4}. Our study is based on the Keras [11] and TensorFlow [1] deep learning frameworks. All experiments are conducted on a computer with the following features: an Intel core i9-9900K (16M Cache, up to 5.00 GHz) CPU and a RTX 2080 Ti GPU. The number of training images in a cross-validation set is around 300 images, with a size of 512×800 pixels. In the preprocessing stage, the training images were cropped into around 1800 overlapping patches (512×128 pixels) to enrich the training size. The total execution time for a single trial, which includes the following steps: preprocessing,

TABLE III
EVALUATION METRICS (%) ON TISSUE PSEUDO LABELS BEFORE AND AFTER THE SPATIAL REGULARISATION ON ADIPOSE SAMPLES.

Method	True Positive Rate	False Positive Rate	Dice Coefficient
Superpixel	71.17 ± 6.36	10.03 ± 2.87	67.09 ± 2.96
Superpixel + Spatial regularisation	71.77 ± 6.72	8.33 ± 2.90	69.70 ± 3.34

TABLE IV
EVALUATION METRICS (%) ON PSEUDO LABELS GENERATED FROM DIFFERENT MODELS ON ADIPOSE SAMPLES.

Model	Precision	True Positive Rate (%)	False Positive Rate (%)	Dice Coefficient (%)
Deeper Model	0.93 ± 0.08	64.13 ± 10.76	5.38 ± 0.97	68.29 ± 7.18
Wider Model	0.92 ± 0.07	63.78 ± 3.98	$\mathbf{5.31 \pm 1.50}$	68.87 ± 1.59
Proposed	$\mathbf{0.93 \pm 0.06}$	$\mathbf{71.77 \pm 6.72}$	8.33 ± 2.90	$\mathbf{69.70 \pm 3.34}$

TABLE V
EVALUATION METRICS (%) OF DIFFERENT MODELS ON WHOLE DATASET. EXCEPT FULLY SUPERVISED METHOD, THE BEST AND SECOND BEST PERFORMANCE ARE MARKED IN RED AND BLUE CORRESPONDINGLY.

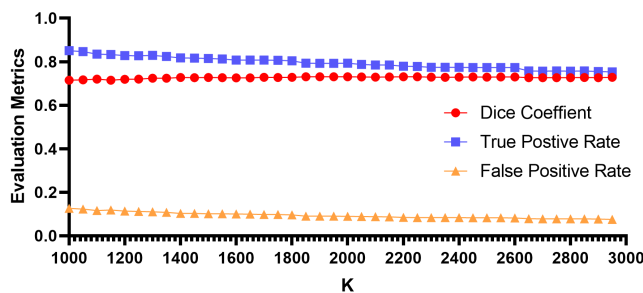|	Method	True Positive Rate	False Positive Rate	Dice Coefficient
Fully supervised baseline	U-Net	$\mathbf{82.32 \pm 5.69}$	$\mathbf{0.99 \pm 0.90}$	$\mathbf{81.57 \pm 5.71}$
Weakly supervised baseline	RRM	76.00 ± 16.36	5.44 ± 2.70	40.39 ± 9.01
	CycleGAN	55.19 ± 17.46	7.23 ± 6.44	64.14 ± 17.21
Proposed framework	U-Net	85.52 ± 5.44	1.35 ± 0.74	72.11 ± 10.12
	FCN	87.77 ± 1.01	1.49 ± 0.85	72.14 ± 8.94
	DeepLab	80.22 ± 4.18	1.45 ± 0.84	67.83 ± 11.25
Ablation study	w/o Boundary masking	72.59 ± 8.46	1.37 ± 1.09	70.91 ± 16.95
	w/o Spatial regularisation	77.84 ± 5.67	1.21 ± 1.08	70.77 ± 12.54
	Adipose seed loss + Dice loss	87.62 ± 8.79	2.24 ± 0.72	64.15 ± 9.82
	CE loss + Dice loss	74.04 ± 15.25	0.77 ± 0.30	74.99 ± 14.72

Sensitivity Analysis on K Values

Fig. 6. Impact of the number of superpixels on pseudo label quality. The quality of pseudo labels only has slight variance on the evaluation metrics (Dice coefficient, true positive rate, and false positive rate) among different K values, which indicates the robustness of our model on pseudo label generation.

structures. These adipose regions are correctly identified using our proposed approach. In Table II, we report the accuracy and precision of adipose seeds before and after boundary masking. As shown, both accuracy and precision are improved after applying the boundary masking algorithm. In particular, the precision of adipose seeds has been significantly increased by approximately 20%, which further demonstrates the effectiveness of boundary masking on accuracy improvement.

In Fig. 4, we show two representative samples that were obtained from different chambers. In each sample, the adipose tissues are located in multiple regions with different appearances. As shown, our model can successfully localize the adipose tissue in a single sample with different adipose appearances. These results indicate that our model is able to

solve the various feature challenges described in Section III.

