Physiology-aware PolySnake For Coronary Vessel Segmentation

Yizhe Ruan[1,2], Lin Gu[2,1], Yusuke Kurose[1,2], Junichi Iho[3], Youji Tokunaga[3],
Makoto Horie[3], Yusaku Hayashi[3], Keisuke Nishizawa[3],
Yasushi Koyama[3,2], Tatsuya Harada[1,2]
[1]The University of Tokyo
[2]RIKEN Center for Advanced Intelligence Project
[3]Sakurabashi Watanabe Advanced Healthcare Hospital
ruanyizhe@mi.t.u-tokyo.ac.jp

Abstract

Coronary artery disease (CAD) is a significant health risk that requires early detection for effective treatment. While recent advances in deep learning have shown promise in automating CAD detection from coronary computed tomography angiography (CCTA) images, the accurate segmentation of coronary vessels remains a challenge, particularly due to the imbalanced presence of plaque in unhealthy vessels. This paper introduces a physiology-aware approach[1] to coronary vessel segmentation that addresses these challenges. Our proposed pipeline consists of three main components. First, a hybrid UNeXt architecture is designed to segment artery boundaries and predict initial boundary contours by leveraging 3D spatial relations among adjacent slices. Second, we introduce multi-class circular convolution for iterative contour deformation, which generates well-connected contour pairs of the artery wall's inner and outer boundaries through iterative refinement. Finally, we propose a focal smooth L1 loss function to handle the implicit class imbalance caused by plaque in unhealthy vessels and to enhance the robustness of the physiology-aware polysnake network by explicitly limiting the accuracy of initial contours. Extensive evaluations demonstrate that our methods significantly improve model performance, achieving state-of-the-art results in coronary vessel segmentation.

1. Introduction

Coronary artery disease (CAD) is a prevalent cardiovascular condition characterized by the accumulation of plaque in the coronary arteries, resulting in diminished blood flow to the heart and representing a significant global cause of mortality [2]. Coronary computed tomography angiography (CCTA) serves as a non-invasive imaging modality commonly utilized for visualizing stenosis within the coronary artery lumen [15,20]. The precise segmentation of the coronary vessel wall from CCTA images is indispensable for accurately diagnosing CAD, as the measurement of stenosis diameter constitutes a critical clinical criterion [3]. Recent advancements in deep learning methods have demonstrated promising results, showcasing their agreement with expert readers in coronary segmentation tasks. [12]

While various approaches, such as the Lucy-Richardson method and the 2D median filter [22, 23], have been proposed for artery wall segmentation, these methods remain susceptible to noise and necessitate substantial fine-tuning. Deep-learning techniques, such as by U-Net [19] and its variants [1, 4, 10, 24, 27], have emerged as promising alternatives. Notably, the hybrid ResU-net architecture has been introduced as a solution to encode 3D volumes of CCTA and decode them into 2D results, achieved through a thoughtful design of the 3D encoder and 2D decoder parameters within the ResU-net [9] structure. The network has demonstrated a robust capacity for segmenting coronary vessels from the CCTA dataset.

Moreover, a recent methodology leveraging the hierarchical ConvLSTM network [12] has been introduced for computing plaque volume and assessing stenosis severity from CCTA data. This approach has demonstrated promise in coronary artery wall segmentation by employing the ConvLSTM Cell to learn 3D consistency and incorporating an attention head to reorganize both 3D and 2D information. While excelling in weighted average evaluation, these methods still face challenges in effectively addressing implicit class imbalance within plaque-containing samples and tend to overlook contour consistency, placing primary emphasis on pixel-wise accuracy.

In recent years, contour-based segmentation methods have experienced rapid growth, although they are mainly applied in common datasets. Deep snake [17] employs cir-

[1]https://github.com/opensourcetorch/Physiology-aware-PolySnake

cular convolution for contour deformation on detected objects, ensuring the smooth continuity of predicted contours. Additionally, PolySnake [6] utilizes a GRU-based [5] network to iteratively deform learnable initial contours into stable states.

To further ensure contour consistency and robust performance in imbalance samples, we propose a Physiology-aware PolySnake pipeline for artery wall segmentation in coronary computed tomography angiography (CCTA), presenting a contour-based segmentation strategy rooted in the PolySnake method that explicitly considers the physiological characteristics of coronary arteries. Our proposed method is structured into three main components.

First, by considering the 3D spatial relations in adjacent CT slices, we designed a hybrid UNeXt architecture to segment artery boundaries. This architecture employs a 3D-encoder-2D-decoder structure for feature learning in adjacent slices, as well as for segmentation and the prediction of initial contour pairs in the target slice. It is worth noting that this architecture can be viewed as a variant of UNeXt [24], a parameter-efficient UNet-like network known for its state-of-the-art performance in medical segmentation tasks.

Secondly, recognizing the tensile structure of the artery wall and the intricate relationship between the exterior and interior contours of coronary artery vessels, we adopt PolySnake method [6] and introduce a refinement to the contour deformation process. We propose multi-class circular convolution as a replacement for the original circular convolution. This modification enables the direct prediction of contour offset pairs, employing multi-class circular convolution to facilitate simultaneous feature learning on both the outer and inner contours.

