Hybrid Focal Loss for Accurate Epicardium Segmentation and Mask Generation via Spherical Harmonics

Ibuki Naka [1], Yinhao Li [1], Yutaro Iwamoto [2], Jain Rahul Kumar [3], Xianhua Han [4], Atsuyuki Wada [5], Yuji Tezuka [5], Kiyosumi Maeda [5], Atsunori Kashiwagi [5], and Yen-Wei Chen [1,*]

[1] College of Information Science and Engineering, Ritsumeikan University, Osaka, Japan
[2] Engineering Informatics, Information and Communication Engineering, Osaka Electro-Communication University, Osaka, Japan
[3] Tiwaki co., Itd, Shiga, Japan
[4] Graduate School of Artificial Intelligence Science, Rikkyo University, Tokyo, Japan
[5] Omi Medical Center, Shiga, Japan
Email: is0569ev@ed.ritsumeai.ac.jp (I.N.); yin-li@fc.ritsumei.ac.jp (Y.L.); yiwamoto@osakaac.ac.jp (Y.I.); rahulkumarjain16@gmail.com (J.R.K.); hanxhua@rikkyo.ac.jp (X.H.); chen@is.ritsumei.ac.jp (Y.-W.C.)
*Corresponding author

***Abstract*—Epicardial Adipose Tissue (EAT) volume has been reported as a clinically relevant biomarker of metabolic risk, yet conventional diabetes assessment relies heavily on biochemical measurements that fluctuate with short-term physiological changes. Computed Tomography (CT)–based EAT quantification offers a more stable alternative; however, accurately delineating the epicardial boundary remains difficult due to its thin, low-contrast structure. Existing Convolutional Neural Network (CNN) approaches trained with Dice-based losses often overlook subtle boundary regions, leading to fragmented contours and degraded downstream EAT estimation. To address this limitation, we propose a Hybrid Focal Loss (HFL) that adaptively balances focal and regional penalties to enhance learning of fine epicardial structures. The loss is embedded in a two-stage framework: Stage 1 performs epicardium segmentation, and Stage 2 reconstructs a smooth mask via spherical harmonics interpolation to enforce global surface consistency. Experimental results on CT images from 15 patients demonstrate that HFL improves segmentation performance by 1.6% in F1-Score compared to Dice Binary Cross-Entropy (BCE) loss and further enhances reconstructed mask accuracy, reducing Hausdorff distance by 6.0 points. When applied to EAT extraction, HFL yields the highest F1 and Intersection over Union (IoU) scores across all evaluated losses, translating into improved agreement with the ground truth of EAT volumes. These findings demonstrate that HFL strengthens both boundary sensitivity and volumetric reliability, suggesting its value for robust EAT quantification and cardiometabolic risk assessment.**

***Keywords*—diabetes, epicardial adipose tissue, epicardium, Computed Tomography (CT) images, segmentation, hybrid focal loss, spherical harmonics**

I. Introduction

Diabetes mellitus is a chronic metabolic disorder characterized by sustained hyperglycemia resulting from impaired insulin secretion or sensitivity. Insulin, secreted by pancreatic β-cells, plays a central role in maintaining glucose homeostasis; its dysfunction leads to a range of systemic complications including retinopathy, nephropathy, and arteriosclerosis. Diabetes is typically classified into two major types: Type 1, caused by autoimmune destruction of β-cells, and Type 2, which is predominantly associated with genetic predisposition and lifestyle factors such as diet and physical inactivity. Although clinical diagnosis is conventionally performed using fasting plasma glucose or glycated hemoglobin (HbA1c) levels, these biochemical indicators are affected by short-term physiological fluctuations and comorbidities, thereby reducing diagnostic robustness. Consequently, the identification of more stable, imaging-based biomarkers is essential for reliable metabolic assessment.

Epicardial Adipose Tissue (EAT), a visceral fat depot located between the myocardium and the visceral pericardium, has recently emerged as a promising imaging biomarker for Type 2 diabetes risk. Unlike traditional blood-based markers, EAT reflects long-term metabolic status and is less sensitive to transient lifestyle changes, enabling consistent evaluation of cardiometabolic health. Furthermore, previous studies have reported a strong correlation between EAT volume and coronary atherosclerosis, cardiovascular events, and overall metabolic burden [1–3]. Hence, accurate EAT quantification is clinically valuable not only for predicting

Manuscript received October 1, 2025; revised October 17, 2025; accepted January 7, 2026; published May 27, 2026.

diabetes onset but also for assessing broader cardiovascular risk.

However, delineating the epicardial boundary in Computed Tomography (CT) images remains a challenging task. The epicardium forms a thin and low-contrast boundary separating EAT from subendocardial fat, often resulting in ambiguous intensity gradients and weak edge definition, as shown in Fig. 1. These factors hinder conventional segmentation algorithms based on Convolutional Neural Networks (CNNs), which typically rely on local intensity features and overlap-loss functions such as Dice BCE loss. Such methods struggle to capture fine boundary structures, leading to inaccuracies in downstream EAT quantification.

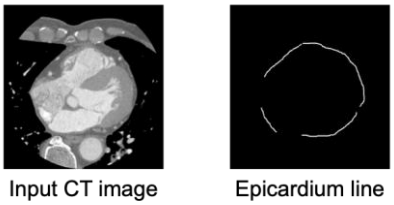

Fig. 1. Epicardium in the CT image.