Pseudo label generation. Figure 5 shows two superpixel segmentation results with accurately segmented pixels marked in blue and false negatives marked in red. In Table III, we provide a quantitative evaluation of our spatial regularisation strategy on samples with adipose tissues. There is an increase of 2% in Dice coefficient after applying it.

Localization network architecture. To avoid overfitting, we apply a small localization network with 32, 32, and 64 channels on CAM generation. To learn the impact of network architecture on pseudo label generation, we conduct sensitivity analysis on models with more parameters: a deeper network with 32, 32, 32, and 64 channels and a wider network with 64, 64, and 64 channels in corresponding convolutional layers. Table IV presents the evaluation metrics on pseudo labels generated from different networks over the adipose samples. The quality of pseudo labels from deeper and wider networks is slightly worse than our proposed network. This discrepancy is probably caused by the noise and artifacts in the OCT images, as these large models might overfit to some non-robust features [24].

Sensitivity analysis on the number of superpixels. The number of superpixels K is the most important hyperparameter in the pseudo label generation module, with larger K values leading to smaller sizes of superpixels. To learn the impact of K on pseudo label quality, we report the evaluation metrics of pseudo labels with different K values in Fig. 6. Along with increased K values, the true positive rate slightly decreases while the false positive rate is improved. Thus, we empirically set $K = 2000$ to balance the true positive rate and false positive rate in our study.
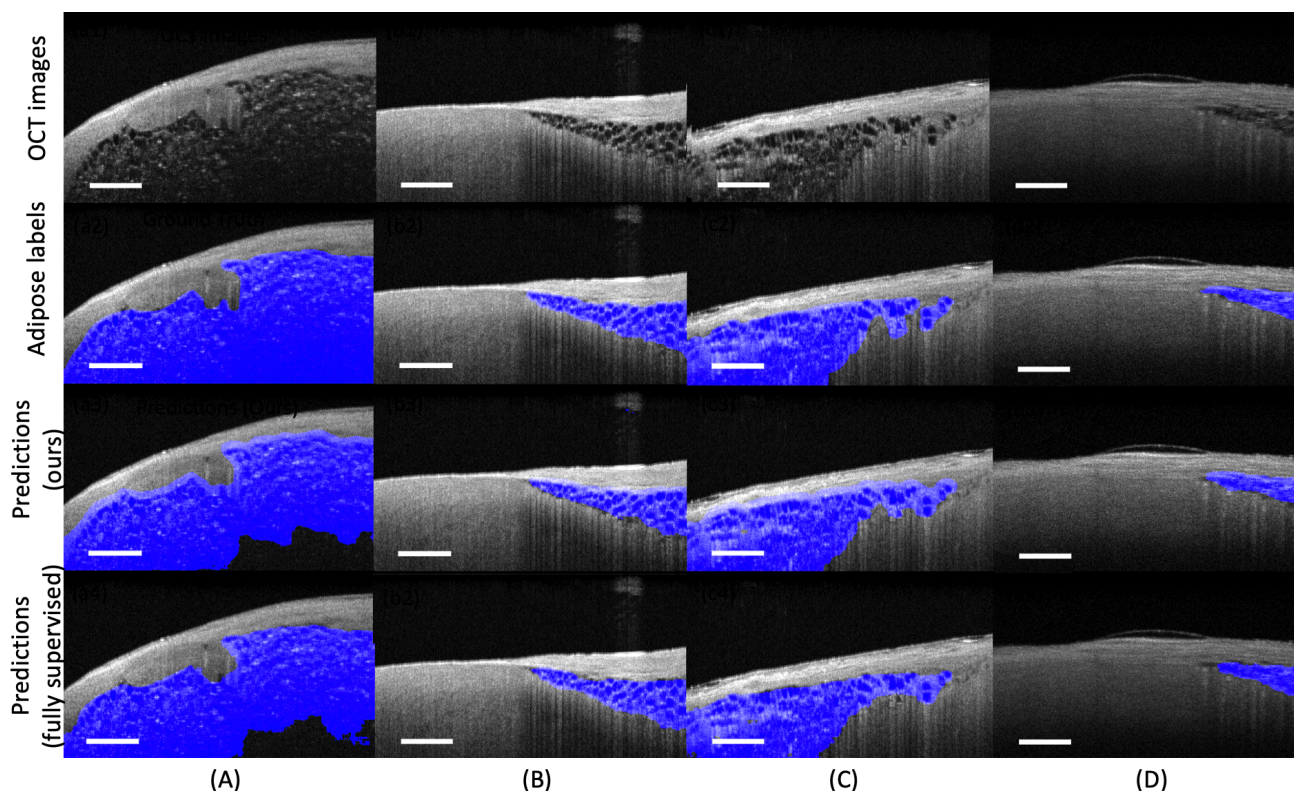
Fig. 7. Representative segmentation results from human atrium and ventricle samples. Our proposed approach accurately identifies the adipose tissues located at different regions with various sizes and shapes. All prediction results are highly consistent with the ground truth labels. Scale bar: **500 μm**

E. Evaluation of segmentation performance

1) Cross Validation Experiments: We use fully supervised models trained from pixel-wise accurate segmentation masks as the baseline for comparison. Table V summarizes the averaged results and the standard deviation of our proposed weakly supervised approach and fully supervised baselines. To fully evaluate our model, we report the results calculated from all samples.

