Article

A Synergistic Multi-scale Attention and Composite Feature Extraction Network for Coronary Artery Segmentation

Long Zhang [1,2,†],Yue Du[1,†], Yunlong Lin[2,†], Yiyuan Li[3],Zhenyu Cheng[1], Boyuan Zhang[1,*] and Shoujun Zhou [1,*]

[1] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Nanshan District, Shenzhen, 518055, China;
[2] Guilin University of Electronic Technology, 1 Jinji Road, Qixing District, Guilin, Guangxi, 541004, China;
[3] Weixian college, Tsinghua University, Haidian District, Beijing, 100084, China;
[*] Correspondence: 17852156693@163.com; sj.zhou@siat.ac.cn;
[†] These authors contributed equally to this work: Long Zhang (Co-first, ranked 1st), Yue Du (Co-first, ranked 2nd), Yunlong Lin (Co-first, ranked 3rd).

[‡] This work was supported by the Key-Area Research and Development Program of Guangdong Province (No. 2025B1111020001), in part by the Natural Science Foundation of Guangdong Province (No. 2023A1515010673), in part by the Shenzhen Science and Technology Innovation Bureau key project (No. SGG20220831110400001, No. CIGIZD20230724093303007, No. KZD20240903101259001), in part by Shenzhen Medical Research Fund (No. D2404001), in part by Shenzhen Engineering Laboratory for Diagnosis & Treatment Key Technologies of Interventional Surgical Robots (No. XMH'T20220104009), and the Key Laboratory of Biomedical Imaging Science and System, CAS, for the Research platform support.

Abstract

Accurate coronary artery segmentation from two-dimensional Digital Subtraction Angiography (DSA) images is paramount for robot-assisted percutaneous coronary intervention (PCI) but is severely challenged by complex background artifacts, the intricate morphology of fine vascular branches, and frequent segmentation discontinuities. These inherent difficulties often render conventional segmentation approaches inadequate for the stringent precision demands of surgical navigation. To address these limitations, we propose a novel deep learning framework incorporating a Composite Feature Extraction Module (CFEM) and a Multi-scale Composite Attention Module (MCAM) within a U-shaped architecture. The CFEM is meticulously designed to capture tubular vascular characteristics and adapt to diverse vessel scales, while the MCAM, strategically embedded in skip connections, synergistically integrates multi-scale convolutions, spatial attention, and lightweight channel attention to enhance the perception of fine branches and model long-range dependencies, thereby improving topological connectivity. Additionally, a combined Dice-Focal loss function is employed to jointly optimize segmentation boundary accuracy and mitigate class imbalance. Extensive experiments on the public ARCADE dataset demonstrate that our method significantly outperforms state-of-the-art approaches, achieving a Dice coefficient of 76.74%, a clDice of 50.30%, and an HD95 of 57.84 pixels. These quantitative improvements in segmentation accuracy, vascular connectivity, and edge precision underscore its substantial clinical potential for providing robust vascular structure information in robot-assisted interventional surgery.

Keywords: Coronary artery vessel segmentation; Deep learning; Multi-scale feature fusion; Attention mechanism; Digital subtraction angiography; Robot-assisted interventional surgery

Received:
Revised:
Accepted:
Published:

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Journal Not Specified* **2025**, *1*, 0. https://doi.org/

1. Introduction

Cardiovascular diseases (CVDs) remain the leading cause of global mortality, accounting for approximately 9.1 million deaths annually according to the World Health Organization's (WHO) Global Health Estimates [1,2]. This pervasive health crisis underscores the urgent need for advanced diagnostic and interventional strategies. Percutaneous coronary intervention (PCI), as the main treatment method for coronary heart disease, has been widely adopted in clinical practice.With the advancement of medical robotics, robot-assisted PCI procedures have emerged as a research hotspot in interventional therapy due to their high operational precision and reduced radiation exposure for physicians [3–7].

In robot-assisted PCI surgeries, accurate coronary artery segmentation results from two-dimensional DSA images is critical, as it serves as the core input for the surgical navigation system.Precise vessel segmentation not only facilitates the accurate planning of interventional instruments (such as guidewires and stents), helping to avoid interference from vascular branches, but also optimizes the imaging acquisition process, thereby reducing the use of contrast agents and minimizing radiation exposure for patients [8–12].However, the imaging characteristics of two-dimensional DSA images pose significant challenges for coronary artery segmentation; First, the background of the images contains artifacts from bone and soft tissues, which overlap with the grayscale intensity of blood vessels, often leading to false detections or missed regions.Second, coronary arteries exhibit a tree-like topology, with vessel diameters varying by more than an order of magnitude between main trunks and fine branches. Traditional segmentation methods struggle to consistently extract vessels across such scales.Third, intersections and curved regions of vessels are especially susceptible to noise and artifacts, resulting in segmentation breaks and compromised vascular connectivity.

In recent years, methods for segmenting coronary arteries have evolved from traditional threshold segmentation and morphological operations to deep learning-based segmentation techniques. U-Net and its variants, with their symmetric encoder-decoder architecture and skip connection design, have been widely applied in the field of medical image segmentation. However, these models exhibit limited adaptability to multi-scale vascular structures and struggle to accurately extract fine vessel branches [13]; AttUNet enhances the focus on vascular regions by introducing an attention mechanism into U-Net, but does not specifically optimize for the tubular morphology and branching structure of blood vessels[14–17];Although Transformer-based segmentation models are capable of capturing long-range dependencies, their high computational cost and large data requirements hinder their application in real-time clinical settings [18].

Despite these advancements, current coronary artery segmentation methodologies still contend with three critical limitations: (1) Existing feature extraction modules often inadequately capture the inherent tubular morphology of blood vessels, exhibiting limited generalization across varying vessel thicknesses. (2) The capacity for modeling long-range dependencies remains insufficient, impeding the preservation of vascular connectivity, particularly within complex branching structures. (3) Loss function designs frequently fail to concurrently optimize for segmentation boundary accuracy and effectively address class imbalance, thereby constraining overall segmentation performance.

In response to the aforementioned issues, this paper proposes a coronary artery segmentation method that integrates multi-scale features. The principal research contributions and innovations are summarized as follows:

1.A meticulously engineered Composite Feature Extraction Module (CFEM): By combining dynamic snake convolution with a dual-path scaling architecture, directional convolutions along the x-axis and y-axis are employed to effectively capture the tubular features of blood vessels. At the same time, the expansion-contraction convolution operation is utilized to expand the feature receptive field, thereby enhancing the model's adaptability to different scales of blood vessels.

2. A novel Multi-scale Composite Attention Module (MCAM): This module is designed to integrate multi-scale convolution, spatial attention, and a lightweight channel attention mechanism. This collective mechanism not only enhances the perception of intricate vascular branch structures with minimal computational overhead but also facilitates robust modeling of long-range dependencies, consequently improving the topological connectivity of segmentation results.

3. Strategic application of a combined Dice-Focal loss function: This loss function is employed to concurrently optimize the overlap measure of segmented regions (via Dice loss) and to effectively mitigate the influence of easily classified samples (via Focal loss). This dual-component strategy is specifically designed to strengthen the learning process for challenging instances, such as fine branches and complex vascular intersection areas, thereby yielding superior segmentation accuracy.

4.Systematic Validation and Empirical Efficacy: Comparative and ablation experiments, rigorously executed on the publicly accessible ARCADE dataset, unequivocally substantiate the profound impact of the newly introduced modules and techniques. These investigations demonstrably reveal a statistically significant enhancement in segmentation accuracy, topological connectivity, and delineative edge precision, thereby furnishing robust empirical evidence for their prospective utility in diverse clinical applications.

