

Article

BRNet: A Dual-Backbone X-Ray Coronary Angiography Segmentation Network Based on Multi-Scale Fusion and Dynamic Detail Reconstruction

Zhan Zhang , Hong Shao  and Wencheng Cui 

School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China; shaoh@sut.edu.cn (H.S.); cuiwc@sut.edu.cn (W.C.)

* Correspondence: zhangz@smail.sut.edu.cn; Tel.: +86-18068873196

Abstract

X-ray coronary angiography remains the gold standard for diagnosing coronary heart disease. However, accurate segmentation is challenged by the subtlety of fine vascular features, topological discontinuities, and blurred boundaries in these images. Existing methods often struggle to capture both the long-range global topology and the fine local details required for robust segmentation. To address these issues, we propose BRNet, a framework that integrates a dual-backbone collaborative mechanism with multi-scale fusion and dynamic detail reconstruction. Our approach first employs a Vascular Local Detail Attention Module that combines ResNet18's local perception with BiFormer's global modeling, using SimAM parameter-free attention to suppress background noise. We then design a Vascular Structure-Coherent Progressive Fusion Module, which uses a top-down pyramid semantic flow to ensure topological coherence across different vascular scales. Finally, a Vascular Enhancement Dynamic Upsampling Module replaces traditional interpolation with content-aware CARAFE operators to achieve high-resolution reconstruction of blurred edges. Experiments on the public ARCADE dataset and the constructed heterogeneous benchmark ZZ-CAHDS show that BRNet achieves superior performance, attaining IoU scores of 0.6305 and 0.6774, and cIDice coefficients of 0.7655 and 0.8355, respectively, and achieving a good balance between segmentation accuracy and computational efficiency. These results highlight BRNet's effectiveness on retrospective segmentation benchmarks, demonstrating its capability for computer-assisted coronary artery segmentation.

Keywords: coronary artery segmentation; dual-backbone network; feature fusion; content-aware upsampling; X-ray coronary angiography images



Received: 9 June 2026

Revised: 1 July 2026

Accepted: 4 July 2026

Published: 10 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Cardiovascular disease is one of the leading causes of death worldwide and poses a significant threat to human beings [1]. Coronary Artery Disease (CAD) is a prevalent form of cardiovascular disease and represents a significant health risk [2]. Coronary artery disease is mainly caused by vascular plaque deposition, which leads to luminal stenosis, which in turn restricts myocardial blood supply and causes clinical symptoms such as ischemia and chest pain [3]. In addition, unstable stenosis can lead to acute occlusion and induce heart attack [4]. Clinically, coronary artery segmentation plays an important role in the prediction, diagnosis, and treatment of CAD.

Although existing studies support the use of non-invasive computed tomographic angiography (CTA)-based methods, such as functional fractional flow reserve analysis,

as a potential alternative to X-ray coronary angiography (XCA) [5,6], and these methods are relatively non-invasive, quick, and highly acceptable to patients, the recordings and analysis of severe and obstructive CAD are still unreliable. This unreliability is due to significantly reduced resolution, the influence of visual artifacts, and the natural movement of vessels in CTA [7,8], which contribute to certain false positive and false negative rates. Therefore, most modern CAD diagnostic studies rely on XCA as the primary quantitative method, and it is often referred to as the gold standard in various sources [9,10]. However, XCA segmentation faces three core challenges: (1) The problem of feature extraction and background interference: the low contrast of the narrow calcified area and the texture similarity of the background soft tissue lead to the submergence of vascular details and the blurring of the boundary; in addition, the coronary artery has many branches and often has two-dimensional image overlap and occlusion. The receptive field of the traditional 2DCNN is limited, and it is difficult to capture the long-distance structural association across branches and overlapping areas. (2) The problem of vascular scale disparity and feature fragmentation: Coronary arteries span an expansive range of scales, with minute branches accounting for a low proportion of pixels. Their feature intensity is significantly weaker than that of the main trunk and is highly susceptible to noise interference. Global features extracted by deep networks lack correlation with local features extracted by shallow networks, resulting in global semantics failing to support the recognition of local details. Features extracted independently at different scales, without targeted fusion, exhibit redundancy. Global information repeats at shallow levels while gaps exist, and shallow-level details are absent at deep levels. Furthermore, mismatched semantic hierarchies lead to feature conflicts. (3) The problem of edge blurring and upsampling distortion: The “gray-scale gradient band” formed by contrast agent diffusion causes blurred vessel boundaries. Traditional fixed-weight upscaling algorithms, such as bilinear interpolation, not only further smooth out high-frequency edge information but also lack content-aware capabilities. This leads to the omission of local feature content information, resulting in semantic mismatches when fusing upscaled features with current-layer features. Ultimately, this disrupts the topological continuity of the vascular tree. The above factors together cause the decline of segmentation accuracy and consistency, and it is difficult to reconstruct a continuous vascular tree for clinical evaluation.

To address these issues, various methods have been explored. Traditional angiographic segmentation primarily includes the filter-based method [11], line-tracking method [12], model-based method [13], region-growing method [14], thresholding method [15], etc. However, these methods are insufficient in robustness in the case of low signal-to-noise ratios and complex vascular structures, and it is difficult to meet the clinical needs. Recently, deep learning technology with Convolutional Neural Networks (CNN) as the core has become the mainstream method of medical image segmentation. Since its introduction in 2015, UNet [16] has achieved outstanding success in restoring image details with its encoder-decoder structure and skip connection design. However, limited by the local receptive field of the convolution operation, U-Net has inherent deficiencies in long-range dependency capture and global semantic modeling, which makes it easy to cause boundary blur and vascular rupture when segmenting complex vascular topology structures. To address this problem, UNet++ [17] introduces a dense skip connection mechanism, which effectively alleviates the semantic gap between deep and shallow features by enhancing feature transfer and multi-scale information fusion. FP-U-Net++ further enhances the feature pyramid design to improve the performance of vascular detail detection [18]. Furthermore, recent engineering-focused optimizations have explored advanced architectures to tackle these challenges. For instance, Alirr [19] proposed a skip-free residual U-Net framework with progressive upsampling to enhance complex medical image segmentation. Similarly,

Wang [20] integrated a multiscale attention mechanism into the network to adaptively suppress background noise and highlight critical fine-grained features. Despite these advancements demonstrating the effectiveness of robust upsampling and attention-guided representation, these improvements have resulted in complex network architectures and heavy computational burdens, and their performance remains limited when handling high-noise data, ambiguous regions, and modeling long-range dependencies.

To overcome the limitations of Convolutional Neural Networks (CNNs) in global modeling, Vision Transformers (ViTs) [21] have been introduced to the field of medical segmentation. TransUNet [22] is a new way to combine CNNs and ViTs to capture both local features and global context at the same time. However, the inherently high computational complexity of the Transformer architecture leads to significant resource consumption when processing high-resolution XCA images. It also tends to overfit on small-sample datasets, resulting in less pronounced advantages under strong noise interference. SwinUnet [23] improves global modeling even more by using a pure Transformer architecture with window attention mechanisms. However, its ability to preserve local texture details remains relatively weak, often yielding suboptimal results when segmenting fine vascular edges.

In view of the limitations of the above Transformer-like methods, BiFormer [24], as a more efficient Transformer variant, introduces a bi-level routing attention mechanism that dynamically sparsifies attention computations in a content-aware manner. This approach effectively reduces computational redundancy while preserving robust global semantic modeling capabilities. So, BiFormer strikes the best balance between performance and resource use when processing targets like coronary arteries, which have a complicated structure but not many pixels covering them. However, relying solely on BiFormer as a baseline still has significant shortcomings: due to the lack of local inductive biases inherent in convolutional networks, its ability to capture details such as vessel edges and minute terminal branches is relatively weak. Furthermore, in low signal-to-noise ratio XCA images, the dynamic attention mechanism is susceptible to interference from complex background noise, leading to the drowning of subtle features. Moreover, its inherent hierarchical structure alone struggles to bridge the semantic gap across vastly differing vascular scales, making the vessel highly susceptible to topological fractures.

At the same time, in the specific field of coronary artery segmentation, in order to cope with the challenges of uneven illumination, low contrast, and complex topology of XCA images, researchers have proposed a variety of improvement schemes. Guo et al. [25] improved the feature pyramid model based on U-Net and optimized the vessel segmentation effect under low contrast; Xu et al. [26] proposed HAM-UNet, which combines hierarchical context aggregation with multi-scale fusion, leveraging the Transformer's advantage in capturing long-range dependencies, it addresses the issue of poor connectivity in fine blood vessels; Zhang et al. [27] proposed the EAG-SAE-ESTF network, which specifically deals with the vessel challenges of large-scale changes and discontinuous connections through entropy-driven feature extraction and elastic topology fusion; Huang et al. [28] combined MedSAM and VM-UNet architecture and introduced the dynamic queue method for vessel stenosis detection and segmentation. Although these specific field methods have their own advantages, it is still difficult to take into account the integrity of the main trunk and the end when facing the extreme scale span and strong background interference in XCA images.

In addition, although general-purpose large visual models such as Mask2Former [29] and SAM 2 [30] have shown strong performance in natural image segmentation, they still face challenges in XCA tasks. The unified mask attention mechanism of Mask2Former is not accurate enough when dealing with the edges of small blood vessels; while SAM 2, although introducing a promptable segmentation mechanism to improve flexibility, its

dependence on prompt information makes it difficult to directly apply to a fully automated clinical pipeline.