Weakly supervised learning vs fully supervised learning. Our weakly supervised model, trained from image-level labels, achieves comparable quality results to the fully supervised model that trained on pixel-wise labels. In addition, our dataset was acquired within a time frame that spanned over five years. During this time frame, imaging setup, such as sample freshness, imaging condition, and tissue preparation, varied among experiments. Thus, our results also demonstrate the generalization ability of our model against imaging condition variance, showing its strong potential on real-world clinical applications.

Comparison with different segmentation models. In addition to the classic UNet model, we also evaluate the performance of our approach with other state-of-the-art segmentation models including DeepLab V3 [9] and FCN [37]. The evaluation metrics of these models are shown in Table V. As shown, all models achieve comparable results (true positive rate>0.80, false positive rate<1.5 over all samples), which indicates the genericity of our framework on weakly supervised tissue identification.

Comparison with existing weakly supervised learning

framework. We compared our method with two existing weakly supervised learning frameworks: reliable region mining (RRM) [57] and CycleGAN [52]. As shown in Table V, our methods generally produce a higher true positive rate and a lower false positive rate than RRM and cycleGAN, independent of segmentation models (i.e., U-Net, FCN, and DeepLab). RRM is based on the concept of conditional random field (CRF). CRF was originally designed for natural images where the boundary of objects was with high contrast, and pixel-wised affinity could be calculated accordingly. However, the blurry boundary of adipose regions within OCT images can add additional challenges in tissue segmentation and thus lead to lower performance of RRM. In contrast, the combination of superpixel design and Markov regularization can contribute to addressing the unique challenges that adipose and speckle noise happen to be similar.

Our approach also achieves higher performance than the CycleGAN-based method [52], which relies heavily on spatial constraints. In our problem setting, the adipose tissue could appear in any region under the endocardium. The morphological change is significant compared to those within retinal images [52]. Furthermore, their model detected the bottom layer (RPE) of the retina to avoid the interruption of background noise. However, in the cardiac OCT dataset, there does exist a distinctive bottom layer to avoid background noise.

Although the adipose tissues and the noise are similar on pixel-wise, they are different on cluster-wise. The walls of adipose cells can be differentiated from the background when the adipose regions are close to the endocardial surface (first
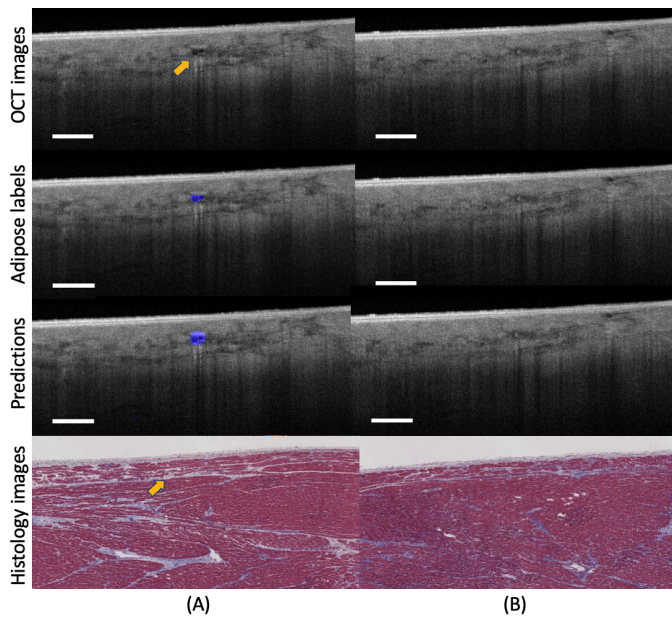
Fig. 8. The prediction results of images obtained from nearby regions. Our approach successfully pinpoints the adipose tissues from other tissue types, showing its strong identification ability on adipose tissues. Scale bar: 500 μm.

layer). As the location of adipose regions appears deeper within the tissue, within the OCT image, it is farther from the focus and thus is blurred and has a lower contrast with the background. The superpixel, as a basic unit, carries more regional information than a single pixel, thus showing promise to distinguish the adipose and noise. In addition, the use of Markov spatial regularization could also help to filter isolated noise by removing noise if the majorities are non-adipose regions and expanding regions if the majorities are adipose regions.