2. Related works

2.1. Analysis of Morphological Characteristics of Coronary Vessels

The morphological characteristics of coronary vessels are the core basis for designing segmentation methods. They mainly include two types: tubular features and branching features, and their characteristics directly influence the design direction of the segmentation methods.

2.1.1. Cylindrical feature

The coronary vessels (especially the aortic segment) present a distinct tubular morphology characterized by thinness, length, continuity, and directional consistency (as shown in the yellow box in Figure 1).The tubular morphology of coronary vessels has the following differences: (1) The diameter of blood vessels shows a uniform change trend with the increase in branching levels. The diameter of the main trunk can reach 2-3 mm, while that of the terminal branches is only 0.2-0.3 mm. (2) The gray value of the blood vessel wall shows a gradient change, which is prone to be confused with background artifacts.The above characteristics require the segmentation model to have strong capabilities in extracting tubular features and adapting to different scales. It must accurately capture the main vessels while avoiding missing the fine branches.

Coronary vessels have a tree-like topological structure. The distal end of the aorta can be divided into the left anterior descending branch, the left circumflex branch, the right coronary artery, etc. as the main branches, and each branch is further divided into multiple levels of fine branches (as shown in Figure 1, the blue box).The projection characteristics of two-dimensional DSA images cause the superimposition of different depths of vascular branches, which alters the uniform change trend of vascular diameters and increases the difficulty of segmentation [19–23]. Moreover, areas with vascular intersections and bends are prone to sudden drops in gray values, making it difficult for traditional models to maintain segmentation connectivity. Therefore, a specially designed long-distance dependent modeling module is required to restore the dendritic structure of the blood vessels.

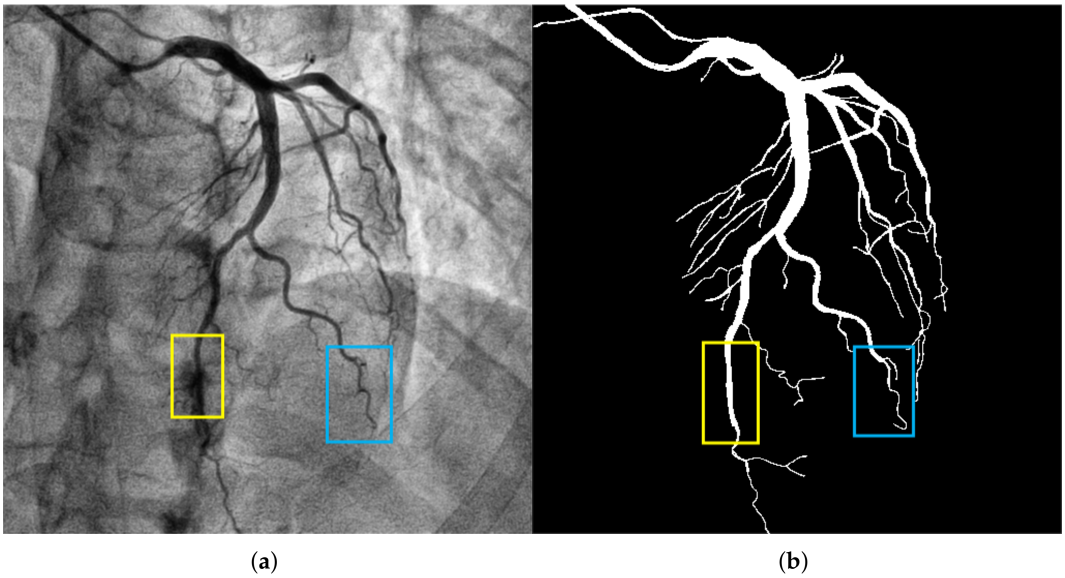


Figure 1. (**a**) Coronary vascular DSA image; (**b**) Coronary vascular mask.

2.2. Key Issues in Coronary Vessel Segmentation

2.2.1. Accuracy Issue

The accuracy issue mainly manifests as false detections and missed detections in pixel-level segmentation, making it difficult to restore the true anatomical shape of coronary vessels.The main reasons include:(1) The gray values of blood vessels are close to those of the background tissues (such as myocardium and veins), resulting in low contrast; (2) The proportion of pixels of small branches (diameter < 0.5mm) is low, and they are easily ignored by the model; (3) The gray values of the lesion areas (such as stenosis and calcification) are abnormal, causing blurred segmentation boundaries.As shown in Figure 2,There are obvious mis-segmentation (the background area is identified as a blood vessel) and missed-segmentation (the blood vessel area is not identified) phenomena in the segmentation results, which seriously affect the reliability of surgical navigation.

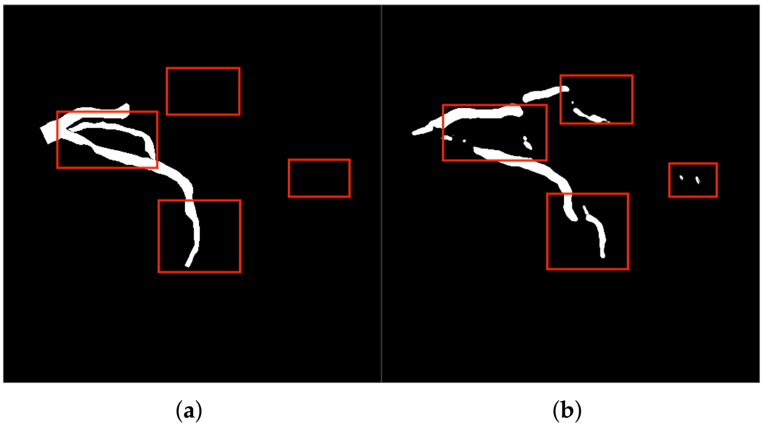
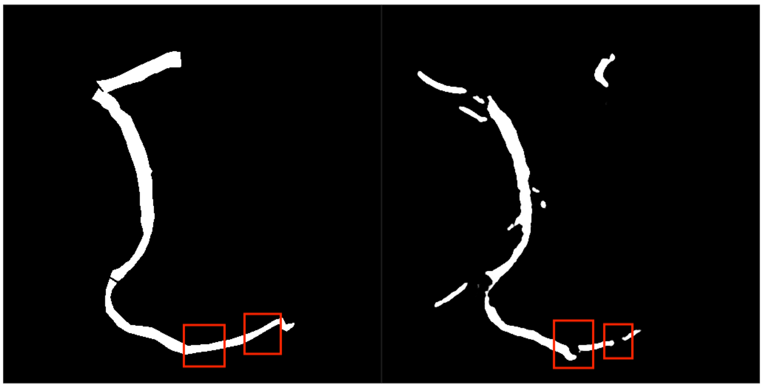


Figure 2. Figure 2. Accuracy issues in coronary vascular segmentation: (**a**) Label; (**b**) Segmentation result.

2.2.2. Connectivity Issues

The connectivity problem manifests as the breakage or incorrect connection between the main vessels and the branch vessels in the segmentation results, thereby disrupting the topological structure of the vascular network (Figure 3).The main reasons include:(1)

The receptive field of traditional convolutional networks is limited, making it difficult to capture long-distance dependencies across branches; (2) The vascular crossing areas are affected by projection superposition, resulting in complex gray value distribution, which leads to model misjudgment of vascular continuity; (3) Noise and motion artifacts interfere, causing discontinuous segmentation of fine branches. The connectivity problem can lead to misjudgment of vascular paths by the surgical navigation system, increasing surgical risks. Therefore, it is a problem that the segmentation method needs to focus on solving.