To address these limitations, we propose a two-stage deep learning framework designed for accurate EAT segmentation and mask generation. In Stage 1, we introduce a Hybrid Focal Loss (HFL) that enhances model sensitivity to small and thin boundaries by combining focal weighting with overlap-based regularization. In Stage 2, a spherical harmonic–based interpolation is employed to reconstruct a smooth and continuous 3D surface from the segmented boundaries, thereby generating a robust binary mask of the enclosed subepicardial region. This combination enables both precise local boundary extraction and global shape consistency. In our previous work [4, 5], we developed a cascaded pipeline integrating CNN-based segmentation (e.g., UNet [6] and DenseUNet [7]) with spherical harmonics reconstruction, which demonstrated the feasibility of this hybrid approach. The present study extends that framework by introducing an improved loss design and a fully integrated 3D reconstruction stage. An overview of the proposed workflow, including Stage 1 segmentation and Stage 2 reconstruction leading to EAT extraction, is illustrated in Fig. 2.

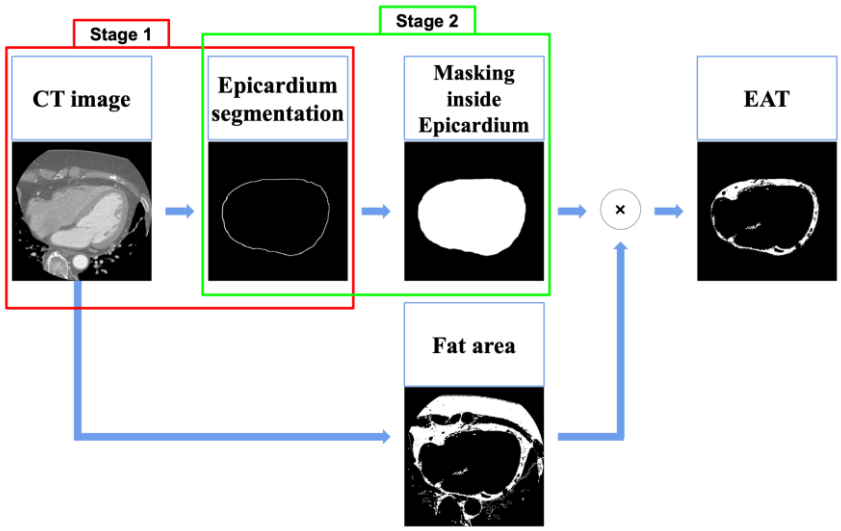

Fig. 2. Overview of our work.

The main contributions of this study are summarized as follows:

- Hybrid Focal Loss (HFL): We propose a boundary-aware loss function to improve sensitivity to small-scale and thin epicardial structures in Stage 1 segmentation.
- Spherical Harmonics Reconstruction: We employ a shape-consistent 3D reconstruction based on spherical harmonics to ensure robust mask generation and spatial continuity of the segmented subepicardial region, which encompasses EAT as well as surrounding non-fat tissues.

II. Related Work

A. Epicardial Adipose Tissue Segmentation

In recent years, there has been growing interest in the automatic quantification of EAT from Computed Tomography (CT) images. Early attempts employed conventional image processing methods, such as thresholding and region growing, to extract EAT regions. However, these approaches were highly sensitive to noise and variations in image contrast, resulting in limited reproducibility and poor boundary precision.

With the emergence of deep learning, CNN-based approaches substantially improved the accuracy and

robustness of EAT segmentation. Zhao *et al.* [4] proposed an automatic framework for epicardial boundary segmentation, demonstrating the feasibility of deep learning-based EAT analysis. Building upon this, Nagata *et al.* [5] introduced a two-stage segmentation network for the quantitative assessment of subepicardial adipose tissue, achieving improved delineation performance. More recently, large-scale studies [8–10] have generalized these approaches to non-contrast CT datasets, enabling more consistent and accurate fat quantification across diverse populations.

In addition to these developments, recent works published over the past three years further strengthened the clinical applicability of deep learning based EAT analysis. For instance, Oikonomou *et al.* [11] demonstrated that automated EAT quantification generalizes effectively across heterogeneous multi-center cohorts, supporting its potential for cardiometabolic risk stratification. Similarly, Jin KN *et al.* [12] reported that opportunistic EAT assessment can be performed using routine or low-dose CT scans, even under varying acquisition conditions, enabling broader integration into clinical workflows. Moreover, Kholi *et al.* [13] showed that segmentation frameworks incorporating uncertainty modeling support more reliable delineation in weakly defined anatomical regions. More recently, a 2024 study that introduced an uncertainty aware adversarial calibration framework specifically for EAT segmentation demonstrated that improving reliability in model optimization remains an active research topic [14], which is consistent with the motivation of our proposed loss design.

Despite these notable advances, precise segmentation of EAT remains challenging, primarily due to the thin and low-contrast nature of the epicardial boundary and the limited spatial extent of EAT regions. These difficulties underscore the need for dedicated architectures and advanced loss functions that can enhance model sensitivity to small and subtle structures.

B. Loss Functions for Segmentation Tasks

A wide range of loss functions has been proposed to enhance segmentation performance, particularly under conditions of class imbalance or when delineating thin and small anatomical structures. The following provides a concise overview of representative loss formulations commonly adopted in medical image segmentation.