Third, inspired by focal loss [13], we introduce a focal smooth L1 loss to enhance model performance in the presence of imbalanced unhealthy samples containing calcified or non-calcified plaques. This loss function also explicitly controls the accuracy of initial contours in training, thereby increasing the model's robustness for contour deformation. Note that, unlike focal loss, we utilize the $\tanh$ function to normalize positive loss values within the range of 0 to 1.

Our approach undergoes testing on a private dataset comprising 50-patient coronary computed tomography angiography scans with curved planar reconstruction (CCTA-CPR) of the arteries. The CT samples are broadly categorized into three classes: healthy slices without plaque, slices with calcified plaque, and slices with non-calcified plaque. Aligned with clinical objectives, we compute and compare the model's macro-average scores across distinct data categories using Hausdorff Distance (HD) and Dice Score (DS). Ablation studies and extensive experiments demonstrate that our method significantly enhances model performance at each stage, achieving state-of-the-art results.

The key contribution of this paper is: we propose Physiology-aware PolySnake network for coronary vessel segmentation and the extensive experiments show our method achieves state-of-the-art performance in artery vessel segmentation.

2. Related Work

2.1. Segmentation In Coronary Artery

Segmentation of coronary arteries is crucial for diagnosing and treating cardiovascular diseases. Most research has focused on 3D segmentation in CCTA. Methods with customized networks [8, 16, 28] have shown promising results in 3D segmentation. However, there is a noticeable gap in research on the cross-sectional segmentation of coronary arteries. Cross-sectional segmentation is essential for detailed visualization of vessel morphology and plaque characterization. Some studies [9, 12] have explored this area, achieving notable results.
While 3D segmentation has been extensively studied, cross-sectional segmentation remains under-explored and offers significant potential for improving cardiovascular imaging and diagnosis.

2.2. Contour-based Instance Segmentation

Contour-based methods aim to predict a series of coordinates that outline object boundaries, commonly used in general object detection tasks. In this paper, we introduce a contour-based approach specifically designed for coronary vessel segmentation, considering its anatomical context.

Recent works [11, 17] utilize Cartesian coordinates to represent vertices, training them to align with the actual object boundaries. Typically, Deep Snake [17] initializes contours based on box predictions and then undergoes several contour deformation steps for segmentation. Expanding on Deep Snake, Dance [14] improves the alignment between predicted and target contours and introduces an attentive deformation mechanism.

In a recent development, E2EC [26] employs a unique learnable contour initialization architecture, demonstrating remarkable performance. Inspired by this, PolySnake incorporates an iterative and progressive mechanism into object contour learning, achieving robust contour estimation in general object detection.

In this paper, we refine the PolySnake network for medical segmentation tasks and introduce the multi-class circular convolution to replace the original circular convolution. This approach simultaneously considers the relationship between inner and outer boundaries of the artery wall, resulting in more accurate and resilient contour predictions.

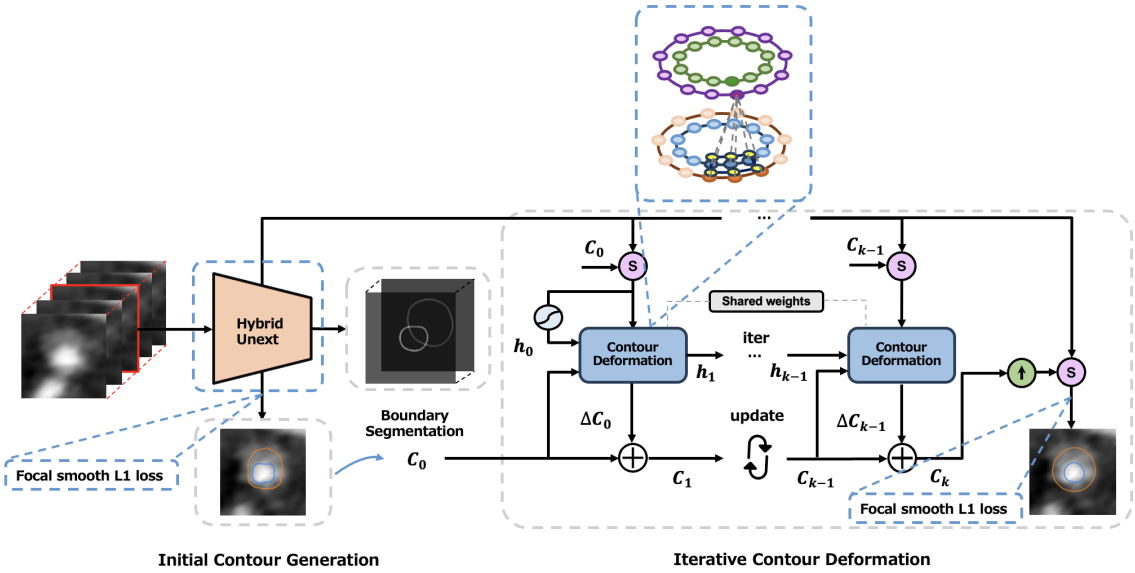

Figure 1. Physiology-aware PolySnake is outlined with three key modules: Boundary Segmentation (BS), Initial Contour Generation (ICG), and Iterative Contour Deformation (ICD). Starting with a 3D CT volume input V, the ICG module, based on the Hybrid UNeXt, initiates a preliminary contour pair $(C_{0(i)}, C_{0(o)})$ representing the inner and outer boundaries for the central CT slice. This initial contour pair is determined by the boundary probability map P from BS and the offset map S. Subsequently, within the ICD module, the obtained contour pair $(C_{0(i)}, C_{0(o)})$ undergoes progressive deformation over K iterations, facilitated by the proposed multi-class circular convolution. Different from PolySnake, our method utilizes the proposed focal smooth L1 loss other than smooth L1 loss during training.