(**a**) (**b**)

Figure 3. Connectivity issues in coronary vascular segmentation: (**a**) Label; (**b**) Segmentation result.

2.3. Classification of Existing Partitioning Methods

Depending on the different technical routes, the existing methods for segmenting coronary vessels can be classified into three categories:

1. Traditional segmentation methods: The traditional methods mainly rely on low-level features such as image gray values and textures for segmentation, including threshold segmentation, edge detection and morphological operations. For instance, Hassouna et al [24] achieved brain vessel segmentation based on a random model, but it was sensitive to noise; Sun et al[25–27] extracted vascular trees by combining morphological multi-scale enhancement with the watershed algorithm, but it was difficult to handle complex background artifacts.

2. Segmentation method based on deep learning: Based on deep learning methods, which possess the ability of end-to-end learning, they have become the mainstream technology for coronary artery segmentation at present. U-Net [28], as the benchmark model for medical image segmentation, integrates high and low-level features through skip connections, but its adaptability to multi-scale blood vessels is insufficient. MedNeXt [29] adopts a scaling architecture to expand the receptive field, enhancing its ability to segment complex structures, but it has not been optimized for vascular tubular features. DSCNet [30] combines dynamic snake convolution to capture tubular features and has a relatively high segmentation accuracy. However, it lacks the ability to model long-distance dependencies and is prone to connectivity issues.

3. Segmentation method based on Transformer: The Transformer architecture achieves long-range dependency modeling through the self-attention mechanism, providing a new approach to solving the vascular connectivity problem. TransUNet [31] combines the Transformer and U-Net architectures, enhancing its ability to perceive the overall structure. Swin-Unet [32] reduces computational costs through hierarchical window attention, but still has problems such as high data requirements and slow inference speed, making it difficult to meet the real-time navigation requirements in clinical settings.

3. Methods

3.1. Overall network architecture

The proposed coronary artery segmentation network is architected upon a U-shaped symmetrical encoder-decoder framework, fundamentally enhanced by the integration of our novel Composite Feature Extraction Module (CFEM) and Multi-scale Composite Attention Module (MCAM) (Figure 4). This architecture is characterized by:

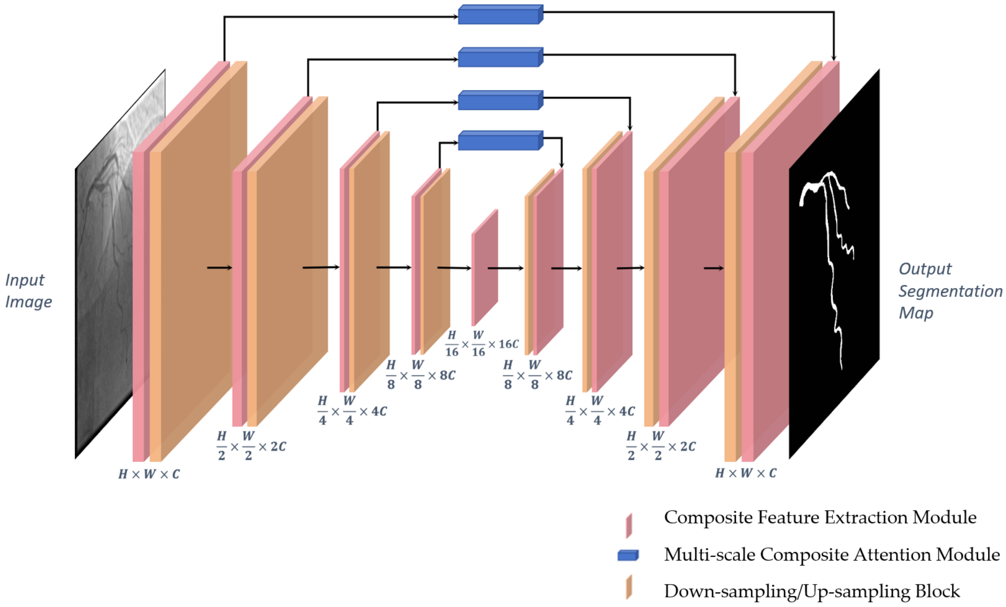

Input Image
$H \times W \times C$
$\frac{H}{2} \times \frac{W}{2} \times 2C$
$\frac{H}{4} \times \frac{W}{4} \times 4C$
$\frac{H}{8} \times \frac{W}{8} \times 8C$
$\frac{H}{16} \times \frac{W}{16} \times 16C$
$\frac{H}{8} \times \frac{W}{8} \times 8C$
$\frac{H}{4} \times \frac{W}{4} \times 4C$
$\frac{H}{2} \times \frac{W}{2} \times 2C$
$H \times W \times C$
Output Segmentation Map
Composite Feature Extraction Module
Multi-scale Composite Attention Module
Down-sampling/Up-sampling Block

Figure 4. Coronary vascular segmentation network.

.Encoder: Comprising four sequential stages, each consisting of a CFEM module followed by a downsampling block. Each downsampling block, implemented via a convolutional operation with a stride of 2, progressively reduces the feature map resolution by half while doubling the channel count, thereby expanding the receptive field and extracting hierarchical global features.

.Decoder: Structured with four corresponding stages, each featuring an upsampling block followed by a CFEM module. Upsampling is achieved through transposed convolution, which restores the feature map resolution. Crucially, multi-scale features from the encoder are integrated into the decoder via skip connections, where the MCAM module plays a pivotal role in refining these features before their fusion with the upsampled features.

.Bottleneck: A single CFEM module is positioned at the network's bottleneck, responsible for extracting the highest-level global features and providing essential contextual information to the decoder.

.Output Layer: The final segmentation probability map is generated by a 1×1 convolutional layer, which maps the feature channels to two (representing foreground vessel and background), followed by a sigmoid activation function.

The input and output dimensions of each module in the network follow the following rules: (1) The downsampling block reduces the feature map size by 1/2 and doubles the number of channels; (2) The upsampling block expands the feature map size by 2 times and halves the number of channels; (3) The CFEM and MCAM modules maintain the feature map size and channel number unchanged to ensure the compatibility of feature fusion between the modules.

3.2. Composite Feature Extraction Module (CFEM)

The CFEM module is designed based on the characteristics of vascular tube shapes and scale variations. It consists of a tube-shaped feature extraction layer and a dual-path scaling architecture , enabling precise extraction of vascular features and scale generalization[36-38].

3.2.1. Cylindrical Feature Extraction Layer

The tubular feature extraction layer captures the tubular morphological characteristics of blood vessels in a directional manner through dynamic snake convolution in the x-axis and y-axis directions. The specific steps are as follows:

1.Group Normalization: Apply Group Normalization (GN) to the input feature maps to reduce the impact of Batch Size on training and stabilize the training process.

2.Directional convolution: The normalized feature maps are respectively input into the dynamic snake-shaped convolutions along the x-axis (9×1 convolution kernel) and y-axis (1×9 convolution kernel), resulting in two directional tubular feature maps.

3.Feature Integration: Concatenate Fx and Fy along the channel dimension, and integrate the features from both directions through a 3×3 ordinary convolution to achieve comprehensive capture of tubular features.

4.Activation function: Perform group normalization on the integrated feature map, combine with the SiLU activation function to enhance the non-linear expression ability of the features, and output tubular feature maps.