To solve the limitations of existing methods in fine blood vessel extraction, edge blur processing, and global semantic modeling, this paper proposes BRNet, which follows the general idea of “feature enhancement-multi-scale fusion-edge optimization” and overcomes the problem of XCA image segmentation through three core modules. The main contributions of this paper are:

- (1) A dual-backbone collaborative perception mechanism is proposed: In response to the problem of feature extraction and background interference, a Vascular Local Detail Attention Module (VLDAM) is designed. This module combines the local inductive bias of ResNet [31] and the global long-range modeling capability of BiFormer [24] and uses SimAM [32] parameter-free attention to adaptively suppress background noise from the physical energy level, which significantly enhances the saliency of vascular features.
- (2) Design a structure-coherent multi-scale progressive fusion strategy: To address the feature fragmentation problem caused by the large span of vascular scales, a Vascular Structure-Coherent Progressive Fusion Module (VSC-PFM) is proposed. The top-down pyramid semantic flow and dense feature interaction effectively bridged the cross-scale semantic gap and ensured the topological coherence of the main trunk and the fine terminal.
- (3) Constructing a content-aware dynamic reconstruction decoder: To address the problems of edge smoothing and detail loss caused by traditional linear interpolation, a Vascular Enhancement Dynamic Upsampling Module (VE-DUSM) was constructed. This module adopted a collaborative decoding strategy that integrated the CARAFE operator [33] and cross-scale concatenation. It used content-aware reorganization to replace fixed-weight interpolation and achieved high-resolution accurate reconstruction of blurred vascular edges and fine branches.
- (4) In addition to evaluating the performance on the standard ARCADE dataset, this paper constructs the ZZ-CAHDS benchmark containing multi-source heterogeneous data. Visualization-based statistical analysis reveals that this benchmark exhibits significant domain shift and higher statistical heterogeneity. The experimental results indicate that BRNet shows better robustness than most compared methods on this challenging benchmark, suggesting its potential value as a robust reference model for future XCA coronary artery segmentation studies under heterogeneous data distributions.

2. Materials and Methods

2.1. Dataset Construction and Distribution Disparity Analysis

In order to comprehensively evaluate the segmentation accuracy of BRNet in standard scenes and its robustness in complex environments, this paper adopts a dual-dataset evaluation strategy of “standard benchmark + heterogeneous benchmark.” Specifically, it includes the publicly available standard dataset ARCADE and the multi-source mixed dataset ZZ-CAHDS constructed in this paper.

2.1.1. Dataset Construction

The ARCADE dataset [34] is the authoritative benchmark in the field of coronary angiography segmentation, containing 1500 images with a resolution of 512×512 . The dataset has undergone strict standardized acquisition and preprocessing and has a unified circular imaging field of view and a relatively consistent contrast distribution. However, some finer structures, such as capillaries, are not annotated in this dataset.

The existing single-center datasets often have the problems of single data distribution and single background, which make it difficult to reflect the “domain shift” challenge brought by multi-center and multi-device in the real clinical environment. To this end, this paper constructs a highly heterogeneous mixed dataset, ZZ-CAHDS. The dataset integrates some samples from DCA1 [35], XCAD [36], CHUAC [37], and the ARCADE original validation set, with a total of 490 images. The original vessel annotations provided by each source dataset were retained and used as the ground truth. In view of the differences in original image resolution among the multi-source datasets, all images were resized to 512×512 pixels using bilinear interpolation to construct a unified input space for batch training and fair comparison across models. The corresponding masks were converted back into binary vessel/background labels after resizing. During dataset construction, each image-mask pair was checked for file integrity, spatial correspondence, resolution consistency, and label validity.

The division of the two datasets and the composition of the ZZ-CAHDS dataset are shown in Tables 1 and 2, and there is no sample overlap between the subsets.

Table 1. Dataset split in ARCADE and ZZ-CAHDS.

Dataset Name	Train	Val	Test	All
ARCADE	1000	200	300	1500
ZZ-CAHDS	390	50	50	490

Table 2. Composition of the ZZ-CAHDS Dataset.

Dataset Name	Number	Image Format	Resolution
ARCADE	200	PNG	512×512
DCA1	134	PGM	300×300
XCAD	126	PNG	512×512
CHUAC	30	PNG	189×189

2.1.2. Analysis of Data Distribution Disparities and Construct Validity

To demonstrate the necessity of constructing ZZ-CAHDS and its value as a challenging evaluation benchmark, this paper quantitatively compares the pixel intensity distribution and statistical heterogeneity of the two datasets using visualization-based statistical methods, as shown in Figure 1.

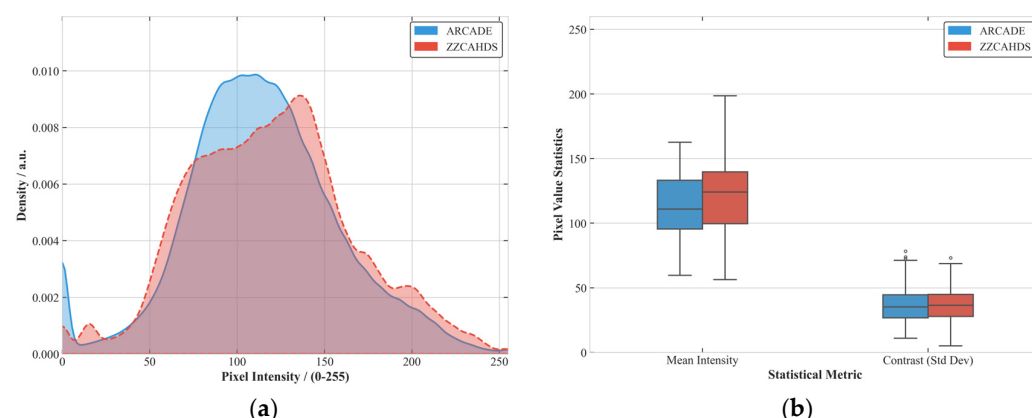


Figure 1. Comparison of Distribution Characteristics Between ARCADE and ZZ-CAHDS Datasets. (a) Pixel Intensity Distribution Density Map. (b) Box Plot for Image Statistical Metrics.

(1) Pixel Intensity Domain Offset

As shown in the pixel intensity density map in Figure 1a, ARCADE and ZZ-CAHDS exhibit obvious distribution differences. From the perspective of preprocessing characteristics, ARCADE shows a pronounced peak at pixel value 0, indicating that the background region has been relatively uniformly masked. In contrast, ZZ-CAHDS presents a flatter distribution near pixel value 0, suggesting that more background information is retained due to differences in image fields of view and preprocessing styles across the source datasets. Regarding brightness distribution, the intensity peak of ARCADE is mainly concentrated in the [90,100] interval with a relatively narrow distribution, whereas the peak of ZZ-CAHDS shifts toward the [130,140] interval and covers a wider high-intensity range in [150,250]. This right-skewed long-tail distribution indicates that ZZ-CAHDS contains samples with more diverse exposure levels, contrast-agent filling conditions, and background appearances, thereby forming a clear domain shift compared with the standardized ARCADE dataset.

(2) Statistical Heterogeneity and Difficult Samples

The boxplot in Figure 1b further quantifies the differences between the two in “average intensity” and “contrast.” In terms of variance comparison, the height of the red box corresponding to ZZ-CAHDS is significantly higher than that of ARCADE, indicating that the fluctuation range of image brightness is larger and the variance is higher. In the analysis of difficult long-tail samples, although the median of the contrast index is similar for both methods, ZZ-CAHDS shows a wider interquartile range. In addition, the small circles outside the whisker line in the figure represent the difficult samples with extremely low contrast or extreme texture in the dataset. Although ARCADE shows more outliers, this just reflects that the main data distribution is too compact and single, resulting in a small deviation being judged as abnormal. In contrast, ZZ-CAHDS shows a significantly longer whisker range in both average intensity and contrast. Especially in the contrast index, the lower whisker line of ZZ-CAHDS extends to the extremely low value area, indicating that the dataset includes long-tail difficult samples, such as extremely low contrast in the conventional distribution, rather than treating them as outliers. This feature of covering a wider dynamic range makes ZZ-CAHDS more heterogeneous than the standardized ARCADE dataset and provides a challenging benchmark for evaluating model behavior under mixed data distributions.

In summary, ZZ-CAHDS is not merely a simple sample expansion but a challenging evaluation benchmark characterized by significant domain shifts in data distribution and greater heterogeneity in statistical features. The results obtained on this benchmark provide supporting evidence that BRNet has the potential to maintain robustness under heterogeneous coronary angiography data.

2.2. Experimental Environment and Evaluation Indicators

All experiments in this study were conducted under the Ubuntu 20.04 operating system on an NVIDIA GeForce RTX 3090 GPU. The software environment was based on Python 3.9.19 and PyTorch 1.13.1. During training, the input image size was set to 512×512 , and the hyperparameter settings are shown in Table 3. To ensure a fair comparison, all competing models were retrained and evaluated under the same experimental conditions, including the same train/validation/test split, input size, preprocessing pipeline, training epochs, optimizer, loss function, hardware platform, and evaluation metrics.

Table 3. Hyperparameter settings.

Hyperparameters	Value
batch size	4
learning rate	0.0001
epoch	100
early stop	20
optimizer	Adam
scheduler	CosineAnnealingLR
loss	BCEDiceLoss
Weight decay	0.001

To comprehensively evaluate the detection performance of the proposed method, this paper selects eight core indicators in the field of image segmentation: intersection over union (IoU), Dice coefficient, precision, accuracy, specificity, Matthews correlation coefficient (MCC), sensitivity, and area under the precision-recall curve (AUPRC) to comprehensively measure the segmentation performance of the model. These indicators can reflect the detection accuracy and robustness of the model from different angles and facilitate the horizontal comparison with the current mainstream methods. Considering the severe class imbalance in coronary artery segmentation, where background pixels are much more abundant than vessel pixels, AUPRC is used to better evaluate the precision-recall trade-off and the model's behavior regarding false-positive predictions.