1) Dice BCE loss

Dice loss, defined in Eq. (1), quantifies the spatial overlap between predicted and ground-truth masks. Here, p_i and g_i denote the predicted and ground-truth values at pixel i, respectively. In Eq. (2), N represents the total number of pixels within the mask, ensuring that Binary Cross-Entropy (BCE) term averages the pixel-wise loss over the entire image domain. While Dice loss effectively evaluates regional similarity, it does not explicitly optimize pixel-level accuracy. To address this limitation, Dice BCE loss, formulated in Eq. (3), integrates Dice and BCE components to jointly capture global overlap and local pixel-wise fidelity.

$$Dice = 1 - \frac{2\sum_i p_i g_i}{\sum_i p_i + \sum_i g_i} \tag{1}$$

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}(g_i \log p_i + (1 - g_i)(1 - \log p_i)) \tag{2}$$

$$Dice\ BCE\ loss = Dice + BCE \tag{3}$$

2) Tversky and focal variants

Tversky loss [15], depicted in Eq. (5), extends Dice loss by introducing weighting parameters α and $\beta = (1 - \alpha)$ to balance penalties for false positives and false negatives. This formulation is particularly effective in handling class imbalance and enhancing boundary delineation in small anatomical structures.

$$Tversky\ Index = \frac{\sum_i p_i g_i}{\sum_i p_i g_i + \alpha\sum_i p_i(1-g_i) + \beta\sum_i (1-p_i) g_i} \tag{4}$$

$$Tversky\ loss = 1 - Tversky\ Index \tag{5}$$

To further emphasize the learning of hard examples, Focal Tversky loss (FTL) [16] introduces a focusing parameter γ that modulates the relative contribution of misclassified pixels. As defined in Eq. (6), this loss function enhances model sensitivity to challenging regions, particularly those with small or ambiguous boundaries.

$$Focal\ Tversky\ loss = \sum(1 - Tversky\ Index)^{\gamma} \tag{6}$$

Similarly, Focal loss (FL) [17], originally proposed for classification tasks, reduces the relative contribution of easily classified samples and concentrates the learning process on hard negatives, where p_i denotes the predicted probability of the true class for the *i*-th sample. As expressed in Eq. (7), the focusing parameter γ controls the strength of this modulation, while α adjusts class balance between foreground and background pixels.

$$Focal\ Loss = -(1 - p_i)^{\gamma} \log(p_i) \tag{7}$$

III. METHOD

A. Hybrid Focal Loss

The Hybrid Focal Loss (HFL), defined as a weighted combination of FTL and FL, is employed to optimize the epicardial segmentation network. While FTL enables explicit control over the trade-off between false positives and false negatives at the regional level, FL mitigates class imbalance at the pixel level by emphasizing hard-to-classify samples. Consequently, the weighted integration of FTL and FL enhances model robustness to class imbalance across both pixel-wise and regional

scales. The mathematical formulation of HFL is given in Eq. (8).

$$HFL = \lambda \times FTL + (1 - \lambda) \times FL \tag{8}$$

B. Spherical Harmonics

The masking of the subepicardial region using spherical harmonics was performed the methodological pipeline outlined below:

1) Epicardial contour extraction

Epicardial contours were delineated on a series of contiguous CT slices, yielding a discrete set of N three-dimensional boundary points, denoted as $\{(x_i, y_i, z_i)\}_{i=1}^{N}$.

2) Centroid alignment

Before converting the boundary points into spherical coordinates, all points were translated such that the geometric centroid of the volume became the origin of the coordinate system. For a volume of size (H, W, D), the centroid was defined as:

$$centroid = \left(\frac{H}{2}, \frac{W}{2}, \frac{D}{2}\right) \tag{9}$$

Each boundary point (x_i, y_i, z_i) was then shifted by subtracting the centroid:

$$(x_i', y_i', z_i') = \begin{pmatrix} x_i - centroid_x, \\ y_i - centroid_y, \\ z_i - centroid_z \end{pmatrix} \tag{10}$$

The spherical coordinates were computed from the translated points (x_i', y_i', z_i').

3) Coordinate transformation

Each boundary point was mapped into spherical coordinates $(r_i, \theta_i, \emptyset_i)$ using the following equations:

$$r_i = \sqrt{x_i'^2 + y_i'^2 + z_i'^2},$$

$$\theta_i = \arccos\left({}^{z_i}/_{r_i}\right),$$

$$\emptyset_i = \arctan\left({}^{y_i}/_{x_i}\right) \tag{11}$$

4) Spherical harmonic approximation

The radial function $r(\theta, \emptyset)$ was approximated by a truncated spherical harmonics series up to degree L:

$$r(\theta, \emptyset) \approx \sum_{l=0}^{L} \sum_{m=-l}^{l} c_l^m Y_l^m(\theta, \emptyset) \tag{12}$$

where

$$Y_l^m(\theta, \emptyset) = N_{lm} P_l^m(\cos\theta) e^{im\emptyset} \tag{13}$$

$$N = \sqrt{\frac{(2l+1)}{4\pi} \frac{(l-m)!}{(l+m)!'}} \tag{14}$$

and P_l^m denotes the associated Legendre function of order l and degree m.