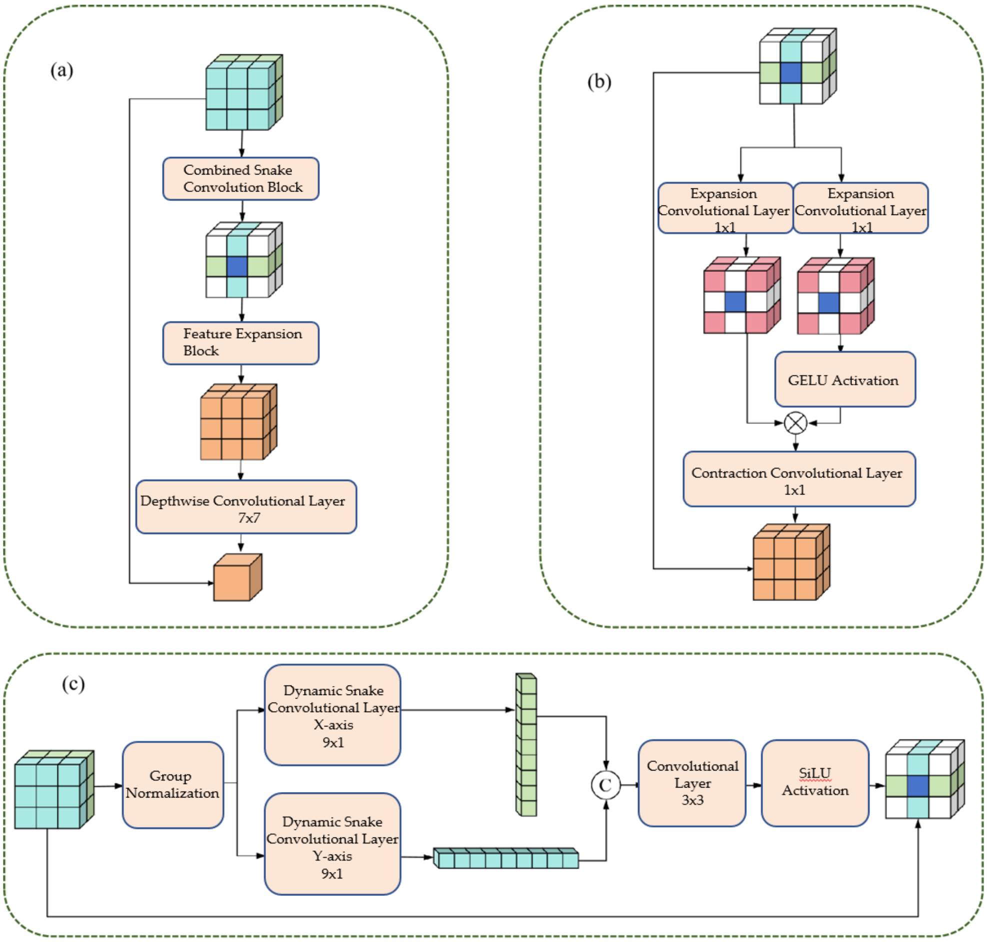

(a)
Combined Snake Convolution Block
Feature Expansion Block
Depthwise Convolutional Layer 7x7
(b)
Expansion Convolutional Layer 1x1
Expansion Convolutional Layer 1x1
GELU Activation
Contraction Convolutional Layer 1x1
(c)
Group Normalization
Dynamic Snake Convolutional Layer X-axis 9x1
Dynamic Snake Convolutional Layer Y-axis 9x1
C
Convolutional Layer 3x3
SiLU Activation

Figure 5. Composite Feature Extraction Module (CFEM).

The above process can be expressed by Formula (1).

$$\begin{aligned} F_x &= \mathrm{DSConv}_x(\mathrm{GN}(F_{\mathrm{in}})) \\ F_y &= \mathrm{DSConv}_y(\mathrm{GN}(F_{\mathrm{in}})) \\ F_{\mathrm{concat}} &= \mathrm{Concat}(F_x, F_y) \\ F_{\mathrm{tube}} &= \mathrm{SiLU}(\mathrm{GN}(\mathrm{Conv}_{3\times 3}(F_{\mathrm{concat}}))) \end{aligned} \tag{1}$$

3.2.2. Dual-path scaling architecture

The dual-path scaling architecture expands and contracts convolution operations to expand the feature receptive field, enhancing the model's adaptability to different scales of blood vessels. The specific steps are as follows:

1.Feature routing: The tubular feature map (Fconcat) is divided into two paths, which are respectively input into the expansion convolution (Exp) layer;

2.Extended convolution: Utilizing a 1×1 convolution kernel with a 4-fold expansion rate, the feature receptive field is expanded by a factor of 4, enabling the capture of large-scale vascular features.

3.Feature Interaction: Apply the GELU activation function to the output of one path, and perform matrix multiplication with the output of another path to achieve feature interaction and scale adaptation.

4.Contraction convolution: Utilizing a 1×1 convolution kernel with a 4-fold contraction rate, the receptive field is restored to its original size, ensuring that the output feature map is consistent with the input size.

5.Deep convolution: By using 7×7 deep convolution to integrate features, the number of parameters is reduced and the computational efficiency is improved;

6.Residual Connection: Before the output of the module, a residual connection is introduced, where the input feature map (Fin) is added to the processed feature map, enhancing the robustness of the network and preventing the vanishing gradient.

The above process can be expressed by Formula 2:

$$\begin{aligned} F_{\mathrm{expl1}} &= \exp(F_{\mathrm{tube}}) \\ F_{\mathrm{exp2}} &= \exp(F_{\mathrm{tube}}) \\ F_{\mathrm{inter}} &= F_{\mathrm{exp1}} \times \mathrm{GELU}(F_{\mathrm{exp2}}) \\ F_{\mathrm{con}} &= \mathrm{Com}(F_{\mathrm{inter}}) \\ F_{\mathrm{depth}} &= \mathrm{DepthConv}_{7\times 7}(F_{\mathrm{con}}) \\ F_{\mathrm{out}} &= F_{\mathrm{depth}} + F_{\mathrm{in}} \end{aligned} \tag{2}$$

3.2.3. Multi-scale Composite Attention Module (MCAM)

The Multi-scale Composite Attention Module (MCAM) is strategically embedded within the skip connection path of the U-shaped network. Its primary function is to refine the features transmitted from the encoder to the decoder by leveraging a synergistic combination of multi-scale feature processing, spatial attention, and lightweight channel attention. Upon receiving the feature map from the skip connection, the MCAM module first performs multi-scale feature extraction internally. This process involves: 1. Channel Compression: Initially, a 1×1 convolution is applied to the input feature map to reduce its channel dimension to 1/4, thereby optimizing computational efficiency. 2. Multi-scale Convolution: The compressed feature map is then branched into multiple parallel paths. Specifically, three distinct paths are processed by convolutional layers employing 1×1, 3×3, and 5×5 kernels, respectively, enabling the extraction of features at varying receptive field scales. 3. Feature Fusion: The outputs from these multi-scale convolutional paths are subsequently concatenated along the channel dimension. This aggregated multi-scale feature representation is then further integrated through group normalization and another 1×1 convolution, yielding a rich, fused feature map (Fms) that captures diverse spatial contexts. This Fms then serves as the input for subsequent spatial and channel attention mechanisms, enhancing the perception of vascular branch structures and ultimately improving segmentation connectivity.