2.3. Methods

2.3.1. Overall Architecture

In this paper, BiFormer [24] is selected as the baseline network, and a dual-backbone collaborative end-to-end segmentation framework, BRNet, is designed. While utilizing its efficient global modeling characteristics, dual-backbone collaboration, multi-scale progressive fusion, and dynamic upsampling strategies are introduced to solve its shortcomings in local detail extraction, noise immunity, and cross-scale connectivity. The overall structure is shown in Figure 2. The input image is first uniformly scaled to 512×512 pixels and sent to the dual-backbone encoder for multi-layer feature extraction. At this stage, the VLDAM is used as the core component to realize the capture of local details of blood vessels and the suppression of background interference by integrating ResNet18 [31], the channel adaptation mechanism, and SimAM [32] parameter-free attention. At the same time, the BiFormer encoder models global semantics and long-range dependencies with the serial process of “downsampling layer-feature modeling layer.” The dual-backbone features are added and fused to provide high discrimination support for subsequent modeling. In the multi-scale modeling stage, the VSC-PFM module adopts a top-down multi-scale progressive fusion path and realizes cross-scale semantic and detail integration through 1×1 convolution dimensionality reduction and 3×3 convolution feature interaction, which alleviates the problem of scale segmentation and indirectly ensures the coherence of vascular structure. In the decoding stage, the VE-DUSM module solves the problems of blurred vascular edges and traditional upsampling distortion through the process of “ 1×1 channel adaptation-CARAFE [33] dynamic upsampling-layer-by-layer concatenation” and strengthens detail recovery. Finally, the optimized enhanced features are integrated by convolution and pixel-level mapping, and the binary segmentation results of blood vessels are output, which ensures the topological integrity of the main trunk and the end of the coronary artery, improves the accuracy of detail recovery, and provides a high-quality segmentation basis for vascular tree reconstruction and coronary heart disease assessment.

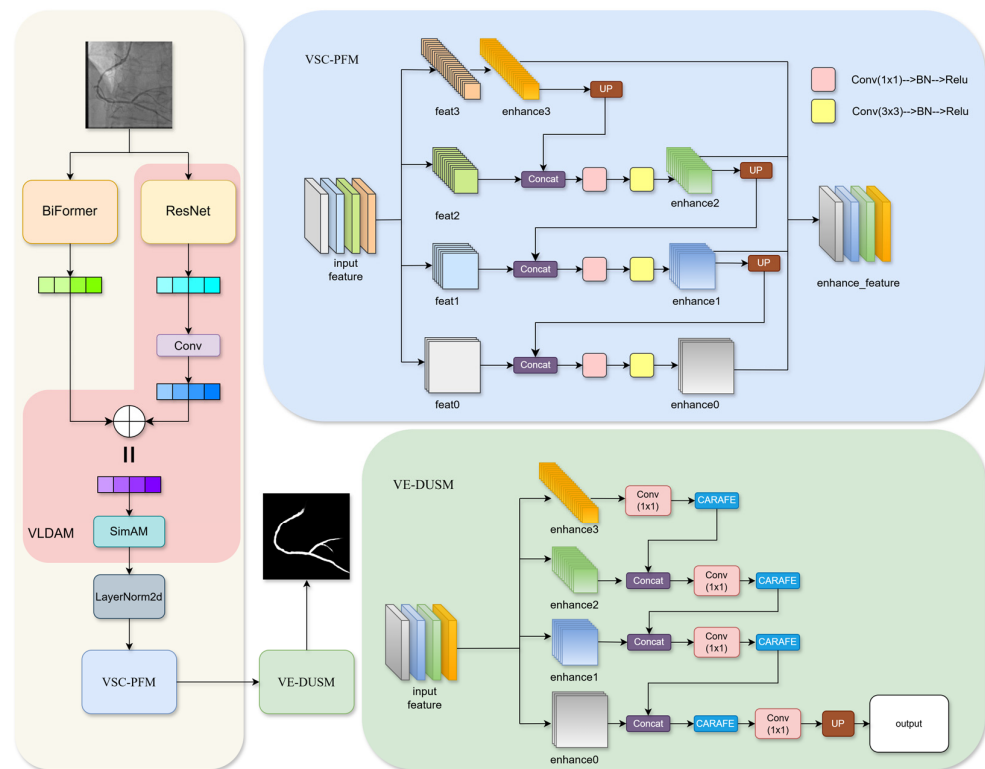


Figure 2. Overall structure of BRNet.

2.3.2. Vascular Local Detail Attention Module (VLDAM)

To address the issues of low contrast between blood vessels and background, overlapping grayscale textures between soft tissue and blood vessels, and insufficient long-distance association capture in traditional networks in XCA images, this encoder adopts a parallel deployment of the VLDAM and BiFormer encoder to construct a complementary feature expression paradigm of “local feature enhancement-global semantic modeling-cross-backbone fusion.” As shown in the VLDAM in Figure 2, the module realizes the accurate complementation and optimization of the dual-backbone features through the organic integration of the ResNet18 residual network, the channel adapter, and the SimAM parameter-free attention mechanism: on the one hand, leveraging the multi-level residual structure of ResNet18, it captures pixel-level local features such as vessel edges, texture details, and fine branches layer by layer after initialization through convolution and max pooling, while simultaneously filtering high-frequency noise in the image; to achieve dimensional alignment with BiFormer features, a channel adapter is introduced. It employs 1×1 convolutions to map ResNet18 features to an embedding dimension consistent with BiFormer, generating local feature maps $\tilde{R}_i (i = 0, 1, 2, 3)$. On the other hand, the BiFormer encoder processes the input image through downsampling and feature modeling layers. It uses the bi-level routing attention mechanism to break through the receptive field limitation of the traditional CNN, gradually aggregating cross-regional information. This allows the model to efficiently capture global semantic information and long-range dependencies. Consequently, it accurately depicts the global topological structure and cross-branch associations characteristics, solves the structural fracture problem caused by overlapping coronary branches, and generates the global feature map $B_i (i = 0, 1, 2, 3)$, which provides semantic context support for local features. After the feature maps of the two networks are dimensionally adapted, the element-by-element addition fusion method

is adopted to realize the deep complementarity of local details and global semantics and avoid feature redundancy and information conflict, and the fusion formula is as follows:

$$F_i^{(0)} = \tilde{R}_i + B_i \quad (1)$$

where $F_i^{(0)}$ is the fused feature map. This method ensures that the feature map retains both the pixel-level edge details and the long-distance topological semantics.

Next, the dynamic weighting characteristics of the SimAM parameter-free energy function are used to adaptively enhance the feature response of the low-contrast blood vessel region based on the difference in feature dispersion between the blood vessel and the background while suppressing background interference information such as myocardial tissue and equipment electronic noise and solving the problem of “blood vessel features being submerged by noise” from the source of feature fusion, providing high discriminant feature support for subsequent multi-scale modeling. First, calculate the mean μ and variance σ^2 of the features:

$$\mu = \sum_{x=1}^H \sum_{y=1}^W F_i^{(0)}(x, y) \quad (2)$$

$$\sigma^2 = \frac{1}{H \times W - 1} \sum_{x=1}^H \sum_{y=1}^W (F_i^{(0)}(x, y) - \mu)^2 \quad (3)$$

where $F_i^{(0)}(x, y)$ represents the value of the i -th feature map at the spatial position (x, y) after the input data is processed by the network layer ($i = 0, 1, 2, 3$). H is the height of the feature map, and W is the width of the feature map. $H \times W - 1$ is the correction factor for unbiased estimation. After that, the energy function and the attention weight Att are generated:

$$E(F_i^{(0)}) = \frac{(F_i^{(0)} - \mu)^2}{4(\sigma^2 + \lambda)} + \frac{1}{2} \quad (4)$$

$$Att_i = \text{sig}(E(F_i^{(0)})) \quad (5)$$

where $E(F_i^{(0)})$ is the energy function ($i = 0, 1, 2, 3$) used to quantify the importance of features; and λ is the regularization constant used to avoid division by zero errors. sig is the activation function Sigmoid, which maps the energy value to the interval $[0, 1]$ to obtain the weight of each pixel. Finally, SimAM multiplies the generated attention coefficient with the fused feature map $F_i^{(0)}$ to obtain the enhanced feature map F_i :

$$F_i = Att_i \odot F_i^{(0)} \quad (6)$$

where F_i is the final output feature of the VLDAM ($i = 0, 1, 2, 3$), and $\odot$ represents the element-wise multiplication operation. Through this process, SimAM can automatically assign a larger weight to the important areas in the feature map, such as the blood vessel area, so as to enhance the response of these areas and suppress the interference of the background or irrelevant areas.

The VLDAM effectively solves the problem of XCA feature extraction and background interference and provides the core support for the subsequent multi-scale fusion module with low noise and strong representation, laying the foundation for the high-precision segmentation of BRNet.