5) Coefficient estimation

The coefficient c_l^m was obtained by minimizing the sum of squared residuals between the sampled radii r_i, and the expansion:

$$\{c_{lm}\} = argmin \sum_{i=1}^{N} \left| r_i - \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{lm} Y_l^m(\theta_i, \emptyset_i) \right|^2 \tag{15}$$

This yields the normal equations in matrix form:

$$c = (Y^H Y)^{-1} Y^H r \tag{16}$$

where $Y \in \mathbb{C}^{N \times M}$ contains sampled harmonics (with $M = (L+1)^2$) and $r \in \mathbb{R}^N$ is the vector of measured radii.

6) Surface interpolation

A dense grid of $(\theta, \emptyset)$ values over the unit sphere, and the interpolated surface, represented as $\hat{r}(\theta, \emptyset)$, was evaluated as:

$$\hat{r}(\theta, \emptyset) = \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{lm} Y_l^m(\theta, \emptyset) \tag{17}$$

7) Cartesian reconstruction

The reconstructed epicardial surface was obtained by mapping $\hat{r}(\theta, \emptyset)$ back to Cartesian coordinates:

$$\hat{x}(\theta, \emptyset) = \hat{r}(\theta, \emptyset) \sin\theta \cos\emptyset,$$

$$\hat{y}(\theta, \emptyset) = \hat{r}(\theta, \emptyset) \sin\theta \sin\emptyset,$$

$$\hat{z}(\theta, \emptyset) = \hat{r}(\theta, \emptyset) \cos\theta \tag{18}$$

thereby producing a smooth, closed surface suitable for subsequent mesh generation and volume rendering.

8) Implementation details

In this study, the order of spherical harmonics was set to L = 8. This value was chosen to balance the representation of high-frequency geometric details and the numerical stability of the reconstruction. In preliminary experiments, higher orders (e.g., $L \geq 10$) frequently caused undesirable oscillations in the reconstructed radius field, resulting in divergence or extreme outliers in predicted radii. Conversely, lower orders (e.g., $L \leq 6$) could not adequately preserve fine anatomical structures. Therefore, L = 8 was adopted as an appropriate compromise.

Topology violations such as self-intersections or non-closed surfaces were not explicitly handled in the current implementation. However, as a preventive measure, we employed a moderate harmonic order (L = 8), which effectively reduces oscillation-induced instabilities. Furthermore, when the reconstructed contour contained multiple connected components or inner holes, only the largest outer contour was retained for the subsequent surface filling process.

IV. Experiments

A. Dataset

The CT dataset used in this study was provided by Omi Medical Center and was acquired using a LightSpeed VCT scanner. The dataset consisted of 2544 axial CT slices (512×512, 0.625-mm slice thickness) collected from a total of 15 patients.

To prevent data leakage across slices, we performed a patient-level split. Among the 15 patients, 12 were assigned to the training set, 2 were assigned to the validation set, and the remaining 1 patient was used exclusively for testing. This allocation resulted in 2030 training slices, 349 validation slices, and 165 consecutive test slices that cover the complete cardiac region of the test subject. An example of the annotated epicardial boundary is shown in Fig. 1.

Although the dataset originates from a single medical institution, its multi-patient structure provides heterogeneous anatomical variation suitable for both model training and patient-level evaluation. A more detailed discussion of dataset limitations and future extensions is provided in Section VI.

B. Implementation Details

The epicardium occupies a very small spatial region compared to the background, leading to severe class imbalance. To address this, Tversky-based and Focal-based losses were employed with the parameter settings summarized in Table I. The Hybrid Focal Loss (HFL) was implemented with a weighting of $\lambda = 0.3$. UNet was trained for epicardial segmentation with Adam optimizer with a learning rate of 0.0001, a batch size of 2, and 100 epochs. All experiments are conducted using a single NVIDIA RTX 8000.

TABLE I. Quantitative Evaluation of the Proposed Method and some Loss Functions of Epicardium Segmentation (Stage 1) and Mask Generation via Spherical Harmonics (Stage 2)

Loss Function	F1 (Stage 1)	F1 (Stage 2)	Hd95 (Stage 2)
Dice BCE	50.9% ± 0.11 95%CI: [0.49, 0.53)	87.0% ± 0.21 95%CI: [0.84, 0.90]	32.3 ± 43.40 95%CI: [25.650, 38.895]
Tversky (β = 0.9)	49.2% ± 0.11 95%CI: [0.48, 0.51]	81.2% ± 0.26 95%CI: [0.77, 0.85]	41.8 ± 56.20 95%CI: [33.22, 50.37]
Focal Tversky (γ = 0.8)	47.4% ± 0.10 95%CI: [0.46, 0.49)	88.6% ± 0.21 95%CI: [0.85, 0.92]	30.9 ± 49.85 95%CI: [23.30, 38.51]
Focal (γ = 2.0)	44.3% ± 0.12 95%CI: [0.43, 0.46]	88.1% ± 0.20 95%CI: [0.85, 0.91]	26.1 ± 38.59 95%CI: [20.20, 31.98]
Ours ($\lambda = 0.3$)	**52.5%** ± 0.10 95%CI: [0.51, 0.54]	**91.2%** ± 0.20 95%CI: [0.88, 0.94]	**20.1** ± 41.08 95%CI: [13.80, 26.34]

C. Results of Epicardium Segmentation and Mask Generation

To evaluate the effectiveness of HFL in epicardial segmentation, we compared it with conventional loss functions, including Dice BCE, Tversky, Focal Tversky, and Focal loss. In Stage 1, epicardium segmentation accuracy was evaluated, and in Stage 2, binary masks were generated via spherical harmonic interpolation based on Stage 1 outputs. The quantitative results are summarized in Table I. HFL achieved the highest overall performance.