Ablation study. Performance drops are observed in Table V when the boundary masking algorithm and spatial regularisation strategy are removed. In particular, after removing the boundary masking algorithm, the true positive rate is severely decreased along with an increased standard deviation. This result indicates the necessity of using a boundary masking algorithm to improve adipose seed quality at the early stage of pseudo-label generation. In comparison with the false positive rate, the true positive rate has notable changes after applying the spatial regularisation, which shows its effectiveness in false negative correction.

We further conduct experiments to assess the influence of different loss functions in our proposed model. As shown in the last four rows of Table V, our model which is optimized with adipose seed loss and Dice loss achieves the best true positive rate, while the version without adipose seed loss (CE loss and Dice loss), achieves the best false positive rate in ablation study. The use of seed loss can notably increase the model detection performance but meanwhile, hinder the false positive rate. In contrast, the cross entropy loss is more efficient in controlling the false alarms. These results show that the use of seed loss can efficiently reduce the impact of false negatives in the pseudo labels. This detection rate and

false alarm trade-off can be balanced through adjusting the weights of seed loss and cross entropy loss.

2) Representative Segmentation Results: In this section, we present the visual output of our proposed weakly supervised model in overall performance, small adipose tissue region detection, and 3D segmentation.

Overall performance. Figure 7 shows the predicted tissue maps on four human cardiac samples. In Fig. 7 (A) and (B), our model accurately localizes the adipose tissue regions in arbitrary shapes. Meanwhile, in Fig. 7 (A), we can also observe the over-segmented regions (regions at the left corner) in the ground truth figure. Human annotators tend to over-segment the regions below the penetration depth, while for the network, it may identify these regions as non-adipose tissues because of the low signal-to-noise ratio. This over-segmentation tendency can lead to decreased values in evaluation metrics. In Fig. 7 (C), our model successfully identifies the adipose tissues in the multiple regions with different penetration depths. In Fig. 7 (D), we show a human atrium sample that is slightly off the focus. Similar to previous results, our models still accurately differentiate the adipose tissues from other tissues, showing their robustness over different image qualities. In all cases, the predicted results are highly consistent with the ground truth labels. These results demonstrate the learning ability of our model via image-level labels, showing its effectiveness in clinical tissue identification.

Identifying small adipose tissue regions. Figure 8 presents two images obtained from nearby regions within the same human heart. As shown, Fig. 8 (A) and (B) are very similar and they all contain large regions of fibrosis tissue. However, in Fig. 8 (A), there is a small cluster of adipose tissue surrounded by the fibrosis tissues, while in Fig. 8 (B), there is no adipose tissue. As shown, this is a very challenging segmentation task due to the size of adipose tissue and the blurry boundary between different tissue types. Our model accurately delineates the adipose tissue regions in Fig. 8 (A) and it does not put any false alarm in Fig. 8 (B). In both cases, our model successfully distinguishes adipose tissues from other cardiac substrates. These results further demonstrate the strong learning ability of our model, as it can learn the most discriminative features via image-level labels, rather than simply memorizing the training samples.

Visualization of 3D segmentation. Figure 9 shows a typical result of 3D visualization of adipose tissue segmentation. We sequentially apply the trained network to segment consecutive Bscans and align segmented Bscans in 3D space. As shown, the segmented boundaries accurately delineate the morphological changes in the adipose tissues. Even though our model is trained on a small quantity of training data with image-level labels, it still successfully segments adipose regions of various sizes. These results indicate that our model has great potential to be applied to assess adipose tissue regions in catheter-based ablation operations.

VI. DISCUSSION

In this study, we propose a weakly supervised learning framework for cardiac adipose tissue segmentation on OCT
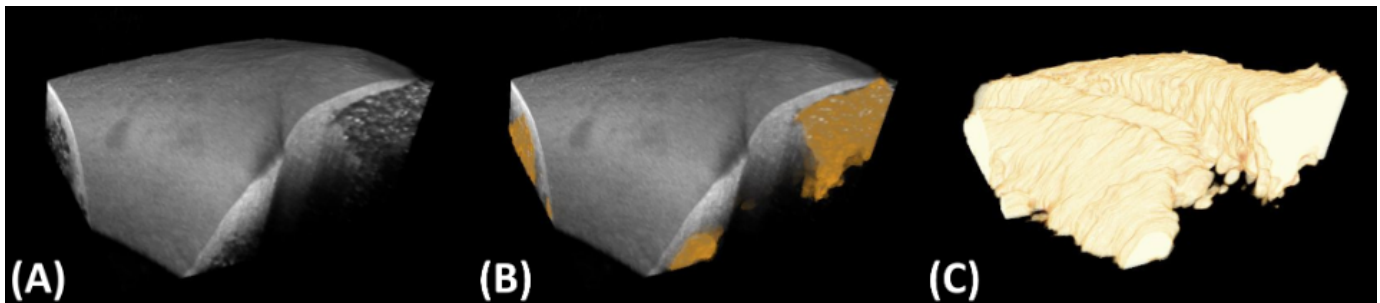
Fig. 9. 3D visualization of adipose tissue segmentation. (A): the original OCT volume; (B): the original volume overlaid with segmented adipose regions; (C): the segmented adipose regions from proposed approach. The segmented boundaries accurately delineate the morphological changes in adipose shape.