2.3.3. Vascular Structure-Coherent Progressive Fusion Module (VSC-PFM)

To address the core challenges of extreme scale variation from the coronary artery trunk to its terminal branches, the lack of correlation between global and local features, and

semantic hierarchy mismatch leading to feature fragmentation, the VSC-PFM module is proposed for multi-scale feature modeling, as illustrated in Figure 2. This module employs a top-down, multi-scale progressive fusion pathway. Leveraging the characteristic of XCA images where the semantic significance of the main vascular trunk is pronounced while distal details are faint, it constructs a three-stage progressive fusion paradigm: “global semantic guidance—local detail supplementation—feature coherence optimization.” The specific workflow is as follows:

In the multi-scale hierarchical capture process, the encoder part of BRNet generates four layers of features: $\{F_{feat3}, F_{feat2}, F_{feat1}, F_{feat0}\}$. The features of each layer are adapted to different scales of vascular structures. Among them, F_{feat3} represents the deepest layer of features, which contains global semantic information and mainly depicts the overall distribution and topological structure of the main blood vessels, providing global contextual guidance for the positioning of small branches; F_{feat2} represents the deep layer of features, which takes into account the local structure and semantic information, and focuses on the morphology of medium-sized blood vessels; F_{feat1} represents shallower-level features, emphasizing details such as vessel edges and branch junctions to support the continuity of vascular structures; F_{feat0} represents the shallowest layer of features, which retains the local details of the highest resolution, accurately captures the pixel-level contours of the terminal small branches, and solves the problem that the features of the terminal small branches are weak and easy to be missed. The core advantage of multi-scale hierarchical capture lies in the targeted design of features at different levels, avoiding the insufficient adaptation of single-scale features to blood vessels of different sizes, ensuring that the features of main blood vessels, medium branches, and terminal branches can be effectively extracted, and laying a foundation for subsequent fusion.

In order to further enhance the fusion effect of multi-scale features, the module introduces a pyramid progressive fusion strategy, which integrates deep global information and shallow local detail information through progressive layers to obtain rich and fine feature expressions. First, the deepest feature map, F_{feat3} , remains unchanged and serves as the third layer of the enhanced feature map, providing a global semantic benchmark for the entire fusion process:

$$F_{enhance3} = F_{feat3} \quad (7)$$

Second, enhanced features are generated layer by layer ($i = 2, 1, 0$): For the enhanced feature map $F_{enhance_i}$ of the i -th layer, the enhanced feature map $F_{enhance_{(i+1)}}$ of the previous layer is upsampled to the same resolution as the current layer by bilinear interpolation and then channel concatenated with the original feature map F_{feat_i} of the current layer to integrate deep semantic and shallow details so as to obtain the spliced feature F_{concat_i} :

$$F_{concat_i} = \text{UpSampling}(F_{enhance_{(i+1)}}) \oplus F_{feat_i} \quad (8)$$

wherein UpSampling represents the bilinear interpolation upsampling operation, $\oplus$ represents channel concatenation, and then the serial processing of “ 1×1 convolution-batch normalization-ReLU- 3×3 convolution-batch normalization-ReLU” is used to achieve the dual goals: On the one hand, by applying 1×1 convolutions to compress redundant channels and unify feature dimensions, followed by 3×3 convolutions that enable deep-layer global semantics and shallow-layer local details to interact fully within pixel-level neighborhoods, this approach bridges the semantic gap across scales and resolves the issue where global features alone cannot support local detail recognition. On the other hand, this approach naturally enhances adjacent pixels by leveraging the local neighborhood coverage property of 3×3 convolutions, while also mitigating redundant representations and semantic conflicts across different scales. This indirectly guarantees structural coher-

ence between primary vessels and peripheral branches, circumventing the limitations of conventional multi-scale fusion—such as semantic dilution of peripheral details or local features that segment the global structure—thereby achieving a uniform representation of vascular characteristics across all scales, thereby obtaining the enhanced feature map $F_{enhance_i}$ of the i -th layer:

$$F_{enhance_i} = ReLU(BN(Conv3 \times 3(ReLU(BN(Conv1 \times 1(F_{concat_i})))))) \quad (9)$$

where $Conv1 \times 1$ represents 1×1 convolution, which is used for feature dimension reduction and de-redundancy, and $Conv3 \times 3$ represents 3×3 convolution, which is used for feature interaction enhancement. BN is Batch Normalization, and $ReLU$ is ReLU activation processing.

The core values of the progressive fusion strategy are: first, through top-down semantic transmission, the shallow features can retain the details of the terminal branches and obtain the global topology guidance of the main blood vessels at the same time so as to avoid missing detection or being free of small branches due to semantic isolation. The second is to alleviate the semantic conflict of different-scale features and strengthen the consistency of features through the de-redundancy and interaction of convolution enhancement blocks. Third, the layer-by-layer fusion ensures the characteristic correlation between the main artery and the terminal branches, improves the continuity of the vascular topology, and effectively solves the problem of branch connection fracture.

Finally, the output of the VSC-PFM module is four independent enhanced feature maps, $F_{enhance_3}$, $F_{enhance_2}$, $F_{enhance_1}$, and $F_{enhance_0}$, which correspond to vascular structures of different scales, which not only maintain the global semantic consistency but also retain the integrity of local details, providing comprehensive and coherent feature support for the subsequent decoding stage.

The VSC-PFM module enhances the transmission of global semantic information while maintaining local details through multi-scale hierarchical capture and the pyramid progressive fusion strategy. BRNet effectively strengthens the segmentation of small blood vessels and ensures the global semantic consistency of the image through the strategy of layer-by-layer progressive fusion and feature concatenation. The fusion of different-scale features ensures the balance between the details and global information of the model, which improves the segmentation accuracy. The method of layer-by-layer upsampling and feature concatenation makes the deep semantic information and shallow detail information effectively integrated and provides a more comprehensive feature representation for the final segmentation result of BRNet.

2.3.4. Vascular Enhancement Dynamic Upsampling Module (VE-DUSM)

In order to solve the problems of 5–8 pixel gray gradient bands at the edge of blood vessels in XCA images, the distortion caused by the smooth features of traditional upsampling, and the edge blur and upsampling distortion caused by insufficient compatibility of cross-scale features, this paper constructs a content-aware dynamic reconstruction decoder, as shown in VE-DUSM in Figure 2. This module employs a collaborative optimization strategy of “channel adaptation—dynamic upsampling—cross-scale synergistic fusion.” It first applies a 1×1 convolution channel adaptation transformation to deep-enhanced features. Through dimensionality reduction and feature refinement, this step aligns the channel dimensionality of deep semantic features with the shallow detail features of the current layer. This eliminates channel discrepancies in cross-scale fusion while preserving the discriminative power of core vascular features. Unlike the encoding stage’s VSC-PFM, which focuses on semantic flow transfer and employs only efficient bilinear interpolation, the decoding end directly determines the edge quality of segmentation masks. Therefore,

we have introduced the CARAFE operator [33], which has a slightly higher computational cost but has the ability of content perception in this key link to achieve pixel edge sharpening with the best performance price ratio to replace the fixed weight strategy of traditional bilinear/bicubic interpolation. CARAFE dynamically adjusts kernel parameters based on local feature distributions in XCA images—such as gray-level gradient bands along vessel edges and sparse features of fine branches—to avoid smoothing suppression of weak details. This enables precise amplification of low-contrast vessel edges and fine branch features. Finally, a layer-by-layer progressive upsampling cross-scale concatenation process is constructed, and the deep semantic features after dynamic upsampling are channel concatenated with the current shallow detail features. Through the complementary integration of feature dimensions, the deep semantic information and shallow detail information are mutually enabled, which can alleviate the semantic gap and insufficient spatial adaptation of cross-scale features and realize the accurate recovery of detail features. The specific process is as follows: The VE-DUSM module takes multi-scale features $F_{enhance_3}$, $F_{enhance_2}$, $F_{enhance_1}$ and $F_{enhance_0}$ from the VSC-PFM module as input. It employs a “deep-to-shallow, layer-by-layer concatenation” strategy to ensure each concatenation stage integrates deep-level semantics with shallow-level details. Features at each layer undergo convolution before being concatenated with the results from the previous layer. First, we process F_{feat3} and F_{feat2} . At this stage, the deep feature map F_{feat3} undergoes a 1×1 convolution operation and CARAFE dynamic upsampling to reconstruct details before being concatenated with the shallow feature map F_{feat2} :

$$F_{concat3} = CARAFE(Conv(F_{feat3})) \oplus F_{feat2} \quad (10)$$

where $Conv$ is 1×1 convolution operation, CARAFE is dynamic upsampling operation, and $\oplus$ is concatenation operation. Secondly, the concatenated features $F_{concat(i+1)}$ and F_{feati} ($i = 1, 2$) are processed, and the deep semantic information is gradually integrated into the shallow high-resolution features:

$$F_{concati} = CARAFE(Conv(F_{concat(i+1)})) \oplus F_{feati} \quad (11)$$

In this process, the layer-by-layer concatenation operation ensures the effective fusion of features at different levels. After completing this operation, the feature map enters the final processing stage, and the details are further restored through CARAFE dynamic upsampling. The feature integration and mapping are performed through the convolution layer, and the original image resolution is restored by using bilinear interpolation. Finally, the output result is generated:

$$F_{carafe_final} = CARAFE(F_{concat1}) \quad (12)$$

$$F_{final} = Conv(F_{carafe_final}) \quad (13)$$

$$\hat{Y} = Bilinear\ UpSampling(F_{final}) \quad (14)$$

where F_{carafe_final} is the final feature map after CARAFE processing, which accurately restores the details. F_{final} is the final feature map after the convolution operation, which generates the final output by integrating the features of different layers. $\hat{Y}$ is the final output segmentation mask map, which restores the image resolution by bilinear interpolation upsampling.

The VE-DUSM module adopts the collaborative design of cross-scale feature concatenation and CARAFE dynamic upsampling, which effectively improves the image resolution, especially the recovery ability in small blood vessels, blurred edges, and low-contrast areas. CARAFE dynamic upsampling is based on local content adaptive generation

of the upsampling kernel, which can accurately match the edge and detail features of blood vessels, and cross-scale feature concatenation provides multi-dimensional semantic support for upsampling. The two work together to achieve the dual enhancement of detail recovery and semantic guidance. This improves the segmentation effect of BRNet in complex medical images.