A paired t-test between proposed HFL ($\lambda = 0.3$) and Dice BCE loss in Stage 1 yielded $t = 3.38$, $p = 0.0009$, and an effect size of Cohen's d = 0.26, confirming a statistically significant improvement. Qualitative examples in Fig. 3 show that HFL produces smoother and more continuous epicardial boundaries with reduced artifacts, and the reconstructed masks exhibit improved boundary continuity and global shape consistency compared to competing methods.

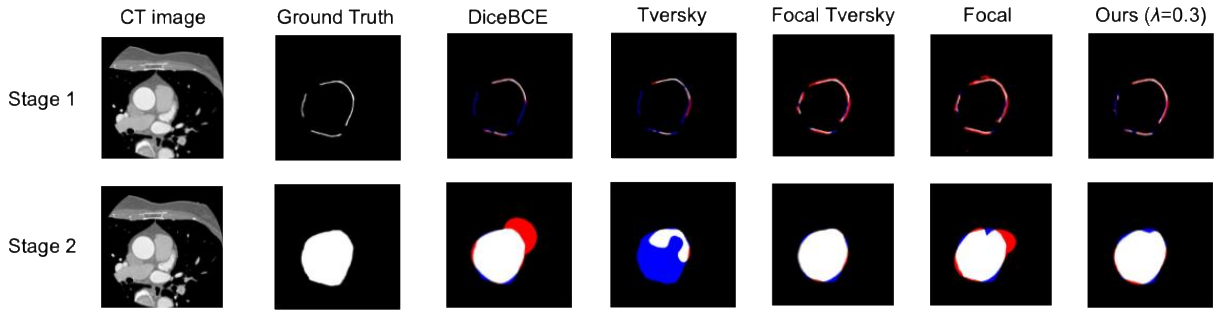

Fig. 3. Qualitative comparison of our two-stage pipeline. Stage 1 shows epicardium boundary segmentation, and Stage 2 shows the generated subepicardial mask. Columns present the CT image, ground truth, and the results of different loss functions (DiceBCE, Tversky, Focal Tversky, Focal, and ours with $\lambda = 0.3$). Predictions are visualized as an error map with respect to the ground truth: white = TP, red = FP, blue = FN.

A paired t-test between the proposed HFL ($\lambda = 0.3$) and Dice BCE loss in Stage 2 yielded $t = 9.35$, $p = 6.30 \times 10^{-17}$, with an effect size of Cohen's d = 0.73, confirming a statistically significant and practically meaningful

improvement in slice-wise F1-Score. For boundary accuracy, HFL also significantly reduced Hd95, with a mean decrease of 8.31 pixels ($t = -7.32$, $p = 1.06 \times 10^{-11}$, Cohen's d = 0.57), indicating improved boundary localization and greater stability in the reconstructed epicardial shape.

D. Discussion of Stage 1 and Stage 2

Although Focal Tversky and Focal losses resulted in slightly lower segmentation scores than Dice Binary Cross-Entropy (BCE) in Stage 1, they outperformed it after spherical harmonic reconstruction in Stage 2. Focal-based losses emphasize hard pixels, such as those belonging to minority classes or difficult boundaries, leading to mild over-segmentation and reduced precision in Stage 1. However, these denser boundary regions provide additional control points for spherical harmonic fitting, stabilizing the reconstruction process and suppressing local irregularities. We attribute the superior Stage 2 performance of HFL to its balanced treatment of pixel-wise and regional errors, achieving both stable interpolation and accurate global shape recovery.

E. Ablation Study of Stage 1 and Stage 2

1) Effect of λ on optimization behavior and segmentation accuracy

We also provide ablation studies on different λ of HFL. The results are summarized in Table II and Fig. 4. When HFL was configured with a weighting coefficient of $\lambda = 0.9$, Stage 1 segmentation performance yielded a lower F1-Score compared with Dice BCE loss. As indicated by Eq. (8), HFL with $\lambda = 0.9$ places stronger emphasis on FTL component. The focusing parameter λ in FTL was set to 0.8, which improved the stability of training by suppressing gradient fluctuations. However, this relatively small λ value also reduced the emphasis on hard regions to classify, resulting in insufficient penalization of misclassified areas. Consequently, the model trained with HFL ($\lambda = 0.9$) achieved more stable but less discriminative learning, leading to slightly inferior accuracy compared with Dice BCE in Stage 1.