images. Our approach contains two powerful modules: the pseudo label generation and the segmentation network training. In the pseudo label generation module, we use the superpixel-based propagation algorithm to address the sparse location seed challenge raised in the CAM results. Benefiting from our boundary masking algorithm and spatial regularisation strategy, the quality of the pseudo labels has been significantly improved for the training guidance. In the segmentation network training module, we introduce a novel loss function to increase adipose tissue detection performance. By evaluating on a human cardiac dataset with cross-validation strategy, our model achieves comparable results with the fully supervised baseline, showing its effectiveness on tissue characterization.

Our study can provide valuable detail on adipose distribution within myocardium regions, holding promise to improve cardiac therapies including ablation procedure and endomyocardial biopsy. Previous research has shown that the accumulation of adipose tissue in ventricular myocardium is associated with severe cardiac arrhythmias such as left ventricular tachycardia and arrhythmogenic right ventricular cardiomyopathy. By pointing out the location of abnormalities, our study could directly improve the radiofrequency ablation treatment, reducing the risk of treatment recurrence and complications. In addition, the knowledge of adipose distribution could also provide referral suggestions on ablation parameters for lesion formation, as the biophysical properties are various among different tissue types. Moreover, our approach also holds promise to be extended for weakly supervised segmentation of adipose tissue in other applications, such as adipose detection in breast cancer classification.

Previous research has shown that the accumulation of adipose tissue in the ventricular myocardium is a substrate for cardiac arrhythmias, which is associated with left ventricular tachycardia and arrhythmogenic right ventricular cardiomyopathy. Additionally, about 75% to 80% of cases of sudden cardiac death are caused by ventricular arrhythmias, and over 80,000 people are detected with supraventricular tachycardia annually [29], [45]. However, there are no existing methods for real-time image guidance of ablation procedures, to the best of our knowledge. Current guidance of ablation procedures is through the measurement of electrograms. Low voltage measurements can be due to increased amounts of collagen (scar or fibrosis) or adipose. Knowledge of the substrate will provide the necessary feedback to determine how to interpret functional measurements. Future OCT-enabled assessment of the presence of adipose (true positive rate) will provide a valuable addition to an electrophysiologist's toolbox for treating cardiac arrhythmias.

Similar to the experiments in RRM, we notice that other weakly supervised methods that were originally developed in computer vision might not be easily applied to our unique OCT task. For instance, SEC [31] is a CRF-based method. Note that the CRF-based method is originally designed for natural images where the boundary of objects is high contrast and continuous. However, some adipose regions in OCT images do not satisfy this condition, especially when the adipose cells are in out-of-focus regions. Likewise, the pairwise affinity-based IRN [3] approach is also hard to handle such subtle feature variation challenges in OCT adipose images. In addition, the IRN framework relies upon multiple instances of CAM, where each object is considered as a single subject for instance segmentation. Such design cannot be easily integrated with adipose tissue which shows a sparse distribution within the myocardium (shown in Fig. 3), as this sparsity could lead to over hundreds of instances in a single OCT image. Moreover, erasing-based methods [50] could not be applied to our problem setting as normal tissue can be likely removed with fat cells due to the blurry boundary of the adipose region.

One limitation of this study is that the results are evaluated on a benchtop OCT system. To move towards the aid of ablation procedures, a catheter-based OCT system is needed, as it can help to optimize the treatment strategy by providing real-time cardiac substrate information. In the future, we will extend our current work into catheter-based *in-vivo* OCT images. Such extension will require further investigation of challenges such as image quality degeneration and motion disturbance. Compared with benchtop OCT images, catheter-based OCT images are with lower image quality, suffering from lower contrast and motion effects. Without special consideration, the decreased image contrast may hinder the performance of the model. Motion disturbance caused by breath and heartbeat is another important factor that could lead to performance degradation. These disturbances could be partially corrected by applying low-pass filters. In the future, we will also extend our proposed weakly supervised framework to other OCT segmentation tasks, such as breast images and retinal images.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3263838

HUANG *et al.*: CARDIAC ADIPOSE TISSUE SEGMENTATION VIA IMAGE-LEVEL ANNOTATIONS

11

Moreover, the segmentation performance is limited by the generated pseudo labels. In the future, we will explore the direction of semi-supervised learning if a separate set of pixel-wise labels is available for training.

VII. CONCLUSION

In this paper, we propose the first weakly supervised learning framework for adipose tissue segmentation in human cardiac OCT images. We design a novel CAM-superpixel segmentation approach that converts the sparse CAM results into pseudo-pixel-wise labels for training. In addition, we also present and analyze the necessity and effectiveness of proposed steps and loss functions. Experimental results on the human cardiac dataset demonstrate that our model achieves comparable performance with models trained under full masks, showing the learning capability of our proposed model on image-level labels.