3. Results

3.1. Ablation Experiments

In order to verify the rationality of the overall architecture design of BRNet and the effectiveness of each core submodule, this paper conducts detailed ablation experiments on the two datasets of ARCADE and ZZ-CAHDS. The experiment starts with the baseline network containing only BiFormer and gradually introduces VLDAM (Model A), VSC-PFM (Model B), and VE-DUSM (Model C). Tables 4 and 5 show the detailed quantitative comparison results.

Table 4. Comparison of ablation experimental performance on the ARCADE dataset.

Model	IoU	Dice	Pre	Acc	Sen	Spe	AUPRC	MCC
Baseline	0.5293 ± 0.0742	0.6891 ± 0.0660	0.6994 ± 0.0859	0.9797 ± 0.0049	0.6839 ± 0.0675	0.9899 ± 0.0035	0.7276 ± 0.0933	0.6800 ± 0.0673
A	0.5837 ± 0.0892	0.7334 ± 0.0733	0.7884 ± 0.0838	0.9844 ± 0.0044	0.7045 ± 0.0886	0.9933 ± 0.0027	0.7971 ± 0.0935	0.7327 ± 0.0737
B	0.5691 ± 0.0766	0.7223 ± 0.0639	0.7314 ± 0.0827	0.9819 ± 0.0045	0.7189 ± 0.0715	0.9909 ± 0.0033	0.7811 ± 0.0836	0.7144 ± 0.0651
C	0.5511 ± 0.0777	0.7073 ± 0.0665	0.7396 ± 0.0816	0.9814 ± 0.0047	0.6828 ± 0.0770	0.9918 ± 0.0029	0.7574 ± 0.0856	0.6998 ± 0.0678
A + B	0.6132 ± 0.0863	0.7566 ± 0.0685	0.8346 ± 0.0764	0.9854 ± 0.0040	0.6978 ± 0.0869	0.9953 ± 0.0024	0.8139 ± 0.0783	0.7542 ± 0.0677
A + C	0.5943 ± 0.0821	0.7421 ± 0.0663	0.8046 ± 0.0844	0.9841 ± 0.0044	0.7098 ± 0.0748	0.9942 ± 0.0032	0.7970 ± 0.0855	0.7363 ± 0.0675
B + C	0.5864 ± 0.0749	0.7364 ± 0.0605	0.7750 ± 0.0806	0.9834 ± 0.0041	0.7064 ± 0.0685	0.9930 ± 0.0028	0.7902 ± 0.0794	0.7302 ± 0.0617
A + B + C	0.6305 ± 0.0876	0.7697 ± 0.0689	0.8040 ± 0.0776	0.9855 ± 0.0042	0.7431 ± 0.0842	0.9938 ± 0.0027	0.8145 ± 0.0911	0.7643 ± 0.0699

Note: Baseline refers to the pure BiFormer architecture. A: VLDAM (Vascular Local Detail Attention Module); B: VSC-PFM (Vascular Structure-Coherent Progressive Fusion Module); C: VE-DUSM (Vascular Enhancement Dynamic Upsampling Module). Bold values indicate the best results.

Table 5. Comparison of ablation experimental performance on the ZZ-CAHDS dataset.

Model	IoU	Dice	Pre	Acc	Sen	Spe	AUPRC	MCC
Baseline	0.5822 ± 0.0304	0.7355 ± 0.0246	0.7108 ± 0.0413	0.9765 ± 0.0053	0.7671 ± 0.0544	0.9861 ± 0.0037	0.8057 ± 0.0369	0.7251 ± 0.0236
A	0.6538 ± 0.0411	0.7899 ± 0.0302	0.7764 ± 0.0410	0.9814 ± 0.0055	0.8080 ± 0.0569	0.9893 ± 0.0040	0.8549 ± 0.0407	0.7814 ± 0.0313
B	0.6146 ± 0.0305	0.7609 ± 0.0238	0.7536 ± 0.0340	0.9793 ± 0.0045	0.7715 ± 0.0481	0.9888 ± 0.0025	0.8389 ± 0.0289	0.7510 ± 0.0233
C	0.5983 ± 0.0334	0.7481 ± 0.0271	0.7316 ± 0.0345	0.9778 ± 0.0054	0.7688 ± 0.0525	0.9873 ± 0.0036	0.8167 ± 0.0424	0.7377 ± 0.0276
A + B	0.6687 ± 0.0359	0.8009 ± 0.0256	0.7756 ± 0.0425	0.9821 ± 0.0050	0.8317 ± 0.0485	0.9891 ± 0.0035	0.8695 ± 0.0339	0.7930 ± 0.0270
A + C	0.6661 ± 0.0387	0.7989 ± 0.0280	0.7706 ± 0.0397	0.9817 ± 0.0052	0.8320 ± 0.0425	0.9885 ± 0.0039	0.8694 ± 0.0331	0.7907 ± 0.0293
B + C	0.6297 ± 0.0321	0.7723 ± 0.0245	0.7673 ± 0.0356	0.9804 ± 0.0043	0.7796 ± 0.0390	0.9895 ± 0.0026	0.8451 ± 0.0275	0.7627 ± 0.0240
A + B + C	0.6774 ± 0.0373	0.8071 ± 0.0268	0.8159 ± 0.0389	0.9834 ± 0.0047	0.8014 ± 0.0461	0.9917 ± 0.0032	0.8678 ± 0.0451	0.7992 ± 0.0270

Note: Baseline refers to the pure BiFormer architecture. A: VLDAM (Vascular Local Detail Attention Module); B: VSC-PFM (Vascular Structure-Coherent Progressive Fusion Module); C: VE-DUSM (Vascular Enhancement Dynamic Upsampling Module). Bold values indicate the best results.

The VLDAM (Model A) module is the core solution for overcoming feature extraction bottlenecks. Experimental data demonstrate that VLDAM is the most critical component for enhancing the model's fundamental performance. On the highly heterogeneous ZZ-CAHDS dataset, the introduction of VLDAM alone achieves a leap in IoU from 0.5822 to 0.6538, accompanied by a significant increase in sensitivity. This substantial gain demonstrates that, when processing low-contrast XCA images, the integration of the ResNet branch and SimAM attention mechanism effectively compensates for the pure Transformer architecture's shortcomings in capturing local details. This significantly reduces the occurrence of small vessel misdetections caused by background noise interference.

The VSC-PFM (Model B) module enhances multi-scale structural coherence. After introducing VSC-PFM, the model achieved approximately a 4% improvement in IoU on the ARCADE dataset compared to the baseline. More importantly, both the Dice coefficient and MCC metrics showed stable increases. From the morphological perspective of segmentation results, this module significantly contributes to the continuity of vascular structures.

Through dense interactions of multi-scale features, the “disconnection” phenomenon that commonly occurs at the junctions between main vessels and fine branches is markedly reduced, demonstrating the effectiveness of the cross-scale fusion strategy in preserving the topological integrity of the vascular tree.

Although the VE-DUSM (Model C) module has a slightly smaller increase in the single numerical value than the previous two, it shows an indispensable “fine-tuning” effect in the combined experiment. In particular, in the combination of A+C, VE-DUSM further improved the IoU and Dice indicators compared with using A alone and performed well in precision. This is reflected in the visual effect; that is, the edge of the blood vessel becomes sharper and smoother, eliminating the jagged artifacts brought by traditional interpolation and realizing the high-resolution accurate reconstruction of the end of the fine branch.

When the three modules are integrated into the complete BRNet, the model achieves the best results on key indicators such as IoU and Dice on the two datasets. This proves that BRNet is not a simple stacking of modules, but an organic whole: VLDAM provides high-quality features, VSC-PFM is responsible for scale integration, and VE-DUSM is responsible for detail restoration. The three work together to solve the three major challenges of background interference, structural fracture and edge blur in XCA vascular segmentation.

To further visually corroborate the conclusions drawn from the quantitative analysis mentioned above, Figure 3 presents the visual segmentation results of representative complex samples selected from the ARCADE dataset. The green, blue, and red regions represent true positive, false negative, and false positive pixels, respectively. From the figure, we can observe the specific roles played by each core module in addressing the specific challenges of XCA images. In the segmentation results of the baseline model, due to the low signal-to-noise ratio of the images, the pure Transformer architecture exhibits limitations in local feature extraction, manifested as numerous false positive background areas and missed detection of fine peripheral branches. After introducing VLDAM, false positive background noise is significantly reduced, and some main vessel features are restored. This indicates that the module can adaptively suppress background interference and enhance the response of local features. Further introduction of VSC-PFM reduces the breakpoints between main vessels and peripheral vessels, and improves the connectivity of fine branches. This suggests that multi-scale feature fusion effectively alleviates the problem of vascular topological discontinuity. Finally, after integrating VE-DUSM to form the complete BRNet, the edge smoothness of the vascular mask is improved, and the fine peripheral structures are further reconstructed. This module mitigates the edge blurring and upsampling distortion phenomena that are often caused by traditional interpolation methods. Overall, the gradual reduction in false positives and false negatives in the visual results objectively reflects the collaborative optimization process of the three modules in enhancing the connectivity and accuracy of vascular segmentation.