2.3. Focal Loss

To effectively tackle class imbalance in object detection tasks, the focal loss was introduced and substantiated through comprehensive experiments. In contrast to employing hard negative mining [21, 25] techniques that involve sampling challenging instances during training or employing intricate sampling/reweighing schemes, focal loss reformulates the loss function to ignore easy examples. This adjustment enables a focused training approach to challenging negatives.

More formally, focal loss adds a modulating factor $(1 - p_t)^\gamma$ to the cross-entropy loss, with a tunable focusing parameter $\gamma \geq 0$. focal loss is defined as

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t) \tag{1}$$

It's crucial to understand two key aspects of focal loss. Firstly, when an example is misclassified and the probability p_t is small, the modulating factor approaches 1, leaving the loss unaffected. Conversely, as p_t approaches 1, the factor diminishes to 0, resulting in the down-weighting of the loss for well-classified examples. Secondly, the focusing parameter γ provides a smooth adjustment to the rate at which easy examples are down-weighted. Specifically, when $\gamma = 0$, the focal loss aligns with cross-entropy (CE), and as γ increases, the impact of the modulating factor is proportionally heightened.

3. Method

We present an overview of the Physiology-aware PolySnake framework in Figure 1. To enhance clarity, we break down the framework into three distinct modules: (a) boundary segmentation, (b) initial contour generation, and (c) iterative contour deformation.

To tailor the framework for medical segmentation, we design the Hybrid UNeXt network for boundary segmentation and initial contour generation. This design takes into account the 3D spatial consistency in adjacent CT slices, utilizing a volume V containing $N + 1$ consecutive slices centered at I_t. Each slice is represented as $I_{t-N/2}...I_t...I_{t+N/2}$. In consideration of the physiological relation between inner and outer boundaries, we further optimize the iterative contour deformation in PolySnake for boundary segmentation using multi-class circular convolution. Our Physiology-aware PolySnake iteratively refines contour pairs to progressively align with the inner and outer boundaries. We adopt PolySnake's recurrent architecture, allowing the estimate of the contour pairs to be progres-

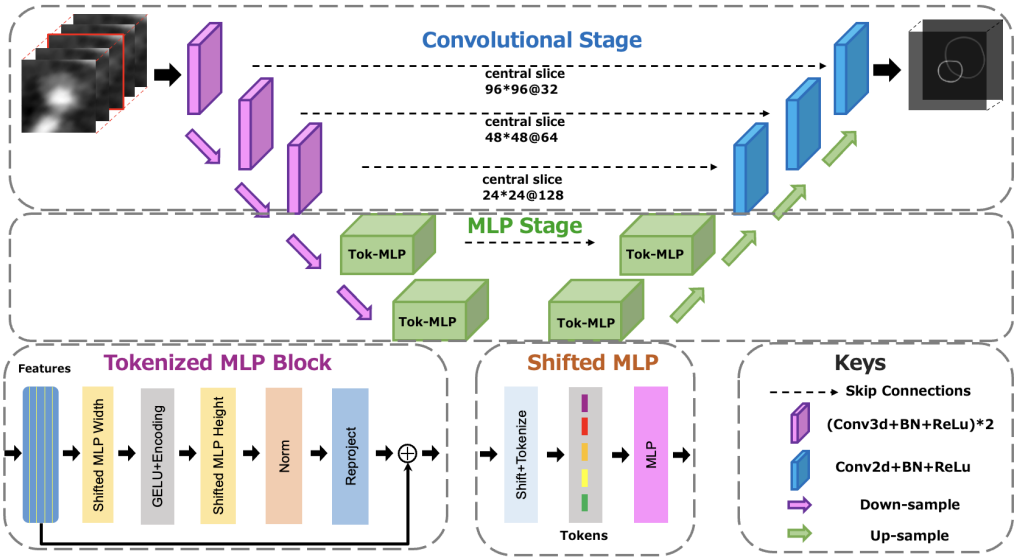

Figure 2. Overview for Hybrid UNeXt. It consists of two stages: Convolutional Stage and MLP Stage. Different from UNeXt architecture, we adopt a 3d-encoder-2d-decoder structure in Convolutional Stage, leveraging the spatial continuity of consecutive slices to detect the boundary of the target slice.

sively updated at each iteration. Through iterative refinements, the contour pairs gradually encompass the inner and outer boundaries, eventually converging to a stable state.

To address the challenge of weak performance in imbalanced classes, particularly in plaque-containing samples, we introduce the focal smooth L1 loss inspired by Focal Loss [13]. Our proposed loss serves two crucial roles during training: (a) it directs the model's attention to challenging negatives by down-weighting well-performed regions, and (b) it explicitly constrains the model's capacity in initial contour generation, thereby enhancing the model's robustness for iterative contour deformation.