In the future, we will extend our work in the following aspects. First, we will evaluate the performance of our current work into catheter-based *in-vivo* OCT images or synthetic *in-vivo* OCT images generated by generative models and image kernels. Second, we will validate the effectiveness of our model on other OCT segmentation tasks to open up opportunities for broader applications.

REFERENCES

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

[2] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.

[3] Jiwoon Ahn, Sunghyun Cho, and Suha Kwak. Weakly supervised learning of instance segmentation with inter-pixel relations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2209–2218, 2019.

[4] Minhaj Alam, David Le, Taeyoon Son, Jennifer I Lim, and Xincheng Yao. AV-Net: deep learning for fully automated artery-vein classification in optical coherence tomography angiography. *Biomedical Optics Express*, 11(9):5249–5257, 2020.

[5] Christina M Ambrosi, Vadim V Fedorov, Igor R Efimov, Richard B Schuessler, and Andrew M Rollins. Quantification of fiber orientation in the canine atrial pacemaker complex using optical coherence tomography. *Journal of Biomedical Optics*, 17(7):071309, 2012.

[6] Christina M Ambrosi, Nader Moazami, Andrew M Rollins, and Igor R Efimov. Virtual histology of the human heart using optical coherence tomography. *Journal of Biomedical Optics*, 14(5):054002, 2009.

[7] Justus MB Anumonwo and Todd Herron. Fatty infiltration of the myocardium and arrhythmogenesis: potential cellular and molecular mechanisms. *Frontiers in Physiology*, 9:2, 2018.

[8] Maria Bonou, Sophie Mavrogeni, Chris J Kapelios, George Markousis-Mavrogenis, Constantina Aggeli, Evangelos Cholongitas, Athanase D Protogerou, and John Barbetseas. Cardiac adiposity and arrhythmias: the role of imaging. *Diagnostics*, 11(2):362, 2021.

[9] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[10] Liyi Chen, Weiwei Wu, Chenchen Fu, Xiao Han, and Yuntao Zhang. Weakly supervised semantic segmentation with boundary exploration. In *European Conference on Computer Vision*, pages 347–362, 2020.

[11] François Chollet et al. *Keras*. GitHub: https://github.com/fchollet/keras, 2015.

[12] Mohammed El Mahdiui, Judit Simon, Jeff M Smit, Jurrien H Kuneman, Alexander R van Rosendael, Ewout W Steyerberg, Rob J van der Geest, Lili Szársaz, Szilvia Herczeg, Nándor Szegedi, et al. Posterior left atrial adipose tissue attenuation assessed by computed tomography and recurrence of atrial fibrillation after catheter ablation. *Circulation: Arrhythmia and Electrophysiology*, 14(4):e009135, 2021.

[13] Leyuan Fang, David Cunefare, Chong Wang, Robyn H Guymer, Shutao Li, and Sina Farsiu. Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative amd patients using deep learning and graph search. *Biomedical Optics Express*, 8(5):2732–2744, 2017.

[14] Christine P Fleming, Hui Wang, Kara J Quan, and Andrew M Rollins. Real-time monitoring of cardiac radio-frequency ablation lesion formation using an optical coherence tomography forward-imaging catheter. *Journal of Biomedical Optics*, 15(3):030516, 2010.

[15] Yu Gan, Theresa H Lye, Charles C Marboe, and Christine P Hendon. Characterization of the human myocardium by optical coherence tomography. *Journal of Biophotonics*, 12(12):e201900094, 2019.

[16] Yu Gan, David Tsay, Syed B Amir, Charles C Marboe, and Christine P Hendon. Automated classification of optical coherence tomography images of human atrial tissue. *Journal of Biomedical Optics*, 21(10):101407, 2016.

[17] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[18] Craig J Goergen, Harsha Radhakrishnan, Sava Sakadžić, Emiri T Mandeville, Eng H Lo, David E Sosnovik, and Vivek J Srinivasan. Optical coherence tractography using intrinsic contrast. *Optics Letters*, 37(18):3882–3884, 2012.

[19] Menglin Guo, Mei Zhao, Allen MY Cheong, Houjiao Dai, Andrew KC Lam, and Yongjin Zhou. Automatic quantification of superficial foveal avascular zone in optical coherence tomography angiography implemented with deep learning. *Visual Computing for Industry, Biomedicine, and Art*, 2(1):1–9, 2019.

[20] Meghna Gupta, Andrew M Rollins, Joseph A Izatt, and Igor R Efimov. Imaging of the atrioventricular node using optical coherence tomography. *Journal of Cardiovascular Electrophysiology*, 13(1):95–95, 2002.