Furthermore, to verify the effectiveness of selecting the SimAM mechanism in the VLDAM, we conducted horizontal comparative experiments on different attention mechanisms within the VLDAM based on the complete model. As shown in Table 6, compared to the baseline without an attention mechanism, although SE [38] and CBAM [39] bring certain gains, they still struggle to effectively distinguish weak blood vessels from similar soft tissues under the low-contrast background of XCA images. In contrast, SimAM exhibits significant advantages; the model not only achieves optimal performance in core metrics such as IoU and Dice coefficient, but its Sensitivity also jumps to 0.7431. This demonstrates that SimAM’s adaptive weighting mechanism, based on the physical energy function, can fundamentally and more acutely capture and enhance low-contrast fine vascular features. Simultaneously, it effectively suppresses complex background interference, thereby minimizing the misdetection of small branches to the greatest extent.

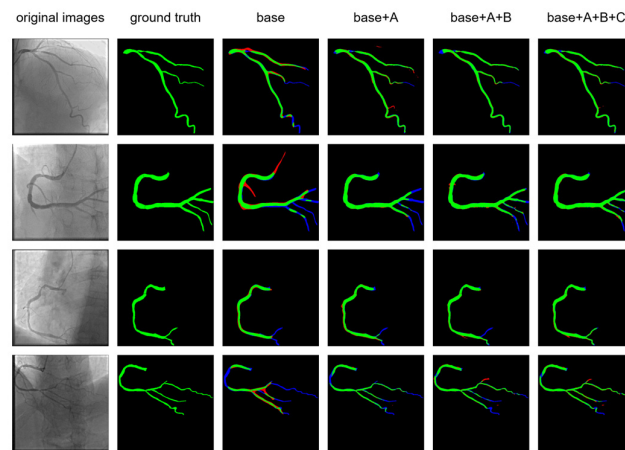


Figure 3. ARCADE Dataset Ablation Visualization Chart. The green, blue, and red areas represent true positive, false negative, and false positive pixels, respectively.

Table 6. Comparison of experimental performance of different attention mechanisms in VLDAM.

Attention	IoU	Dice	Pre	Acc	Sen	Spe	AUPRC	MCC
None	0.6056	0.7500	0.8141	0.9847	0.7016	0.9945	0.8018	0.7463
Se [38]	0.6133	0.7575	0.8164	0.9850	0.7124	0.9944	0.8054	0.7534
CBAM [39]	0.6090	0.7538	0.7989	0.9845	0.7184	0.9937	0.7951	0.7484
SimAM [32]	0.6305	0.7697	0.8040	0.9855	0.7431	0.9938	0.8145	0.7643

Note: Bold values indicate the best results.

3.2. Comparative Experiment

In order to prove the effectiveness of the BRNet proposed in this paper, the method in this paper is compared with Unet, Unet++, TransUNet, UTNet, SwinUnet, Mask2Former, VM-UNet, BiFormer, and EAG-SAE-ESTF on the public dataset ARCADE and the ZZ-CAHDS dataset constructed in this paper. In order to ensure the comparability of the experiments, the above comparison algorithms refer to the author's publicly available code and are retrained and tested in the same experimental environment as the model in this paper to avoid deviations caused by environmental differences. A total of 100 epochs are trained, and the algorithm performance is compared on the test set.

As shown in Table 7, the proposed BRNet achieves the best comprehensive segmentation performance on the ARCADE dataset. Compared with the baseline BiFormer, BRNet improves IoU and Dice by 10.12% and 8.06%, respectively. Compared with the second-best model EAG-SAE-ESTF, BRNet also achieves higher IoU and Dice, with improvements of 2.64% and 2.46%, respectively. These results indicate that BRNet is effective in improving vessel completeness and overall segmentation quality. Meanwhile, BRNet achieves the highest sensitivity among all compared models, suggesting that it has a stronger ability to detect subtle and low-contrast vascular structures. However, its precision and specificity are slightly lower than those of some competing models, indicating that a small number of ambiguous non-vascular pixels may be classified as vessels. This phenomenon is related to the design of the final BRNet architecture. The dual-backbone feature representation, multi-scale fusion, and dynamic upsampling modules are designed to enhance weak vascular responses and recover fine vascular branches. Compared with simpler models that tend to generate more conservative predictions, BRNet is more sensitive to boundary-blurred and low-contrast vessel-like patterns. Therefore, while the final architecture improves the detection of fine vessels and reduces missed branches, it may also enhance some background structures or artifact-like regions with similar intensity patterns, leading to slightly lower precision and specificity. Nevertheless, the higher IoU, Dice, and sensitivity indicate

that BRNet achieves stronger overall segmentation performance and better vessel integrity on ARCADE.

Table 7. Comparison of performance metrics of different models on the ARCADE dataset.

Model	IoU	Dice	Pre	Acc	Sen	Spe	AUPRC	Mcc
Unet [16]	0.5653 ± 0.0937 *	0.7176 ± 0.0795 *	0.8053 ± 0.0808	0.9832 ± 0.0047 *	0.6544 ± 0.1022 *	0.9946 ± 0.0024	0.7338 ± 0.1093 *	0.7154 ± 0.0782 *
Unet++ [17]	0.5832 ± 0.0886 *	0.7327 ± 0.0730 *	0.7707 ± 0.0842 *	0.9831 ± 0.0049 *	0.7040 ± 0.0880 *	0.9928 ± 0.0031 *	0.7351 ± 0.1086 *	0.7265 ± 0.0742 *
TransUNet [22]	0.5637 ± 0.0979 *	0.7157 ± 0.0849 *	0.8084 ± 0.0875	0.9832 ± 0.0050 *	0.6513 ± 0.1103 *	0.9947 ± 0.0027	0.6953 ± 0.1142 *	0.7145 ± 0.0830 *
UTNet [40]	0.5730 ± 0.0934 *	0.7239 ± 0.0782 *	0.7583 ± 0.0884 *	0.9826 ± 0.0049 *	0.6982 ± 0.0914 *	0.9924 ± 0.0031 *	0.7534 ± 0.1085 *	0.7173 ± 0.0793 *
SwinUnet [23]	0.5593 ± 0.1443 *	0.7056 ± 0.1291 *	0.7569 ± 0.1495 *	0.9820 ± 0.0094 *	0.6811 ± 0.1483 *	0.9927 ± 0.0054 *	0.7817 ± 0.1541 *	0.7034 ± 0.1227 *
Mask2Former [29]	0.5855 ± 0.0838 *	0.7350 ± 0.0691 *	0.7557 ± 0.0840 *	0.9830 ± 0.0046 *	0.7208 ± 0.0810 *	0.9920 ± 0.0030 *	0.7740 ± 0.0961 *	0.7279 ± 0.0706 *
VM-UNet [41]	0.5916 ± 0.0782 *	0.7403 ± 0.0635 *	0.7819 ± 0.0822 *	0.9837 ± 0.0043 *	0.7061 ± 0.0637 *	0.9933 ± 0.0028 *	0.7467 ± 0.0912 *	0.7339 ± 0.0651 *
BiFormer [24]	0.5293 ± 0.0742 *	0.6891 ± 0.0660 *	0.6994 ± 0.0859 *	0.9797 ± 0.0049 *	0.6839 ± 0.0675 *	0.9899 ± 0.0035 *	0.7276 ± 0.0933 *	0.6800 ± 0.0673 *
EAG-SAE-ESTF [27]	0.6041 ± 0.0879 *	0.7451 ± 0.0711 *	0.8111 ± 0.0937	0.9845 ± 0.0047 *	0.7018 ± 0.0776 *	0.9943 ± 0.0033 *	0.8136 ± 0.0933	0.7451 ± 0.0728 *
ours	0.6305 ± 0.0876	0.7697 ± 0.0689	0.8040 ± 0.0776	0.9855 ± 0.0042	0.7431 ± 0.0842	0.9938 ± 0.0027	0.8145 ± 0.0911	0.7643 ± 0.0699

Note: Bold values indicate the best results. * denotes BRNet statistical significance compared with other competing methods using the paired *t*-test ($p < 0.05$). Data are presented as mean ± standard deviation.

As shown in Table 8, BRNet also achieved strong comprehensive performance on the ZZ-CAHDS dataset. Compared with the baseline BiFormer, BRNet significantly improved IoU and Dice by 17.13% and 13.64%, respectively, demonstrating strong feature extraction ability under heterogeneous data conditions. Compared with the runner-up model EAG-SAE-ESTF, BRNet achieved higher IoU, Dice, precision, specificity, accuracy, and MCC, indicating that it produced cleaner vessel masks with better false-positive control on this heterogeneous benchmark. We note that EAG-SAE-ESTF obtained slightly higher sensitivity and AUPRC than BRNet, suggesting that it has a stronger tendency to identify potential vessel pixels and achieves slightly better precision-recall curve performance. However, this recall-oriented behavior may also be accompanied by more scattered or block-like false-positive regions in some cases. In contrast, BRNet achieved a more balanced performance between vessel detection and false-positive suppression on ZZ-CAHDS, while maintaining better boundary consistency and vascular structural coherence. Overall, although BRNet does not achieve the highest sensitivity or AUPRC, it maintains competitive precision-recall performance while achieving the best IoU and Dice scores on this heterogeneous benchmark.

Table 8. Comparison of performance metrics of different models on the ZZ-CAHDS dataset.