3.1. Boundary Segmentation

The boundary segmentation module divides the inner and outer boundaries of the coronary artery wall. Note that while the boundary segmentation doesn't have a direct impact on the final prediction during the inference stage, it plays a crucial role in extracting finer features that enhance the perception of object boundaries.

3.1.1 Hybrid UNeXt

To capture the continuous spatial relationships between consecutive images, we crafted the Hybrid UNeXt network. Similar to UNeXt, this network adopts an encoder-decoder architecture featuring a convolutional stage and a tokenized MLP stage. As illustrated in Figure 2, the convolutional stage is constructed with a 3D encoder using residual blocks and a 2D decoder, while the tokenized MLP stage incorporates a tokenized MLP block-based encoder and decoder.

For a given input volume V, the encoder in the convolutional stage processes it initially with three consecutive 3D residual blocks. These blocks consist of two stages: batch normalization, ReLU activation, and 3D convolution, as shown in Figure 2. To enable hybrid processing with both 3D and 2D, we pass the central feature slice from the 3D encoder, representing the features for the target slice, to the next stage or the corresponding 2D decoder. In the tokenized MLP stage, we follow the same structure as UNeXt [24], receiving the central 2D feature slice from the last 3D encoder in the convolutional stage.

Specifically, for the given input 3d CT volume $V \in R^{D\times H\times W}$ with depth D, width W and Height H, we first feed it to Hybrid UNeXt network and a 2d-convolutional head outputs the feature map $F_b \in R^{H\times W\times C}$, where $C = 3$ is the category number representing background, inner boundary, outer boundary.

3.2. Initial contour generation

The Initial Contour Generation (ICG) module focuses on creating the initial contour pairs for inner and outer boundaries. Similar to PolySnake, we adopt the approach from E2EC [26], directly regressing an ordered set of vertices as the initial contour. This is in contrast to the method employed by Deep Snake [17] and Dance [14], which initialize a contour from the detected bounding box with a rectangle in their pipelines.

The Hybrid UNeXt network introduces an additional

convolutional head for initial contour generation, producing a feature map $F_i \in R^{\frac{H}{R} \times \frac{W}{R} \times 2C_v N_v}$, where $C_v = 2$ represents the boundary category (inner and outer boundaries), and N_v denotes the vertex number to form a contour. With given center coordinate pairs $(x^i, y^i), (x^o, y^o)$, the contour pairs are obtained by adding the predicted offset pairs $\{(\Delta x_k^i, \Delta y_k^i), (\Delta x_k^o, \Delta y_k^o) | k = 1, 2, ..., N_v\}$ at $\{(x^i, y^i), (x^o, y^o)\}$ of S.

3.3. Iterative contour deformation

With the initial contour pairs for inner and outer boundaries, the Iterative Contour Deformation (ICD) module continuously refines a single estimation of contour pairs through iterations. This iterative process leads the initial contour pairs to eventually converge to a stable state, tightly encompassing the inner and outer boundaries. Our approach in the ICD module aligns with the strategy employed by PolySnake. However, we distinguish ourselves by utilizing multi-class circular convolution instead of 1D circular convolution for concurrent feature learning of inner and outer boundaries, considering the physiological characteristics of the artery wall.

3.4. Multi-class circular convolution

For object detection tasks in general datasets, Deep Snake and PolySnake adopt a "detect then segment" strategy to predict final contour offsets. This means the network first detects an object instance to initialize a contour, and then progressively deforms it using 1D circular convolution [17]. It's important to note that the network deforms contours solely based on their initial states, without considering class differences. This can lead to confusion, especially when object contours are concentric, resulting in unsatisfactory performance.

In this paper, we introduce multi-class circular convolution for simultaneous feature learning of inner and outer boundaries, as illustrated in Figure 3. In essence, the features of inner and outer contour pairs can be treated as a 2D discrete signal $f : Z \rightarrow R^{M \times D}$ where $M = 2$ in this case, and processed using standard convolution. To preserve the contour's topology, we extend f to be a periodic signal defined as:

$$(f_N)_i = \sum_{j=-\infty}^{\infty} (f_{i-jN}) \tag{2}$$

and propose to encode the periodic features by the multi-class circular convolution defined as:

$$(f_N * k)_i = \sum_{j=-r}^{r} \sum_{c=1}^{M} (f_{N,c})_{i+j} k_{j,c} \tag{3}$$

where $k \rightarrow R^{M \times D}$ represents a learnable kernel function, the operator $*$ denotes standard convolution, c refers to class index, and M equals 2 in this paper. Similar to standard convolution, we design a network layer using multi-class circular convolution for feature learning, making it easy to integrate into modern network architectures.

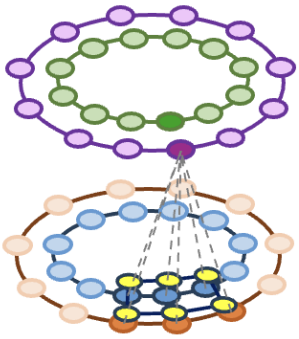

Figure 3. Multi-class Circular Convolution: The orange and blue nodes are the input features defined on contours on two different class(inner boundary and outer boundary), the yellow nodes represent the kernel function, and the purple and green nodes are the output features. The highlighted green and purple nodes are the inner product between the kernel function and the highlighted orange and blue nodes, which is the same as the standard convolution. The output features of circular convolution have the same length as the input features.