Spatial attention enhances the spatial weights of vascular regions and suppresses background artifacts interference. The specific steps are as follows:1.Pooling operation: Perform average pooling and maximum pooling on the fused feature maps (F_{ms}) respectively, obtaining two spatial mapping images F_{avg} and F_{max}. 2.Feature concatenation: Concatenate F_{avg} and F_{max} along the channel dimension to obtain a 2-channel feature map. 3.Convolutional activation: Through a 7×7 convolutional layer, spatially dependent features are extracted. Combined with the SiLU activation function and the sigmoid function, a spatial attention map M_s is generated. 4.Feature weighting: Multiply the spatial attention map M_s element-wise with the fused feature map F_{ms} to obtain the spatially weighted feature map F_s.

The above process can be expressed by formula 3:

$$\begin{aligned}
F_{\text{avg}} &= \text{AvgPool}(F_{ms}) \\
F_{\text{max}} &= \text{MaxPool}(F_{ms}) \\
F_{\text{cat}} &= \text{Concat}(F_{\text{avg}}, F_{\text{max}}) \\
M_s &= \text{Sigmoid}(\text{SiLU}(\text{Conv}_{7\times 7}(F_{\text{cat}}))) \\
F_s &= F_{ms} \square M_s
\end{aligned} \tag{3}$$

The channel attention adopts the lightweight SimAM module. By calculating the energy function of pixels within the channel, it generates channel weights, enhancing the key feature channels. The specific steps are as follows: 1. Statistical calculation: For each channel of the feature map input to the MCAM module, calculate the pixel mean and variance;2. Energy function calculation: Based on the mean and variance, the energy value Ec for each channel is calculated using formula (4). The smaller the energy value, the higher the importance of that channel for segmentation.

$$E_c = \frac{(\mu_c - t)^2}{\sigma_c^2 + \epsilon} \tag{4}$$

Weight Generation: The energy values are mapped to channel weights M_c through the sigmoid function, with the weight range being [0, 1]. Feature weighting: Multiply the channel weights (M_c) with the input feature map (F_{in}) channel by channel to obtain the channel-weighted feature map (F_c).

The above process can be expressed by formula 5:

$$\begin{aligned}
M_c &= \text{Sigmoid}(-E_c) \\
F_c &= F_{\text{in}} \odot M_c
\end{aligned} \tag{5}$$

4. Experiments and Result Analysis

4.1. Dataset Introduction

The experiment utilized the ARCADE public dataset [36], a comprehensive collection of 1500 pairs of coronary artery Digital Subtraction Angiography (DSA) images meticulously labeled by medical experts [31]. Each image, originally sized at 512×512 pixels, encompasses diverse patient cases and various types of vascular lesions, rendering it highly suitable for the intricate task of coronary artery segmentation. To rigorously assess the method's generalization capability, the dataset was systematically partitioned into a training set, validation set, and test set, adhering to a ratio of 1000:200:300, respectively. Furthermore, to augment data diversity and bolster the model's robustness against variations, the training set underwent a three-stage image enhancement protocol, incorporating brightness and contrast adjustments, CLAHE equalization, and a suite of geometric transformations including rotation, flipping, blurring, and affine transformations.

4.2. Experimental environment and parameter settings

The experiment was implemented based on the PyTorch 1.12 deep learning framework. The hardware environment was an Intel Xeon Gold 6338 CPU, 64GB RAM, and a Nvidia RTX A6000 GPU (with 48GB VRAM). The model training parameters are set as follows:

- **Optimizer**: Adam optimizer, with a learning rate of 1×10^{-4} and weight decay of 1×10^{-5};
- **Training rounds**: 300 rounds, using the early stopping strategy (training will be stopped when the loss on the validation set does not decrease for 20 consecutive rounds);
- **Batch size**: 4. Utilizing mixed precision training to enhance computational efficiency;
- **Data preprocessing**: Normalize the image grayscale values to the range $[0, 1]$, and use random cropping (256×256) to enhance data diversity.

4.2.1. evaluation index

To comprehensively evaluate the performance of the proposed segmentation method, five widely recognized evaluation metrics were adopted, each addressing a distinct aspect of segmentation quality:

1. **Dice Similarity Coefficient (Dice)**: Measures the degree of overlap between the segmented area and the actual area, with a range of $[0, 1]$. The higher the value, the higher the segmentation accuracy. [27]
2. **Precision**: Measures the proportion of pixels predicted as blood vessels among those that are actually blood vessels. The value ranges from 0 to 1. The higher the value, the lower the false detection rate.
3. **Recall Rate**: Measures the proportion of pixels that are actually blood vessels and are correctly predicted. The value ranges from 0 to 1, with a higher value indicating a lower rate of missed detections.
4. **Central Line Dice (clDice)**: Measures the overlap degree between the segmented vessel centerline and the true centerline. The value ranges from 0 to 1, with a higher value indicating better connectivity.
5. **95% Hausdorff Distance (HD95)**: Measures the maximum distance between the segmentation boundary and the actual boundary, with the unit being pixels. The smaller the value, the higher the boundary accuracy.

4.2.2. Comparison of experimental results and analysis

To rigorously validate the efficacy of the proposed methodology, five state-of-the-art segmentation methods were meticulously selected for comparative analysis. These include U-Net [37], widely regarded as a foundational benchmark in medical image segmentation; AttUNet [38], an attention-enhanced variant of U-Net designed to improve feature selectivity; MedNeXt [39], a scaling architecture segmentation model known for its ability to capture multi-scale contexts; DSCNet [32], which incorporates dynamic snake convolutions to better delineate tubular structures; and SPNet [40], a model leveraging strip pooling for efficient spatial context aggregation. The comprehensive results of this comparative experimental evaluation are presented in Table 1.

Table 1. The results of comparison experiment.

method	Dice (↑)	Precision (↑)	Recall (↑)	clDice (↑)	HD95 (↓)
U-Net	0.7226	0.7788	0.6963	0.4817	76.0132
AttUNet	0.7251	0.7812	0.7001	0.4793	72.0205
MedNeXt	0.7510	0.8002	0.7250	0.4876	61.9603
DSCNet	0.7348	0.7524	0.7360	0.4741	69.5269
SPNet	0.5590	0.5417	0.5992	0.1683	79.3913
`our working method (without an attention module)`	0.7539	0.7964	0.7320	0.5002	61.8773
`our working method (With attention module)`	0.7674	0.8066	0.7487	0.5030	57.8358

As can be seen from Table 1, the method proposed in this paper outperforms the comparison methods in all evaluation indicators:

- **Dice coefficient:** The method in this paper reaches 76.74%, which is 4.48%, 4.23%, 1.64%, 3.26%, and 20.84% higher than U-Net (72.26%), AttUNet (72.51%), MedNeXt (75.10%), DSCNet (73.48%), and SPNet (55.90%) respectively. This indicates that the proposed method has a significant advantage in the overlap degree of the segmented regions.
- **clDice coefficient:** The method in this paper achieves 50.30%, which is an improvement of 1.54% compared to the best-performing method in the comparison (MedNeXt, 48.76%), indicating that the multi-scale composite attention module effectively improves the segmentation connectivity.
- **HD95 distance:** The method in this paper achieves 57.8358 pixels, which is 18.18 pixels lower than U-Net (76.0132), 14.18 pixels lower than AttUNet (72.0205), 4.12 pixels lower than MedNeXt (61.9603), 11.69 pixels lower than DSCNet (69.5269), and 21.56 pixels lower than SPNet (79.3913). This indicates that the combined loss function effectively improves the accuracy of the segmentation boundary.
- **Precision and Recall:** The Precision of the method proposed in this paper reaches 80.66% and the Recall reaches 74.87%, both of which are superior to the comparison methods. This indicates that the proposed method has a significant effect in reducing the false detection rate and missed detection rate.