[21] Zilong Huang, Xinggang Wang, Jiasi Wang, Wenyu Liu, and Jingdong Wang. Weakly-supervised semantic segmentation network with deep seeded region growing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7014–7023, 2018.

[22] Ziyi Huang, Yu Gan, Theresa Lye, Darnel Theagene, Spandana Chintapalli, Simeran Virdi, Andrew Laine, Elsa Angelini, and Christine P Hendon. Segmentation and uncertainty measures of cardiac substrates within optical coherence tomography images via convolutional neural networks. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1–4. IEEE, 2020.

[23] Ziyi Huang, Yu Gan, Theresa Lye, Haofeng Zhang, Andrew Laine, Elsa D Angelini, and Christine Hendon. Heterogeneity measurement of cardiac tissues leveraging uncertainty information from image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 782–791. Springer, 2020.

[24] Andrew Ilyas, Shibani Santurkar, Dimitris Tsipras, Logan Engstrom, Brandon Tran, and Aleksander Madry. Adversarial examples are not bugs, they are features. *Advances in Neural Information Processing Systems*, 32, 2019.

[25] Mohammad Shafkat Islam, Jui-Kai Wang, Samuel S Johnson, Matthew J Thurtell, Randy H Kardon, and Mona K Garvin. A deep-learning approach for automated OCT en-face retinal vessel segmentation in cases of optic disc swelling using multiple en-face images as input. *Translational Vision Science and Technology*, 9(2):17–17, 2020.

[26] Zhanghexuan Ji, Yan Shen, Chunwei Ma, and Mingchen Gao. Scribble-based hierarchical weakly supervised learning for brain tumor segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 175–183. Springer, 2019.

[27] Zhe Jiang, Zhiyu Huang, Bin Qiu, Xiangxi Meng, Yunfei You, Xi Liu, Mufeng Geng, Gangjun Liu, Chuanqing Zhou, Kun Yang, et al. Weakly supervised deep learning-based optical coherence tomography angiography. *IEEE Transactions on Medical Imaging*, 40(2):688–698, 2020.

[28] Hoel Kervadec, Jose Dolz, Shanshan Wang, Eric Granger, and Ismail Ben Ayed. Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision. In *Medical Imaging with Deep Learning*, pages 365–381. PMLR, 2020.

[29] Shaan Khurshid, Seung Hoan Choi, Lu-Chen Weng, Elizabeth Y Wang, Ludovic Trinquart, Emelia J Benjamin, Patrick T Ellinor, and Steven A Lubitz. Frequency of cardiac rhythm abnormalities in a half million

adults. *Circulation: Arrhythmia and Electrophysiology*, 11(7):e006273, 2018.

[30] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

[31] Alexander Kolesnikov and Christoph H Lampert. Seed, expand and constrain: Three principles for weakly-supervised image segmentation. In *European Conference on Computer Vision*, pages 695–711. Springer, 2016.

[32] Kyungmoo Lee, Alexis K Warren, Michael D Abràmoff, Andreas Wahle, S Scott Whitmore, Ian C Han, John H Fingert, Todd E Scheetz, Robert F Mullins, Milan Sonka, et al. Automated segmentation of choroidal layers from 3-dimensional macular optical coherence tomography scans. *Journal of Neuroscience Methods*, 360:109267, 2021.

[33] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, Timo Aila, et al. Noise2noise: Learning image restoration without clean data. In *International Conference on Machine Learning*. PMLR, 2018.

[34] Chao Li, Haibo Jia, Jinwei Tian, Chong He, Fang Lu, Kaiwen Li, Yubin Gong, Sining Hu, Bo Yu, and Zhao Wang. Comprehensive assessment of coronary calcification in intravascular OCT using a spatial-temporal encoder-decoder network. *IEEE Transactions on Medical Imaging*, 41(4):857–868, 2022.

[35] Xiaoming Liu, Quan Yuan, Yaozong Gao, Kelei He, Shuo Wang, Xiao Tang, Jinshan Tang, and Dinggang Shen. Weakly supervised segmentation of COVID19 infection with scribble annotation on CT images. *Pattern Recognition*, 122:108341, 2022.

[36] Julian Lo, Morgan Heisler, Vinicius Vanzan, Sonja Karst, Ivana Zadro Matovinović, Sven Lončarić, Eduardo V Navajas, Mirza Faisal Beg, and Marinko V Šarunić. Microvasculature segmentation and intercapillary area quantification of the deep vascular complex using transfer learning. *Translational Vision Science & Technology*, 9(2):38–38, 2020.

[37] 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.

[38] Theresa H Lye, Charles C Marboe, and Christine P Hendon. Imaging of subendocardial adipose tissue and fiber orientation distributions in the human left atrium using optical coherence tomography. *Journal of Cardiovascular Electrophysiology*, 30(12):2950–2959, 2019.