Throughout all experiments, the kernel size of the multi-class circular convolution is consistently fixed at 2×7.

3.5. Training Objectives

For object detection tasks in general datasets, both Deep Snake and PolySnake train the model using smooth L1 loss [7], which doesn't explicitly address the issue of class imbalance. To enhance the model's performance in imbalanced samples, we introduce focal smooth L1 loss as a replacement for smooth L1 loss during training, drawing inspiration from Focal Loss [13].

3.5.1 Focal Smooth L1 Loss

The focal smooth L1 loss is crafted specifically for scenarios involving contour offset prediction, particularly in cases where there's a substantial imbalance between healthy and unhealthy CT samples. Taking inspiration from Focal Loss, we aim to reshape the loss function, giving less weight to easy examples and emphasizing training on challenging negatives. Note that, during the initial contour generation stage, our proposed focal smooth L1 loss can also control the accuracy of offset predictions by adjusting the temperature parameter τ. Similar to Focal Loss, we introduce a modulating factor $\tanh(\frac{\mathcal{L}_{sml1}(p^n, \hat{p}^n)}{\tau})^\gamma$ into the smooth L1 loss, where γ is a tunable focusing parameter ($\gamma \geq 0$) and τ is a positive temperature parameter.

$$\mathcal{L}_F(P,\hat{P}) = \sum_{n=1}^{N} \tanh(\frac{\mathcal{L}_{sml1}(p^n,\hat{p}^n)}{\tau})^{\gamma}\mathcal{L}_{sml1}(p^n,\hat{p}^n) \quad p^n \in P, \hat{p}^n \in \hat{P} \tag{4}$$

We observe two characteristics of the focal smooth L1 loss. (a) When τ is small, 0.1 in our experiment, the focal smooth L1 loss is well-suited for accurate contour prediction. The factor approaches 0 when the smooth L1 loss is sufficiently small, down-weighting the loss for well-predicted coordinates. (b) When τ is large, 2 in our experiment, the focal smooth L1 loss is well-suited for initial contour generation. The factor approaches 0 when the smooth L1 loss reaches a rough range, down-weighting the loss for coordinates that aren't accurate enough compared to the ground truth. Consequently, the accuracy of the initial contours is constrained, enhancing the model's robustness for iterative contour deformation.

The training is divided into two stages. Concretely, the ICG and ICD modules are optimized with the objective as follows,

$$\mathcal{L} = \mathcal{L}_{ICG} + \mathcal{L}_{ICD} \tag{5}$$

(a) The $\mathcal{L}_{ICG}$ consists of two terms as follows

$$\mathcal{L}_{ICG} = \mathcal{L}_B + \mathcal{L}_S \tag{6}$$

where $\mathcal{L}_S$ is the loss on offset map S and $\mathcal{L}_B$ is the loss on boundary probability map. The offset map is shaped on $\frac{H}{R} \times \frac{W}{R} \times 2C_vN_v$ to directly regress the initial contour pairs of inner and outer boundaries with $2C_vN_v$ vertices. The boundary probability map is shaped on $H \times W \times C$, where C = 3 is the category number representing background, inner boundary, outer boundary. We use weighted hausdorff distance loss [18] for boundary segmentation

(b) The computation of ICD involves accumulating the loss over all K iterations, outlined as follows,

$$\mathcal{L}_{ICD} = \sum_{k=1}^{K} \lambda^{K-k}(\mathcal{L}_R^{(k)} + \alpha\mathcal{L}_P^{(k)}) \tag{7}$$

where $\mathcal{L}_R^{(k)}$and $\alpha\mathcal{L}_P^{(k)}$ are the coordinate regression loss and our proposed contour shape loss at the k th iteration, λ is the temporal weighting factor and α is the weight of contour shape loss. Here λ is less than 1, which means the weight of the loss increases exponentially with the iteration. We follow PolySnake to use coordinate regression loss $\mathcal{L}_R^{(k)}$ and the shape loss $\alpha\mathcal{L}_P^{(k)}$. Note that we replace all smooth L1 losses with our focal smooth L1 losses using different temperatures during training.

4. Experiments

4.1. Datasets & Metric

We assess and compare our methodology using a proprietary CCTA-CPR dataset, encompassing volumetric scans from 50 individual patients. Within each patient's dataset, three distinct artery volumes were examined: the right coronary artery (RCA), the left anterior descending (LAD) coronary artery, and the left circumflex (LCX) coronary artery. These CCTA scans were conducted using a 256-slice MDCT scanner (Brilliance iCT, Philips Medical Systems), featuring 128 mm × 0.625 mm collimation, a gantry rotation time of 270 ms, and an effective tube current ranging from 320 to 840 mA, with adjustments for obese patients.

During image acquisition, cardiac images covered the split from the pulmonary artery to the heart's apex. This took place at the 75th cardiac phase, with adjustments made for heart rates exceeding 75 beats per minute. The acquisition phase was set at 45 of one RR interval on the ECG. The reconstructed images achieved an in-plane resolution between 0.38 and 0.56 millimeters. The slice thickness remained constant at 0.9 millimeters, with an increment of 0.45 millimeters.Following this, the Cardiac Analysis software was employed for multi-planar reformation of coronary arteries, ultimately generating the CPR (Curved Planar Reformation) views.