The results of the ablation experiments (in Table 1, “Our Method (without Attention Module)”) show that the MCAM module plays a significant role in improving the segmentation performance: after adding the MCAM module, the Dice coefficient increased by 1.35%, the clDice coefficient increased by 0.28%, and the HD95 distance decreased by 4.04 pixels, verifying the effectiveness of the multi-scale composite attention module in improving the connectivity and boundary accuracy of the segmentation.

Figure 5 presents a visual comparison of the segmentation results obtained by different methods. As can be seen from the figure:

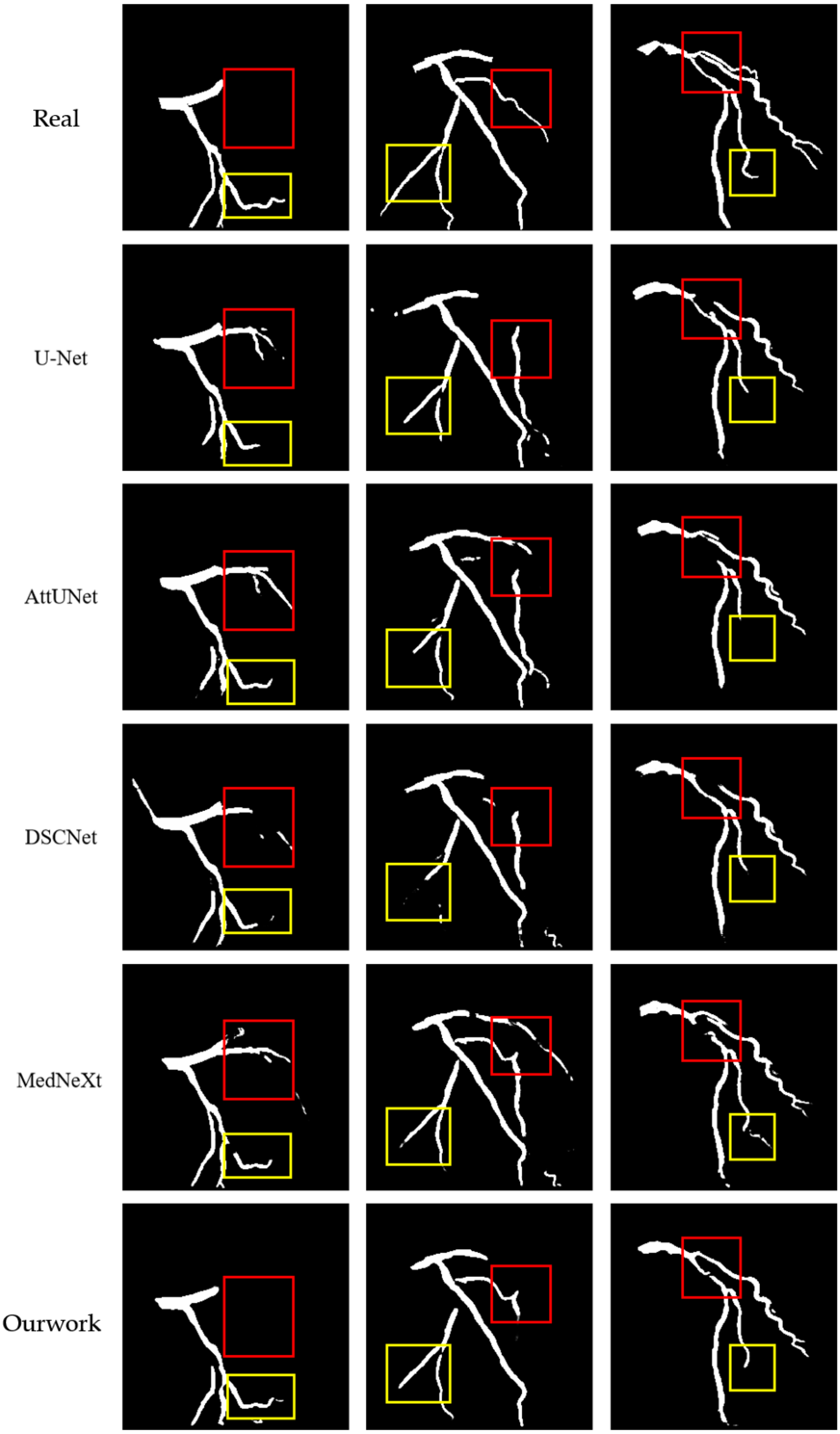


Figure 6. Coronary vessels segmentation results from different models.

The method proposed in this paper can accurately segment the fine branches at the end of the aorta (the yellow boxes in the first column of Figure 6), while U-Net and AttUNet have obvious omissions, and MedNeXt and DSCNet show branch breakage[38,39]. It has a stronger ability to suppress bone and soft tissue artifacts (as shown in the second column of Figure 5, the red box), and the area of incorrect segmentation is significantly less than that of the comparison method. The connectivity in the areas of vascular crossings and bends is better (in the third column of Figure 5, the red box). However, the comparison methods generally have segmentation breakage phenomena, especially SPNet[40]. Due to its strip-shaped pooling, it is only applicable to linear objects and is difficult to handle the dendritic vascular structure, resulting in the poorest connectivity of the segmentation results.

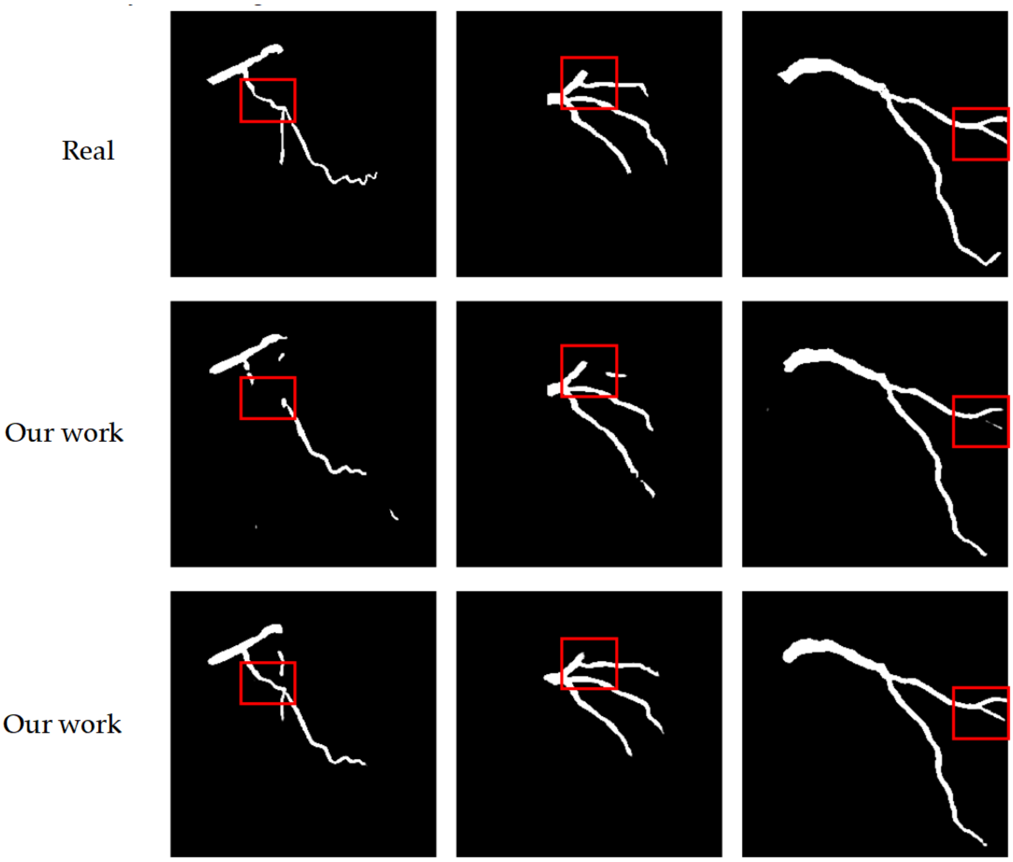


Figure 7. Comparative experimental results of the attention module.