Model	IoU	Dice	Pre	Acc	Sen	Spe	AUPRC	Mcc
Unet [16]	0.6340 ± 0.0310 *	0.7756 ± 0.0229 *	0.7565 ± 0.0461 *	0.9801 ± 0.0050 *	0.7999 ± 0.0461	0.9884 ± 0.0036 *	0.8324 ± 0.0293 *	0.7666 ± 0.0238 *
Unet++ [17]	0.6364 ± 0.0349 *	0.7772 ± 0.0264 *	0.7703 ± 0.0352 *	0.9806 ± 0.0047 *	0.7885 ± 0.0590 *	0.9896 ± 0.0021 *	0.8116 ± 0.0441 *	0.7683 ± 0.0275 *
TransUNet [22]	0.6477 ± 0.0364 *	0.7856 ± 0.0266 *	0.7737 ± 0.0537 *	0.9811 ± 0.0047 *	0.8033 ± 0.0512	0.9896 ± 0.0029 *	0.7842 ± 0.0472 *	0.7774 ± 0.0276 *
UTNet [40]	0.5078 ± 0.1485 *	0.6592 ± 0.1490 *	0.6859 ± 0.1641 *	0.9693 ± 0.0208 *	0.6486 ± 0.1488 *	0.9843 ± 0.0161 *	0.7139 ± 0.1338 *	0.6477 ± 0.1547 *
SwinUnet [23]	0.4469 ± 0.0885 *	0.6124 ± 0.0902 *	0.5990 ± 0.1388 *	0.9654 ± 0.0116 *	0.6644 ± 0.1264 *	0.9789 ± 0.0118 *	0.6650 ± 0.1170 *	0.6048 ± 0.0855 *
Mask2Former [29]	0.6454 ± 0.0477 *	0.7834 ± 0.0367 *	0.7866 ± 0.0356 *	0.9815 ± 0.0046 *	0.7837 ± 0.0619 *	0.9905 ± 0.0026 *	0.8587 ± 0.0382 *	0.7747 ± 0.0365 *
VM-UNet [41]	0.6572 ± 0.0344 *	0.7926 ± 0.0250 *	0.7834 ± 0.0293 *	0.9817 ± 0.0053 *	0.8039 ± 0.0420	0.9898 ± 0.0034 *	0.8528 ± 0.0282 *	0.7836 ± 0.0266 *
BiFormer [24]	0.5822 ± 0.0304 *	0.7355 ± 0.0246 *	0.7108 ± 0.0413 *	0.9765 ± 0.0053 *	0.7671 ± 0.0544 *	0.9861 ± 0.0037 *	0.8057 ± 0.0369 *	0.7251 ± 0.0236 *
EAG-SAE-ESTF [27]	0.6732 ± 0.0271 *	0.8044 ± 0.0193	0.7845 ± 0.0344 *	0.9827 ± 0.0043 *	0.8280 ± 0.0403	0.9899 ± 0.0025 *	0.8750 ± 0.0302	0.7963 ± 0.0201
ours	0.6774 ± 0.0373	0.8071 ± 0.0268	0.8159 ± 0.0389	0.9834 ± 0.0047	0.8014 ± 0.0461	0.9917 ± 0.0032	0.8678 ± 0.0451	0.7992 ± 0.0270

Note: Bold values indicate the best results. * denotes BRNet statistical significance compared with other competing methods using the paired *t*-test ($p < 0.05$). Data are presented as mean ± standard deviation.

Furthermore, we evaluated the model complexity and performance in terms of the centerline Dice coefficient (cIDice) [42], the number of parameters (Params), floating-point operations (FLOPs), frames per second (FPS), inference latency, and peak GPU memory consumption. As shown in Table 9, models such as U-Net and UTRNet incur high computational overheads, reaching 223.94 and 325.48 GFLOPs, respectively, which may be unfavorable for deployment on resource-constrained clinical devices. Meanwhile, although the emerging VM-UNet reduces the computational load to 30.25 GFLOPs, its parameter count increases to 44.27 M due to its complex symmetric decoder, and its inference speed is relatively low at 30.64 FPS, with a peak GPU memory consumption of 0.63 GB. In contrast, the proposed BRNet strikes a favorable balance, achieving the highest cIDice scores on both datasets while maintaining a moderate computational cost of 33.88 GFLOPs, 35.71 M parameters, 101.57 FPS, 9.85 ms/frame inference latency, and only 0.25 GB peak GPU memory consumption. Although BRNet is not the lightest model in terms of raw computational complexity, its advantages in segmentation accuracy, topological coherence preservation, inference efficiency, and memory consumption make this computational cost acceptable. Therefore, from the perspective of overall performance, the proposed method holds practical value for XCA coronary artery segmentation tasks.

Table 9. Comparison of computational complexity and segmentation performance among different models.

Model	cIDice (ARCADE)	cIDice (ZZ-CAHDS)	FLOPs (G)	Params (M)	FPS (frames/s)	Peak GPU Memory (GB)	Inference Latency (ms/frame)
Unet [16]	0.7101 ± 0.0896 *	0.7881 ± 0.0448 *	223.94	31.39	103.17	0.28	9.69
Unet++ [17]	0.7214 ± 0.0834 *	0.7934 ± 0.0466 *	554.64	36.63	38.76	0.51	25.8
TransUNet [22]	0.7044 ± 0.1056 *	0.8139 ± 0.0441 *	154.64	98.37	36.13	0.81	27.68
UTNet [40]	0.7119 ± 0.0938 *	0.6777 ± 0.1320 *	325.48	36.63	43.56	1.04	22.96
SwinUnet [23]	0.6925 ± 0.1390 *	0.6030 ± 0.1048 *	46.8	41.38	73.35	0.46	13.63
Mask2Former [29]	0.7308 ± 0.0778 *	0.8075 ± 0.0619 *	56.39	37.69	31.38	0.29	31.87
VM-UNet [41]	0.7406 ± 0.0750 *	0.8136 ± 0.0419 *	30.25	44.27	30.64	0.63	32.64
BiFormer [24]	0.6787 ± 0.0715 *	0.7510 ± 0.0557 *	19.76	22.40	131.65	0.17	7.60
EAG-SAE-ESTF [27]	0.7423 ± 0.0830 *	0.8277 ± 0.0406 *	196.20	53.34	73.80	0.87	13.55
ours	0.7655 ± 0.0854	0.8355 ± 0.0521	33.88	35.71	101.57	0.25	9.85

Note: Bold values indicate the best results. * denotes BRNet statistical significance compared with other competing methods using the paired *t*-test ($p < 0.05$). Data are presented as mean ± standard deviation. FPS, inference latency, and peak GPU memory were measured with an input size of 512×512 and batch size of 1 on the same GPU platform.

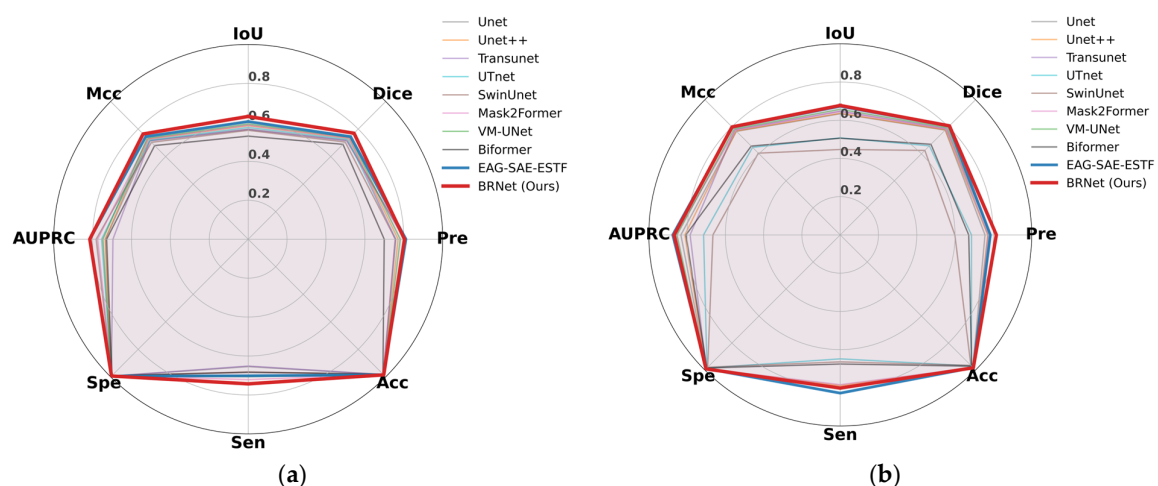
To further strengthen the statistical reliability of the reported improvements, we additionally calculated 95% confidence intervals (95% CIs) for the main topology-related metric, cIDice. Since all models were evaluated on the same test images, paired *t*-tests were used to compare BRNet with the corresponding competing methods based on per-image cIDice values. In addition, we specifically compared BRNet with the strongest competing method, EAG-SAE-ESTF, which achieved the highest cIDice among the non-BRNet models on both ARCADE and ZZ-CAHDS. As shown in Table 10, BRNet achieved higher mean cIDice scores than EAG-SAE-ESTF on both datasets. On ARCADE, BRNet improved the cIDice from 0.7423 to 0.7655, with a 95% CI of [0.7558, 0.7752]. On ZZ-CAHDS, BRNet improved the cIDice from 0.8277 to 0.8355, with a 95% CI of [0.8207, 0.8503]. These additional statistical results indicate that the performance gains of BRNet are not only reflected in the mean values but are also supported by uncertainty estimation and paired statistical testing.

The radar chart of evaluation indexes of each model on the two datasets is shown in Figure 4. BRNet exhibits the largest overall enclosed area on both datasets, intuitively reflecting its superior comprehensive segmentation performance. The model performs outstandingly on key metrics such as IoU and Dice coefficient, demonstrating its low missed detection rate and excellent segmentation integrity.

Table 10. Statistical comparison between BRNet and the strongest competing method in terms of cIDice.