For the development of our model, we partitioned the dataset randomly into training, validation, and test sets. Specifically, 35 patients (equivalent to 33,676 slices) were allocated for training, 5 patients (constituting 3,231 slices) for validation, and the remaining 10 patients (comprising 11,515 slices) for the final testing phase. To ensure uniformity and enhance computational efficiency, all slices underwent consistent pre-processing and augmentation.

This pre-processing included the conversion of Hounsfield Units (HU) to floating-point values within the range of 0 to 255, random central cropping with dimensions between 192 and 256 to eliminate extraneous information while preserving all artery voxels, image resizing to a dimension of 96 × 96, random rotations in 90-degree increments, and random horizontal and vertical flips. Note that cardiac experts provided corresponding annotation of segmentation masks for each slice, labeling from 0 to 4, representing the background, lumen regions of arteries, artery wall regions, calcified plaque, and non-calcified plaque in order. Therefore, we preprocess the ground-truth label for outer and inner boundaries by extracting the contour of the lumen region and the exterior contour of the artery wall region as the final ground-truth label for offset prediction.

We present the quantitative results of the proposed framework using the Dice Similarity Coefficient (Dice), and Hausdorff Distance (HD).

Table 1. Quantitative comparison in Hausdorff Distance. Methods are named in A+B format, where A denotes the backbone network, and B signifies the pipeline for contour deformation. Specifically, "HConvLstm" corresponds to the hierarchical ConvLSTM [12]. Furthermore, "F-Poly-Snake" represents PolySnake utilizing focal smooth L1 loss, while "Fmc-PolySnake" signifies PolySnake using both focal smooth L1 loss and multi-class circular convolution.

Method	Hdf^{h}_{ave}	Hdf^{cal}_{ave}	$\text{Hdf}^{noncal}_{ave}$	Avg_{macro}
Hybrid ResUNet +Deep snake	4.397	6.198	7.697	6.097
UNeXt + Deep snake	4.450	6.550	7.873	6.291
TransUNet + Deep snake	4.348	6.243	8.001	6.197
HConvLstm + Deep snake	4.400	6.144	7.339	5.961
Hybrid UNeXt + Deep snake	4.158	6.184	7.405	5.916
Hybrid UNeXt + DANCE	4.107	6.202	7.330	5.880
Hybrid UNeXt + E2EC	4.255	6.240	7.313	5.936
Hybrid-UNeXt + Poly-Snake	4.230	6.269	7.287	5.928
Hybrid-UNeXt + mc-Poly-Snake	4.243	6.235	7.090	5.856
Hybrid-UNeXt + F-Poly-Snake	4.081	6.027	6.793	5.633
Hybrid-UNeXt + Fmc-Poly-Snake	**4.081**	**5.903**	**6.590**	**5.524**

Table 2. Quantitative comparison in Dice.

Method	Dice^{h}_{ave}	Dice^{cal}_{ave}	$\text{Dice}^{noncal}_{ave}$	Avg_{macro}
Hybrid ResUNet +Deep snake	86.8	72.0	66.8	75.2
UNeXt + Deep snake	85.9	71.8	67.0	74.9
TransUNet + Deep snake	86.2	72.6	66.4	75.1
HConvLstm + Deep snake	87.5	72.3	66.7	75.5
Hybrid UNeXt + Deep snake	88.9	71.9	67.5	76.1
Hybrid UNeXt + DANCE	89.0	72.3	66.8	76.0
Hybrid UNeXt + E2EC	88.3	71.7	67.3	75.7
Hybrid-UNeXt + Poly-Snake	88.6	72.1	67.1	75.9
Hybrid-UNeXt + mc-Poly-Snake	88.4	72.7	67.9	76.3
Hybrid-UNeXt + F-Poly-Snake	89.2	73.3	68.4	76.9
Hybrid-UNeXt + Fmc-Poly-Snake	**89.6**	**74.5**	**68.8**	**77.6**

4.2. Implementation Detail

In the training phase, we utilized an Adam optimizer with a weight decay of 0.0005 and an initial learning rate set at 0.001. The learning rate underwent decay by a factor of 0.9 after every 10 epochs. This training protocol remained consistent for both our proposed model and comparable approaches. The entire training process spanned 50 epochs, employing a batch size of 96 to facilitate model convergence and enable robust performance evaluation.

4.3. Ablation Study

In this section, we perform ablations to validate the efficacy of key components within our proposed Physiology-aware pipeline. These components include boundary segmentation, initial contour generation, and iterative contour deformation. All ablations are conducted on our proprietary dataset comprising 50 individual patients.

4.3.1 Backbone Network

Initially, we assess the performance of the Hybrid UNeXt network for contour deformation. In Table 1, 2, we present a comparative analysis of the performance of the Hybrid UNeXt network serving as the backbone network, as opposed to various backbone networks used in the structures of Deep Snake. The overall macro-average score across all metrics demonstrates that our Hybrid UNeXt network enhances contour performance when employed as the backbone network.