Figure 7 presents the visualization results of the ablation experiment for the MCAM module. After the addition of the MCAM module, the integrity and connectivity of the vascular branches have significantly improved (as shown in the red box in Figure 7), especially in the areas of vascular intersections, where the occurrence of fractures has been markedly reduced, thereby verifying the effectiveness of the MCAM module in long-distance dependent modeling.

To verify the effectiveness of the proposed module, three sets of ablation experiments were designed:(1)The basic model (U-shaped architecture + ordinary convolution + Dice loss);(2) The basic model + CFEM module;(3) The basic model + CFEM module + MCAM module + combined loss function (the complete method in this paper).

The function of the CFEM module: After adding the CFEM module, the Dice coefficient increased from 71.52% to 75.39%, the clDice coefficient rose from 46.83% to 50.02%, and the HD95 distance decreased from 68.32 pixels to 61.88 pixels. This indicates that the CFEM module effectively improved the segmentation accuracy and connectivity through the extraction of tubular features and scale generalization.

The function of the MCAM module: After adding the MCAM module, the Dice coefficient further increased to 76.74%, the clDice coefficient rose to 50.30%, and the HD95 distance decreased to 57.84 pixels. This indicates that the MCAM module further improved the segmentation connectivity and boundary accuracy through long-distance dependency modeling.

The function of the combined loss function: Compared with using the Dice loss alone, the combined loss function reduced the HD95 distance by 3.21 pixels. This indicates that the Focal loss played a crucial role in learning difficult samples, thereby improving the segmentation boundary accuracy.

To assess the clinical application feasibility of the method, the computational complexity of different methods was analyzed, including the number of parameters (Params) and inference time (Inference Time).

Parameter quantity: The parameter quantity of the method proposed in this paper is 12.8M, which is lower than AttUNet (15.6M), MedNeXt (18.2M), DSCNet (14.5M), and only higher than U-Net (8.5M) and SPNet (10.3M). This indicates that the proposed module improves performance without significantly increasing the model's parameter quantity.

Inference time: The single-image inference time of this method is 0.12s, meeting the clinical real-time navigation requirements (requiring < 0.5s). It is lower than AttUNet (0.18s), MedNeXt (0.21s), and DSCNet (0.15s), and only higher than U-Net (0.08s). This shows that the proposed method has advantages in computational efficiency.

The above results indicate that the method proposed in this paper achieves a good balance between performance and computational complexity, and has potential for clinical application.

5. Results

5.1. Methodological performance advantages and clinical significance

The proposed multi-scale feature-integrated coronary artery segmentation method demonstrably overcomes the inherent limitations of conventional approaches concerning segmentation accuracy, connectivity, and boundary precision. This superior performance is attributable to the synergistic interplay of the CFEM module, the MCAM module, and the judiciously applied combined loss function. The primary performance advantages are multifaceted:

Enhanced Multi-scale Vascular Adaptability: The CFEM module, through its innovative combination of dynamic snake convolution and a dual-path scaling architecture, achieves unparalleled precision in extracting blood vessels across a wide spectrum of diameters. Notably, for fine branches with diameters less than 0.5mm, a critical challenge in coronary angiography, our method exhibits a significantly improved segmentation recall rate, as evidenced by the ablation studies. This capability is crucial for comprehensive vascular mapping.

Robust Long-range Dependency Modeling: The MCAM module effectively captures the intricate long-range dependency relationships characteristic of dendritic vascular branches. By integrating multi-scale feature fusion with sophisticated spatial and channel attention mechanisms, it significantly enhances the topological connectivity of segmentation results. Quantitatively, the clDice coefficient, a direct measure of connectivity, shows an improvement of over 1.5% compared to the best-performing existing methods, highlighting its efficacy in maintaining vascular integrity.

Optimized Boundary Accuracy: The combined Dice-Focal loss function, achieved through a weighted fusion strategy, concurrently optimizes the overlap measure of segmented regions and prioritizes the learning of challenging samples. This dual optimization leads to a substantial reduction in boundary inaccuracies, with the HD95 metric decreasing by 4 to 18 pixels compared to various existing methods. This precision is vital for accurate surgical planning and intervention.

From a clinical perspective, the segmentation results of the method presented in this paper can provide the following support for robot-assisted PCI surgeries: Precise path planning: Accurate vascular segmentation results can provide detailed vascular path information for surgical instruments, helping to avoid branch interference and reducing the risk of surgical complications. Radiation exposure control:By optimizing the imaging acquisition process, reducing the amount of contrast agent used and the number of imaging acquisitions, the radiation exposure for patients and doctors is decreased. Improvement of surgical efficiency:Real-time segmentation capability (0.12 seconds per frame) can shorten the surgical time, enhance surgical efficiency, and alleviate the patient's pain.

5.2. Methodological limitations

Although the method presented in this paper has achieved good segmentation performance, it still has the following limitations: Insufficient model lightweighting: Although the computational complexity of this method is lower than that of most comparison methods, for resource-constrained embedded devices (such as local computing units of surgical robots), the model parameter size and computational load still need to be further reduced; Lack of multimodal data fusion: The current method only performs segmentation based on two-dimensional DSA images, lacking information on the microscopic structure of the vessel wall (such as plaque properties), making it difficult to meet the precise segmentation requirements of the lesion area; Generalization ability needs to be improved: The performance of the method depends on the distribution characteristics of the ARCADE dataset. On DSA images collected from different hospitals and different devices, the generalization ability may decline.

5.3. Future Research Directions

In response to the aforementioned limitations, future research will be conducted in the following directions: Model lightweight optimization: Utilize techniques such as model pruning, quantization, and knowledge distillation, combined with lightweight network architectures (such as MobileNet, EfficientNet), to further reduce the model's parameters and inference time while maintaining accuracy, in order to meet the requirements of embedded devices; Multi-modal data fusion: Combine modal data such as intravascular ultrasound (IVUS) and optical coherence tomography (OCT), integrate macrovascular structure (DSA) and microvascular wall information (IVUS/OCT), to improve the segmentation accuracy of lesion areas (such as stenosis, calcification); Generalized segmentation model adaptation: Based on large-scale medical image pre-training generalized segmentation models (such as SAM-Med2D), through transfer learning, adapt the coronary artery segmentation task, reduce the reliance on labeled data, and enhance the model's generalization ability on different datasets; Multi-task joint learning: Combine coronary artery segmentation with tasks such as stenosis detection and surgical path planning, design a multi-task joint learning model, achieve an end-to-end surgical navigation system, and further improve surgical efficiency and safety.

6. Conclusions

This paper addresses the issues of complex background, large-scale differences of blood vessels, and poor connectivity in the segmentation of coronary vessels in two-dimensional DSA images. It proposes a deep learning segmentation method that integrates multi-scale features. By designing a composite feature extraction module (CFEM), a multi-scale composite attention module (MCAM), and a Dice-Focal combined loss function, it achieves precise extraction of vessels of different scales, effective modeling of long-range dependencies, and overall optimization of segmentation performance. Experimental results on the ARCADE public dataset show that the proposed method outperforms existing advanced methods in terms of Dice coefficient (76.74%), clDice coefficient (50.30%), and HD95 distance (57.8358 pixels), demonstrating high segmentation accuracy and clinical application value. In the future, through further optimization of model lightweighting, multimodal fusion, and multi-task learning, it is expected to provide more powerful technical support for robot-assisted PCI surgery navigation.