[39] Theresa H Lye, Darnel Theagene, Charles C Marboe, and Christine P Hendon. Quantifying differences in cardiac tissue structure related to patient characteristics using an optical coherence tomography cardiac tissue atlas. In *Diagnostic and Therapeutic Applications of Light in Cardiology 2020*, volume 11215, page 1121508. SPIE, 2020.

[40] Andrew Mamalis, Derek Ho, and Jared Jagdeo. Optical coherence tomography imaging of normal, chronologically aged, photoaged and photodamaged skin: a systematic review. *Dermatologic Surgery: Official Publication for American Society for Dermatologic Surgery [et al.]*, 41(9):993, 2015.

[41] Saleha Masood, Ruogu Fang, Ping Li, Huating Li, Bin Sheng, Akash Mathavan, Xiangning Wang, Po Yang, Qiang Wu, Jing Qin, et al. Automatic choroid layer segmentation from optical coherence tomography images using deep learning. *Scientific Reports*, 9(1):1–18, 2019.

[42] William Meiniel, Yu Gan, Christine P Hendon, Jean-Christophe Olivo-Marin, Andrew Laine, and Elsa D Angelini. Sparsity-based simplification of spectral-domain optical coherence tomography images of cardiac samples. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 373–376. IEEE, 2016.

[43] Reza Mirshahi, Pasha Anvari, Hamid Riazi-Esfahani, Mahsa Sardarinia, Masood Naseripour, and Khalil Ghasemi Falavarjani. Foveal avascular zone segmentation in optical coherence tomography angiography images using a deep learning approach. *Scientific reports*, 11(1):1–8, 2021.

[44] Arata Miyazawa, Young-Joo Hong, Shuichi Makita, Deepa Kasaragod, and Yoshiaki Yasuno. Generation and optimization of superpixels as image processing kernels for jones matrix optical coherence tomography. *Biomedical Optics Express*, 8(10):4396–4418, 2017.

[45] Leonardo A Orejarena, Humberto Vidaillet, Frank DeStefano, David L Nordstrom, Robert A Vierkant, Peter N Smith, and John J Hayes. Paroxysmal supraventricular tachycardia in the general population. *Journal of the American College of Cardiology*, 31(1):150–157, 1998.

[46] Maria A Pabon, Kevin Manocha, Jim W Cheung, and James C Lo. Linking arrhythmias and adipocytes: insights, mechanisms, and future directions. *Frontiers in Physiology*, page 1752, 2018.

[47] M. Pekala, N. Joshi, T.Y. Alvin Liu, N.M. Bressler, D. Cabrera De-Buc, and P. Burlina. Deep learning based retinal OCT segmentation. *Computers in Biology and Medicine*, 114:103445, 2019.

[48] Mike Pekala, Neil Joshi, TY Alvin Liu, Neil M Bressler, D Cabrera DeBuc, and Philippe Burlina. Deep learning based retinal OCT segmentation. *Computers in Biology and Medicine*, 114:103445, 2019.

[49] 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.

[50] Kunyang Sun, Haoqing Shi, Zhengming Zhang, and Yongming Huang. ECS-Net: Improving weakly supervised semantic segmentation by using connections between class activation maps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7283–7292, 2021.

[51] Hui Wang, Wei Kang, Austin P Bishop, Andrew M Rollins, Thomas Carrigan, Noah Rosenthal, and Mauricio Arruda. In vivo intracardiac optical coherence tomography imaging through percutaneous access: toward image-guided radio-frequency ablation. *Journal of Biomedical Optics*, 16(11):110505, 2011.

[52] Jing Wang, Wanyue Li, Yiwei Chen, Wangyi Fang, Wen Kong, Yi He, and Guohua Shi. Weakly supervised anomaly segmentation in retinal OCT images using an adversarial learning approach. *Biomedical Optics Express*, 12(8):4713–4729, 2021.

[53] Juan Wang and Bin Xia. Bounding box tightness prior for weakly supervised image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 526–536. Springer, 2021.

[54] Juan Xu, Hiroshi Ishikawa, Gadi Wollstein, and Joel S Schuman. 3D optical coherence tomography super pixel with machine classifier analysis for glaucoma detection. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3395–3398, 2011.

[55] Xinwen Yao, Yu Gan, Charles C Marboe, and Christine P Hendon. Myocardial imaging using ultrahigh-resolution spectral domain optical coherence tomography. *Journal of Biomedical Optics*, 21(6):061006, 2016.

[56] Tae Keun Yoo, Joon Yul Choi, and Hong Kyu Kim. Feasibility study to improve deep learning in OCT diagnosis of rare retinal diseases with few-shot classification. *Medical & Biological Engineering & Computing*, 59(2):401–415, 2021.

[57] Bingfeng Zhang, Jimin Xiao, Yunchao Wei, Kaizhu Huang, Shan Luo, and Yao Zhao. End-to-end weakly supervised semantic segmentation with reliable region mining. *Pattern Recognition*, 128:108663, 2022.

[58] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2921–2929, 2016.