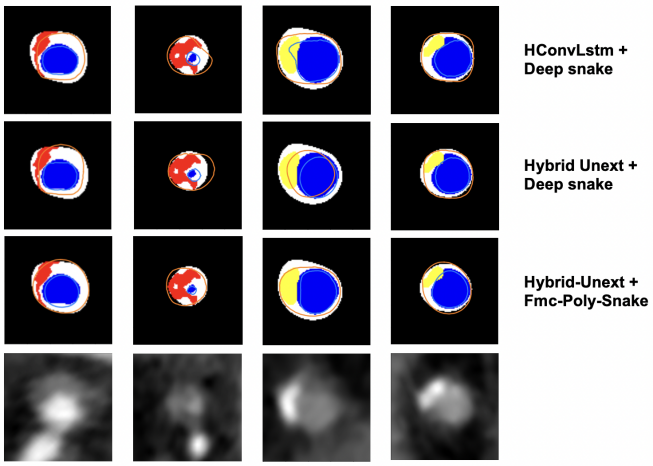

Figure 4. Quantitative Vision Comparison

4.3.2 Contour Deformation

In evaluating contour deformation, our focus lies on comparing the performance of Deep Snake, DANCE, E2EC, PolySnake, and our Physiology-aware PolySnake, all utilizing the same backbone network, as illustrated in Table 1, 2. For the Deep Snake network, we adhere to its approach of employing octagons based on extreme points as initial contours for deformation. In the case of PolySnake, the model is trained using smooth L1 loss, whereas our Physiology-aware PolySnake is trained with focal smooth L1 loss.

4.4. Comparisons with State-of-the-art Methods

To assess the effectiveness of our methodology, we conducted a comparative analysis against state-of-the-art boundary detection techniques leveraging deep learning. Noteworthy models included in our benchmark are hybrid ResUNet [9], Trans-UNet [4], UNeXt [24], Hierarchical ConvLSTM [12] for backbone network and Deep Snake [17], DANCE [14], E2EC [26], Polysnake [6] for contour deformation. The results of our evaluation encompass both quantitative metrics, detailed in Table 1, 2, and qualitative observations, presented in Figure 4.

4.5. Quantitative Visual Comparison

The results of three distinct methodologies are showcased in Figure 4. The first three rows present the outputs from various models: hierarchical convlstm with Deep Snake, hybrid UNeXt with Deep Snake, and our proposed Physiology-aware PolySnake, followed by the final row featuring test CCTA images in grayscale. The displayed images highlight the alignment between ground-truth segmentation regions and predicted boundary contours. In the visual representation, black corresponds to the ground-truth background, white signifies the ground-truth artery wall, blue indicates the ground-truth lumen, yellow represents the ground-truth calcified plaque, and red corresponds to the ground-truth non-calcified plaque. Additionally, the predicted exterior artery wall boundary is delineated by an orange contour, while the predicted interior artery wall boundary is outlined with a light blue contour.

Our proposed approach consistently outperforms competing methods by accurately depicting vessel wall shapes across diverse test samples with varying conditions. This superiority is evident not only in overall metrics but also in the trend of covering more plaque regions. These qualitative results serve to reinforce the efficacy of our method for coronary wall segmentation.

5. Conclusion

In this paper, we propose a novel Physiology-aware PolySnake method designed for coronary vessel wall segmentation, acknowledging the physiological intricacies of the coronary artery. Our contributions encompass three key aspects: (a) Hybrid UNeXt Network Design: We present a Hybrid UNeXt network tailored for artery vessel segmentation and initial contour prediction, taking into account the 3D spatial coherence in adjacent CT slices. (b) Multi-class Circular Convolution for Iterative Contour Deformation: We propose the integration of multi-class circular convolution within the iterative contour deformation module. This addition allows for the iterative refinement of initial contour pairs, considering the interplay between the inner and outer boundaries of the artery wall simultaneously. (c) Focal Smooth L1 Loss for Improved Model Robustness: Addressing class imbalance in samples with arterial plaque, we introduce a focal smooth L1 loss, drawing inspiration from Focal Loss. This loss function explicitly constrains the accuracy of initial contour predictions, enhancing the model's robustness.

Our comprehensive evaluation attests to the state-of-the-art performance of our approach when compared to existing methods. The effectiveness of our proposed components positions our methodology as a promising candidate for real-world clinical applications. This study contributes valuable insights into the development of physiology-aware networks for enhanced coronary vessel wall analysis.

Acknowledgments

This work was partially supported by Cross-ministerial Strategic Innovation Promotion Program (SIP) on "Integrated Health Care System" Grant Number JPJ012425JST, Moonshot R&D Grant Number JPMJPS2011, CREST Grant Number JPMJCR2015, Basic Research Grant (Super AI) of Institute for AI and Beyond of the University of Tokyo and JSPS KAKENHI Grant Number 23K16990.