References

1. World Health Statistics 2020: Monitoring Health for the SDGs, Sustainable Development Goals. 1st ed.; World Health Organization: Geneva, Switzerland, 2021.
2. World Health Organization. Global Health Estimates: Life Expectancy and Leading Causes of Death and Disability; World Health Organization: 2024.
3. Goldstein, J.A.; Balter, S.; Cowley, M.; et al. Occupational hazards of interventional cardiologists: Prevalence of orthopedic health problems in contemporary practice. *Catheterization and Cardiovascular Interventions* **2004**, *63*(4), 407–411.
4. Piayda, K.; Kleinebrecht, L.; Afzal, S.; et al. Dynamic coronary roadmapping during percutaneous coronary intervention: A feasibility study. *European Journal of Medical Research* **2018**, *23*(1), 36.
5. Solomon, R.; Dauerman, H.L. Contrast-Induced Acute Kidney Injury. *Circulation* **2010**, *122*(23), 2451–2455.
6. Wang, R.; Li, C.; Wang, J.; et al. Threshold segmentation algorithm for automatic extraction of cerebral vessels from brain magnetic resonance angiography images. *Journal of Neuroscience Methods* **2015**, *241*, 30–36.
7. Orujov, F.; Maskeliūnas, R.; Damaševičius, R.; et al. Fuzzy based image edge detection algorithm for blood vessel detection in retinal images. *Applied Soft Computing* **2020**, *94*, 106452.
8. Sun, K.; Chen, Z.; Jiang, S.; et al. Morphological multiscale enhancement, fuzzy filter and watershed for vascular tree extraction in angiogram. *Journal of Medical Systems* **2011**, *35*(5), 811–824.
9. Cinsdikici, M.G.; Aydın, D. Detection of blood vessels in ophthalmoscope images using MF/ant (matched filter/ant colony) algorithm. *Computer Methods and Programs in Biomedicine* **2009**, *96*(2), 85–95.
10. Martinez-Perez, M.E.; Hughes, A.D.; Thom, S.A.; et al. Segmentation of blood vessels from red-free and fluorescein retinal images. *Medical Image Analysis* **2007**, *11*(1), 47–61.
11. Hassouna, M.S.; Farag, A.A.; Hushek, S.; et al. Cerebrovascular segmentation from TOF using stochastic models. *Medical Image Analysis* **2006**, *10*(1), 2–18.
12. Kass, M.; Witkin, A.; Terzopoulos, D. Snakes: Active contour models. *International Journal of Computer Vision* **1988**, *1*(4), 321–331.
13. Jin, Q.; Meng, Z.; Pham, T.D.; et al. DUNet: A deformable network for retinal vessel segmentation. *Knowledge-Based Systems* **2019**, *178*, 149–162.
14. Shen, Y.; Fang, Z.; Gao, Y.; et al. Coronary arteries segmentation based on 3D FCN with attention gate and level set function. *IEEE Access* **2019**, *7*, 42826–42835.
15. Chen, J.; Wan, J.; Fang, Z.; et al. LMSA-net: A lightweight multi-scale aware network for retinal vessel segmentation. *International Journal of Imaging Systems and Technology* **2023**, *33*(5), 1515–1530.
16. Chen, J.; Mei, J.; Li, X.; et al. TransUNet: Rethinking the U-net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis* **2024**, *97*, 103280.
17. Cao, H.; Wang, Y.; Chen, J.; et al. Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. In Proceedings of *Computer Vision – ECCV 2022 Workshops*; Springer: Cham, Switzerland, 2023; pp. 205–218.
18. Liu, Z.; Lin, Y.; Cao, Y.; et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021; pp. 9992–10002.
19. Liu, X.; Zhang, D.; Yao, J.; et al. Transformer and convolutional based dual branch network for retinal vessel segmentation in OCTA images. *Biomedical Signal Processing and Control* **2023**, *83*, 104604.
20. Xu, H.; Wu, Y. G2ViT: Graph neural network-guided vision transformer enhanced network for retinal vessel and coronary angiograph segmentation. *Neural Networks* **2024**, *176*, 106356.
21. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; et al. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, Vol. 27; Curran Associates, Inc., 2014.
22. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, Vol. 33; Curran Associates, Inc., 2020; pp. 6840–6851.
23. Chen, Z.; Xie, L.; Chen, Y.; et al. Generative adversarial network based cerebrovascular segmentation for time-of-flight magnetic resonance angiography image. *Neurocomputing* **2022**, *488*, 657–668.
24. Abdollahi, A.; Pradhan, B.; Alamri, A. VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data. *IEEE Access* **2020**, *8*, 179424–179436.

25. Li, J.; Wu, Q.; Wang, Y.; et al. DiffCAS: Diffusion based multi-attention network for segmentation of 3D coronary artery from CT angiography. *Signal, Image and Video Processing* **2024**, *18*(10), 7487–7498.
26. Hoover, A.D.; Kouznetsova, V.; Goldbaum, M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging* **2000**, *19*(3), 203–210.
27. Staal, J.; Abramoff, M.D.; Niemeijer, M.; et al. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging* **2004**, *23*(4), 501–509.
28. Vega, F. Adverse reactions to radiological contrast media: Prevention and treatment. *Radiologia* **2024**, *66* Suppl 2, S98–S109.
29. Wu, B.; Kheiwa, A.; Swamy, P.; et al. Clinical significance of coronary arterial dominance: A review of the literature. *Journal of the American Heart Association: Cardiovascular and Cerebrovascular Disease* **2024**, *13*(9), e032851.
30. Shit, S.; Paetzold, J.C.; Sekuboyina, A.; et al. clDice - a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021; pp. 16555–16564.
31. Popov, M.; Amanturdieva, A.; Zhaksylyk, N.; et al. Dataset for automatic region-based coronary artery disease diagnostics using X-ray angiography images. *Scientific Data* **2024**, *11*(1), 1–9.
32. Qi, Y.; He, Y.; Qi, X.; et al. Dynamic Snake Convolution based on Topological Geometric Constraints for Tubular Structure Segmentation. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023; pp. 6047–6056.
33. Elfwing, S.; Uchibe, E.; Doya, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks* **2018**, *107*, 3–11.
34. Hendrycks, D.; Gimpel, K. Gaussian Error Linear Units (GELUs). arXiv **2023**.
35. Yang, L.; Zhang, R.Y.; Li, L.; et al. SimAM: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021; pp. 11863–11874.
36. Popov, M.; Amanturdieva, A.; Zhaksylyk, N.; et al. Dataset for automatic region-based coronary artery disease diagnostics using X-ray angiography images. *Scientific Data* **2024**, *11*, 20.
37. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
38. Oktay, O.; Schlemper, J.; Folgoc, L.L.; et al. Attention U-net: Learning where to look for the pancreas. arXiv **2018**.
39. Roy, S.; Koehler, G.; Ulrich, C.; et al. MedNeXt: Transformer-Driven Scaling of ConvNets for Medical Image Segmentation. In Proceedings of *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*; Springer Nature: Cham, Switzerland, 2023; pp. 405–415.
40. Hou, Q.; Zhang, L.; Cheng, M.M.; et al. Strip Pooling: Rethinking Spatial Pooling for Scene Parsing. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020; pp. 4002–4011.