TABLE II. QUANTITATIVE EVALUATION OF THE ABLATION STUDY WITH DIFFERENT VALUES OF $\boldsymbol{\lambda}$ FOR EPICARDIUM SEGMENTATION (STAGE 1) AND MASK GENERATION VIA SPHERICAL HARMONICS (STAGE 2)

Loss Function	F1 (Stage 1)	F1 (Stage 2)	Hd95 (Stage 2)
HFL ($\lambda = 0.1$)	51.9% ± 0.09 95%CI: [0.51, 0.53]	89.6% ± 0.21 95%CI: [0.86, 0.93]	22.3 ± 43.15 95%CI: [15.70, 28.87]
HFL ($\lambda = 0.5$)	**52.7% ± 0.09** 95%CI: [0.41, 0.510]	90.0% ± 0.22 95%CI: [0.87, 0.93]	23.0 ± 45.82 95%CI: [16.01, 30.00]
HFL ($\lambda = 0.7$)	51.6% ± 0.10 95%CI: [0.50, 0.53]	90.7% ± 0.21 95%CI: [0.88, 0.94]	21.9 ± 43.82 95%CI: [15.19, 28.57]
HFL ($\lambda = 0.9$)	49.5% ± 0.09 95%CI: [0.48, 0.51]	90.7% ± 0.21 95%CI: [0.87, 0.94]	**17.9 ± 32.02** 95%CI: [12.99, 22.77]
Ours ($\lambda = 0.3$)	52.5% ± 0.10 95%CI: [0.51, 0.54]	**91.2% ± 0.20** 95%CI: [0.88, 0.94]	20.1 ± 41.08 95%CI: [13.80, 26.34]

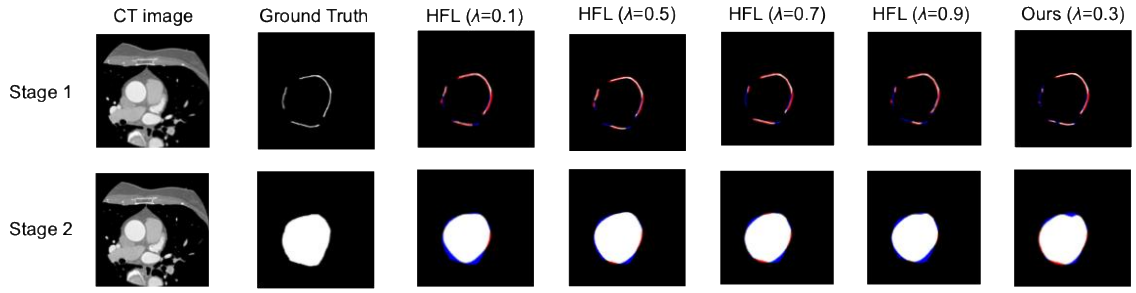

Fig. 4. Ablation study of Hybrid Focal Loss (HFL) with different \lambda values in our two-stage pipeline. Stage 1 corresponds to epicardium boundary segmentation, and Stage 2 corresponds to subepicardial mask generation. Columns present the CT image, ground truth, and the color-coded error maps for HFL with λ = 0.1, 0.5, 0.7, and 0.9, as well as our final configuration (λ = 0.3). Error maps are shown against the ground truth: white = TP, red = FP, blue = FN.

2) Trade-off between contour smoothness and overlap accuracy

Interestingly, although $\lambda = 0.9$ tends to produce visually smoother and fully connected epicardial contours, the predicted boundary becomes systematically thicker than the ground truth, extending slightly beyond the true epicardial surface on both the inner and outer sides. This boundary drift increases false-positive regions and leads to poorer overlap performance. In contrast, λ = 0.3 preserves fine boundary variations, resulting in contours that more closely align with the ground truth. Fig. 5 illustrates that λ = 0.3 retains local structural variability more faithfully than λ = 0.9, despite its less visually uniform appearance. Consistent with this visual tendency, F1-Score on this slice increases by 3.0% when λ = 0.3 is used instead of $\lambda = 0.9$.

F. Epicardial Adipose Tissue Segmentation

After the mask image was generated through spherical harmonic reconstruction in Stage 2, the epicardial adipose tissue region was extracted by performing a logical and operation between the generated mask and the fat tissue map obtained from the CT images. As summarized in Table III, the proposed method, HFL (λ = 0.3), achieved the highest segmentation accuracy among all conventional loss functions.

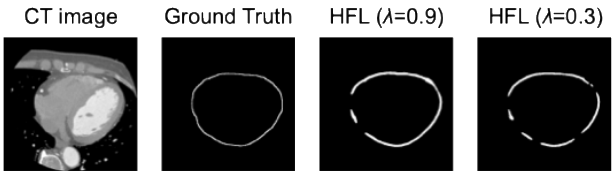

Fig. 5. Trade-off between contour smoothness and overlap accuracy with different λ values in HFL. From left to right: CT image, ground truth, and predictions obtained with λ = 0.9 and λ = 0.3. While λ yields a smoother and more connected contour, it produces a systematically thicker boundary that drifts beyond the ground truth, increasing false positives. In contrast, λ = 0.3 preserves fine boundary variations and aligns more closely with the ground truth; the F1-Score on this slice improves by 3.0% compared with λ = 0.9.