References

[1] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M Taha, and Vijayan K Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*, 2018. 1

[2] Armin Arbab-Zadeh and Valentin Fuster. The myth of the "vulnerable plaque" transitioning from a focus on individual lesions to atherosclerotic disease burden for coronary artery disease risk assessment. *Journal of the American College of Cardiology*, 65(8):846–855, 2015. 1

[3] Davide Cao, Rishi Chandiramani, Davide Capodanno, Jeffrey S Berger, Matthew A Levin, Mary T Hawn, Dominick J Angiolillo, and Roxana Mehran. Non-cardiac surgery in patients with coronary artery disease: risk evaluation and periprocedural management. *Nature Reviews Cardiology*, 18(1):37–57, 2021. 1

[4] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021. 1, 8

[5] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014. 2

[6] Hao Feng, Wengang Zhou, Yufei Yin, Jiajun Deng, Qi Sun, and Houqiang Li. Recurrent contour-based instance segmentation with progressive learning. *arXiv preprint arXiv:2301.08898*, 2023. 2, 8

[7] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015. 5

[8] Dongdong Hao, Song Ding, Linwei Qiu, Yisong Lv, Baowei Fei, Yueqi Zhu, and Binjie Qin. Sequential vessel segmentation via deep channel attention network. *Neural Networks*, 128:172–187, 2020. 2

[9] Kaikai Huang, Antonio Tejero-de Pablos, Hiroaki Yamane, Yusuke Kurose, Junichi Iho, Youji Tokunaga, Makoto Horie, Keisuke Nishizawa, Yusaku Hayashi, Yasushi Koyama, et al. Coronary wall segmentation in ccta scans via a hybrid net with contours regularization. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1743–1747. IEEE, 2020. 1, 2, 8

[10] Simon Kohl, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R Ledsam, Klaus Maier-Hein, SM Eslami, Danilo Jimenez Rezende, and Olaf Ronneberger. A probabilistic u-net for segmentation of ambiguous images. *Advances in neural information processing systems*, 31, 2018. 1

[11] Justin Liang, Namdar Homayounfar, Wei-Chiu Ma, Yuwen Xiong, Rui Hu, and Raquel Urtasun. Polytransform: Deep polygon transformer for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9131–9140, 2020. 2

[12] Andrew Lin, Nipun Manral, Priscilla McElhinney, Aditya Killekar, Hidenari Matsumoto, Jacek Kwiecinski, Konrad Pieszko, Aryabod Razipour, Kajetan Grodecki, Caroline Park, et al. Deep learning-enabled coronary ct angiography for plaque and stenosis quantification and cardiac risk prediction: an international multicentre study. *The Lancet Digital Health*, 4(4):e256–e265, 2022. 1, 2, 7, 8

[13] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 2, 4, 5

[14] Zichen Liu, Jun Hao Liew, Xiangyu Chen, and Jiashi Feng. Dance: A deep attentive contour model for efficient instance segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 345–354, 2021. 2, 4, 8

[15] Riccardo Marano, Giuseppe Rovere, Giancarlo Savino, Francesco Ciriaco Flammia, Maria Rachele Pia Carafa, Lorenzo Steri, Biagio Merlino, and Luigi Natale. Ccta in the diagnosis of coronary artery disease. *La radiologia medica*, 125:1102–1113, 2020. 1

[16] Lei Mou, Yitian Zhao, Huazhu Fu, Yonghuai Liu, Jun Cheng, Yalin Zheng, Pan Su, Jianlong Yang, Li Chen, Alejandro F Frangi, et al. Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging. *Medical image analysis*, 67:101874, 2021. 2

[17] Sida Peng, Wen Jiang, Huaijin Pi, Xiuli Li, Hujun Bao, and Xiaowei Zhou. Deep snake for real-time instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8533–8542, 2020. 1, 2, 4, 5, 8

[18] Javier Ribera, David Guera, Yuhao Chen, and Edward J Delp. Locating objects without bounding boxes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6479–6489, 2019. 6

[19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 1

[20] Patrick W Serruys, Hironori Hara, Scot Garg, Hideyuki Kawashima, Bjarne L Nørgaard, Marc R Dweck, Jeroen J Bax, Juhani Knuuti, Koen Nieman, Jonathon A Leipsic, et al. Coronary computed tomographic angiography for complete assessment of coronary artery disease: Jacc state-of-the-art review. *Journal of the American College of Cardiology*, 78(7):713–736, 2021. 1

[21] Kah-Kay Sung. Learning and example selection for object and pattern detection. 1996. 3

[22] Mazhar B Tayel, MA Massoud, and Y Farouk. A modified segmentation method for determination of iv vessel boundaries. *Alexandria Engineering Journal*, 56(4):449–457, 2017. 1

[23] Mazhar B Tayel, Magdy A Massoud, and YF Shehata. An automatic segmentation for determination of iv vessel boundaries. *International Journal of Bioscience, Biochemistry and Bioinformatics*, 4(4):218, 2014. 1

[24] Jeya Maria Jose Valanarasu and Vishal M Patel. Unext: Mlp-based rapid medical image segmentation network.

In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 23–33. Springer, 2022. 1, 2, 4, 8

[25] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I. Ieee, 2001. 3

[26] Tao Zhang, Shiqing Wei, and Shunping Ji. E2ec: An end-to-end contour-based method for high-quality high-speed instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4443–4452, 2022. 2, 4, 8

[27] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018. 1

[28] Xiliang Zhu, Zhaoyun Cheng, Sheng Wang, Xianjie Chen, and Guoqing Lu. Coronary angiography image segmentation based on pspnet. *Computer Methods and Programs in Biomedicine*, 200:105897, 2021. 2