TABLE III. QUANTITATIVE EVALUATION OF EPICARDIAL ADIPOSE TISSUE IMAGES GENERATED THROUGH AND OPERATION BETWEEN MASK IMAGES FROM STAGE 2 AND FAT IMAGES

Loss Function	F1	IoU
Dice BCE	88.1% ± 0.15 95%CI: [0.86, 0.90]	81.1% ± 0.19 95%CI: [0.78, 0.84]
Tversky (β=0.9)	87.1% ± 0.14 95%CI: [0.85, 0.89]	79.3% ± 0.18 95%CI: [0.77, 0.82]
Focal Tversky (γ=0.8)	89.0% ± 0.11 95%CI: [0.87, 0.91]	81.6% ± 0.15 95%CI: [0.79, 0.84]
Focal (γ=2.0)	90.2% ± 0.10 95%CI: [0.89, 0.92]	83.3% ± 0.13 95%CI: [0.81, 0.85]
Ours (λ = 0.3)	**92.7%** ± 0.09 95%CI: [0.91, 0.94]	**87.4%** ± 0.12 95%CI: [0.86, 0.89]

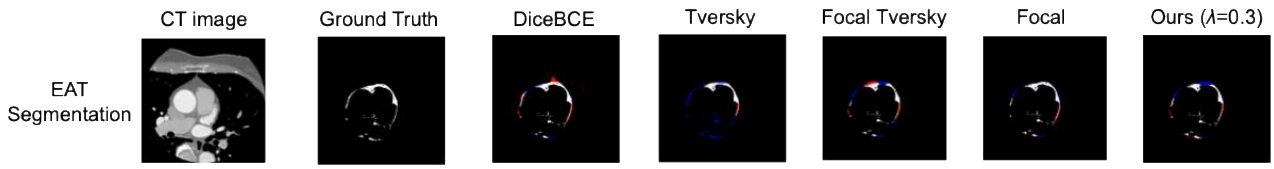

Fig. 6. Qualitative comparison of Epicardial Adipose Tissue (EAT) segmentation. From left to right: CT image, ground truth, and color-coded error maps for DiceBCE, Tversky, Focal Tversky, Focal, and our method (HFL, λ = 0.3). Error maps are shown with respect to the ground truth: white = True Positives (TP), red = False Positives (FP), and blue = False Negatives (FN).

To quantify the statistical significance of these improvements, paired t-test were conducted between HFL and Dice BCE across all 165 slices. For F1-Score, HFL achieved a significantly higher performance, yielding an absolute improvement of 0.046 (mean 4.6%; 95%CI: 0.031 – 0.061; t = 6.16, p = 5.47×10^{-9}; Cohen's effect = 0.48). A similar gain was observed for IoU, with an absolute improvement of 0.063 (mean 6.3%; 95%CI: 0.043 – 0.083; t = 6.17, p = 5.05×10^{-9}; Cohen's d = 0.48). These results indicate that HFL provides statistically significant enhancement in both overlap accuracy and region stability, thereby improving the reliability of EAT extraction.

Quantitative examples in Fig. 6 further show that HFL produces cleaner boundaries and reduces false detection relative to competing methods. This improvement in segmentation accuracy directly enhances the precision of EAT volume quantification, reducing inter-case variability and enabling more reliable estimation of patient-specific cardiometabolic risk. Because EAT volume is a clinically recognized biomarker for metabolic and cardiovascular disorders, the proposed method contributes to improving the consistency and clinical interpretability of EAT-based risk assessment.

G. Volume-Based Analysis and Clinical Relevance

To evaluate whether the improvements in segmentation accuracy translate into clinically meaningful differences, we conducted a slice-wise EAT volume analysis using 165 axial slices from the test patient. The physical voxel spacing ($0.3516\times0.3516\times0.625$ mm^3) was used to convert the binary masks into volumetric measurements. A Bland-Altman analysis was performed to compare Dice BCE-based and HFL-based estimates with the ground truth.

Dice BCE loss exhibited a bias of −0.0549 mL with 95% Limits of Agreement (LOA) from −0.268 to 0.158 mL. In contrast, the proposed HFL nearly estimated systematic bias (bias = 0.0014 mL) and achieved a

substantially narrower LOA of −0.053 to 0.056 mL. Furthermore, the mean absolute slice-wise volume error decreased markedly from 0.0639 mL (Dice BCE) to 0.0190 mL (HFL), representing a 70% reduction, while the median error decreased from 0.0209 mL to 0.0119 mL.

These findings indicate that HFL not only improves pixel-wise segmentation accuracy but also enhances the stability and clinical reliability of EAT volume estimation. Given that EAT volume is a recognized biomarker for cardiometabolic risk, the improved agreement demonstrated by HFL suggests its strong potential for reliable EAT-based risk stratification in clinical settings.

Although this volumetric analysis was performed on a single test patient, the results highlight the feasibility and clinical promise of the proposed approach. Further work will involve comprehensive validation across multi-patient, multi-center datasets to further establish the generalizability and robustness of the method.

V. Conclusion

In this study, we proposed Hybrid Focal Loss (HFL), a novel loss function designed to enhance the segmentation of thin and ambiguous epicardial boundaries in CT images. In Stage 1, HFL improved the F1-Score by 1.6% over the conventional Dice BCE loss, demonstrating higher sensitivity to hard-to-classify regions. To further enforce anatomical consistency, Stage 2 employed spherical harmonic reconstruction, achieving an additional 2.6% increase in F1-Score and a 6.0-point reduction in Hd95. This two-stage framework effectively combines local boundary refinement with global surface regularization.

Following reconstruction, Epicardial Adipose Tissue (EAT) regions were extracted by applying a logical AND operation between the generated mask and the CT-derived fat map. Under the optimal setting ($\lambda = 0.3$), the proposed pipeline achieved the highest overall EAT segmentation performance, with gains of 2.5% in F1 and 4.1% in IoU, and reduced inter-case variability in EAT volume quantification. Given the strong clinical relevance of EAT volume as a biomarker for metabolic and cardiovascular risk, the proposed framework represents a robust and clinically meaningful solution for epicardial boundary extraction and quantitative EAT analysis.

Despite these promising results, several limitations and future directions remain. First, the present dataset includes 15 patients from a single institution; therefore, future work will extend the evaluation to larger, multi-center cohorts to better assess generalizability across diverse anatomical variations, imaging conditions, and disease backgrounds. Second, although spherical harmonics reconstruction provided stable epicardial surface estimation, explicit correction of topology violations was not implemented in the current pipeline. Future extensions will incorporate dedicated self-intersection detection, mesh-repair procedures, and topological regularization to ensure fully consistent and watertight epicardial surfaces. These improvements are expected to further enhance robustness and clinical applicability of spherical-harmonics-based epicardial modeling and EAT quantification.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Ibuki Naka: Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualization, Writing—original draft. Yinhao Li, Yutaro Iwamoto, Jain Rahul Kumar, Xianhua Han: Methodology, Validation, Writing—review and editing. Atsuyuki Wada, Yuji Tezuka, Kiyosumi Maeda, Atsunori Kashiwagi: Resources (dataset provision), Clinical interpretation, Writing—review and editing. Yen-Wei Chen: Supervision, Project administration, Writing—review and editing. All authors read and approved the final manuscript.

Funding

This work was supported in part by the Grant in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant No. 20KK0234, No. 21H03470, and No. 20K21821.

References

[1] M. Shimabukuro *et al.*, "Epicardial adipose tissue volume and adipocytokine imbalance are strongly linked to human coronary atherosclerosis," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 33, vol. 5, pp. 1077–1084, 2013.

[2] A. A. Mahabadi *et al.*, "Association of epicardial fat with cardiovascular risk factors and incident myocardial infarction in the general population: The Heinz Nixdorf recall study," *Journal of the American College of Cardiology*, vol. 61, no. 13, pp. 1388–1395, 2013.

[3] N. Gaibazzi, M. Rigo, and M. Reverberi, "Epicardial fat thickness predicts cardiovascular events in patients with coronary artery disease," *International Journal of Cardiology*, vol. 167, no. 5, pp. 2232–2238, 2013.

[4] Z. Ziyu *et al.*, "Automatic segmentation of visible epicardium using deep learning in CT image," in *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery*, vol. 1, Springer International Publishing, 2020, pp. 577–584.

[5] T. Nagata, "Quantative analysis of subepicardial adipose tissue using a two-stage segmentation network and its system development," Master's Thesis, Ritsumeikan University, 2021.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference*, Munich, Germany, 2015. pp. 234–241.

[7] L. Xiaomeng *et al.*, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.

[8] T. Commandeur, K. Goeller, M. Betancur *et al.*, "Deep learning for quantification of epicardial and thoracic adipose tissue from non-contrast CT," *Radiology: Artificial Intelligence*, vol. 1, no. 6, e190045, 2019.

[9] H. Wen, S. Sun, J. He *et al.*, "Automated quantification of epicardial adipose tissue on non-contrast cardiac CT using a deep learning approach," *European Heart Journal—Digital Health*, vol. 2, no. 3, pp. 363–371, 2021.

[10] F. Liu, Y. Guo, L. Wang *et al.*, "Epicardial fat segmentation and quantification in CT using deep learning," *Medical Physics*, vol. 49, no. 3, pp. 1554–1565, 2022.

[11] E. K. Oikonomou, M. Marwan, S. Kotanidis *et al.*, "Deep learning enables automated quantification of epicardial adipose tissue and its association with cardiometabolic risk," *European Heart*

Journal—Cardiovascular Imaging, vol. 24, no. 3, pp. 345–357, 2023.
[12] K. N. Jin, J. Lee, M. S. Park *et al.*, "Opportunistic epicardial fat quantification from routine chest CT using deep learning for cardiometabolic risk assessment," *Scientific Reports*, vol. 12, no. 1, pp. 1–10, 2022.
[13] S. Kohl, D. Bonekamp, H. Biedermann *et al.*, "Probabilistic U-Net: Deep learning-based segmentation with uncertainty estimation for medical image analysis," *Nature Machine Intelligence*, vol. 2, pp. 684–692, 2020.
[14] Z. Kai *et al.*, "Uncertainty-driven and adversarial calibration learning for epicardial adipose tissue segmentation," in *Proc. 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, IEEE, 2024. pp. 1561–1566.
[15] S. S. M. Salehi, D. Erdogmus, A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," in *International Workshop on Machine Learning in Medical Imaging*, Cham: Springer International Publishing, 2017, pp. 379–387.
[16] N. Abraham and N. M. Khan, "A novel focal Tversky loss function with improved attention u-net for lesion segmentation," in *Proc. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE, 2019, pp. 683–687.
[17] L. Tsung-Yi *et al.*, "Focal loss for dense object detection," in *Proc. Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.