Dataset	Metric	Strongest Competing Method	Strongest Method Mean $\pm$ SD	95% CI	BRNet Mean $\pm$ SD	95% CI	Improvement	Significance
ARCADE	cIDice	EAG-SAE-ESTF	0.7423 $\pm$ 0.0830	[0.7329, 0.7517]	0.7655 $\pm$ 0.0854	[0.7558, 0.7752]	+0.0232	$p < 0.05$
ZZ-CAHDS	cIDice	EAG-SAE-ESTF	0.8277 $\pm$ 0.0406	[0.8162, 0.8392]	0.8355 $\pm$ 0.0521	[0.8207, 0.8503]	+0.0078	$p < 0.05$

Note: The strongest competing method refers to the non-BRNet model with the highest mean cIDice on each dataset. Data are presented as mean $\pm$ standard deviation. The 95% CI was calculated using the corresponding test-set size, with $n = 300$ for ARCADE and $n = 50$ for ZZ-CAHDS. Statistical significance was assessed using paired t -tests on per-image results.

**Figure 4.** Radar charts of various models on different datasets. (a) Radar Chart on the ARCADE Dataset. (b) Radar Chart on the ZZ-CAHDS Dataset.

To further illustrate the effectiveness of the proposed BRNet, this paper presents the visualization segmentation results on the two datasets of ARCADE and ZZ-CAHDS in Figure 5, where the green, blue, and red areas represent true positive, false negative, and false positive pixels, respectively. It can be clearly observed from the figure that BRNet shows the best segmentation integrity and robustness under different data distributions. In the ARCADE dataset, compared with the obvious fractures at the end of blood vessels in Unet and TransUNet, BRNet can completely capture the subtle branch structure and effectively repair the topological discontinuity problem existing in the baseline model BiFormer. This advantage is particularly significant in the more complex ZZ-CAHDS dataset. Facing the intricate blood vessel tree structure and low-contrast images, models such as SwinUnet and BiFormer suffer from severe false-positive red background noise and large areas of blue structure loss. In contrast, BRNet demonstrates good anti-interference and feature extraction capabilities. It not only maintains a high degree of green connectivity at complex blood vessel intersections but also effectively suppresses background misjudgments. Compared with the suboptimal model EAG-SAE-ESTF, the mask boundary generated by BRNet is clearer, and there is less stray noise. This suggests that the model maintains a favorable balance between high sensitivity and structural accuracy, thereby contributing to its competitive performance on comprehensive indicators such as IoU and Dice.

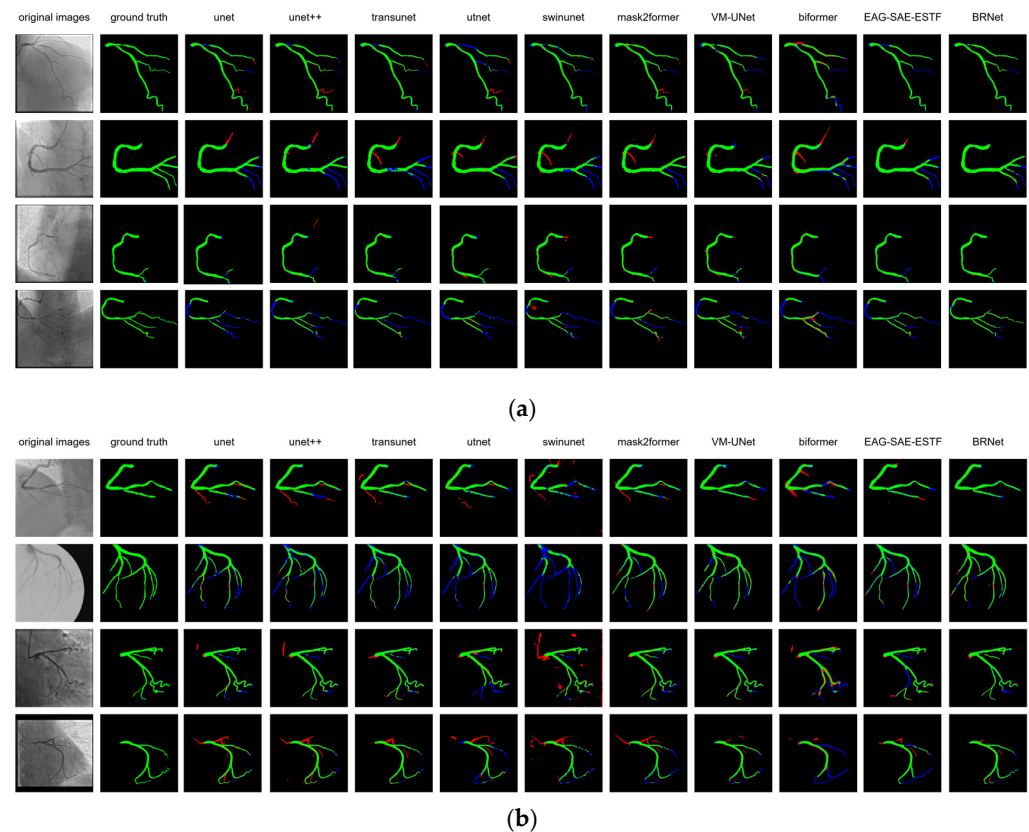


Figure 5. Visualization results of various models on different datasets (a) Segmentation results under the ARCADE dataset. (b) Segmentation results under the ZZ-CAHDS dataset. In (b), the first to fourth rows correspond to representative samples from ARCADE, DCA1, XCAD, and CHUAC, respectively. The green, blue, and red areas represent true positive, false negative, and false positive pixels, respectively.

4. Discussion

This study aims to solve the three major challenges of background noise interference, feature fragmentation, and edge blur in X-ray coronary angiography image segmentation by constructing the BRNet framework. Based on the experimental evidence of ARCADE and self-built ZZ-CAHDS datasets, this section will deeply explore the adaptation mechanism of the model on heterogeneous data, the effective mechanism of each module, and the limitations.

4.1. Robustness of Models and Analysis of Inductive Bias Under Heterogeneous Data

The experimental results reveal a noteworthy phenomenon: BRNet and some CNN-based models perform better on the more heterogeneous ZZ-CAHDS dataset than on the standardized ARCADE dataset. This phenomenon may be partly related to the annotation characteristics of the two datasets. Specifically, the ARCADE dataset does not annotate some very fine vascular structures, which may cause correctly detected fine vessels to be counted as false positives, thereby limiting the IoU score. In contrast, the annotations in ZZ-CAHDS are more consistent with visible vascular structures, which may better reflect the model's ability to capture fine vessels. However, the adaptability of different architectures to heterogeneous data varies significantly: pure Transformer architectures such as SwinUnet are difficult to converge on ZZ-CAHDS with small samples and high heterogeneity due to the lack of inherent inductive bias of convolution, while the hybrid architecture UTNet also shows a decline in performance; the reason is that the shallow attention mechanism is introduced too early, and the robust feature anchoring provided

by the deep pure CNN backbone is lacking when facing significant domain offset, which makes the model susceptible to heterogeneous background noise interference. In contrast, BRNet relies on the dual-backbone collaborative mechanism of ResNet local perception and BiFormer global modeling. Specifically, the ResNet branch uses the inherent inductive bias of convolution as a stable anchor for feature extraction to ensure the robust capture of high-frequency details such as blood vessel edges under small samples; the BiFormer branch introduces the bi-level routing attention with sparse perception characteristics, which can automatically filter out large areas of irrelevant backgrounds in XCA images, such as lung fields, myocardium, etc., and only perform long-range modeling on sparse vascular topology. This complementary design, where the CNN is responsible for local robustness and the sparse Transformer is responsible for global topology, is the fundamental reason why BRNet outperforms both full convolution and full Transformer architectures.

4.2. Mechanism Effectiveness Analysis of Core Modules

Ablation experiments have confirmed the necessity of the targeted module design, and the underlying mechanism of action can be summarized as follows: First, the VLDAM provides the most critical anti-noise capability. The essential challenge of XCA images is that vascular features are often submerged in the myocardial background or electronic noise. Different from the traditional attention mechanism that only relies on statistical feature weighting, the SimAM introduced by VLDAM is based on neuroscience theory and distinguishes the foreground from the background by evaluating the energy difference in feature points. This mechanism based on physical energy enables the model to adaptively suppress the high-energy background noise response, thereby solving the problem of not being able to see clearly at the source of feature extraction. Second, the VSC-PFM module solves the topological fracture caused by the semantic gap. Traditional FPN often directly integrates deep and shallow features, which can easily lead to semantic ambiguity, while the dense convolutional interaction in VSC-PFM acts as a semantic calibrator, ensuring that deep topological information can accurately guide the recovery of shallow terminal details, thereby maintaining the connectivity of the vascular tree. Finally, the VE-DUSM module solves the edge distortion from the perspective of signal processing. The coronary artery presents a gray gradient band with blurred edges under X-ray. The traditional bilinear interpolation is essentially a low-pass filter, which will further blur the edges. VE-DUSM uses the CARAFE operator to realize content-aware feature reorganization, dynamically generates an upsampling kernel according to the gradient distribution of local pixels, introduces edge-preserving characteristics in the decoding stage, and realizes pixel-level edge sharpening.

Furthermore, the interaction among VLDAM, VSC-PFM, and VE-DUSM was further analyzed based on the ablation results. VLDAM produces the largest individual improvement, indicating that local detail enhancement and background suppression are the dominant factors for improving the baseline representation. On this basis, VSC-PFM provides additional gains by improving cross-scale feature coherence and reducing vascular discontinuities between the main vessels and peripheral branches. Although VE-DUSM shows a relatively smaller individual gain, it plays a complementary role when combined with VLDAM and VSC-PFM by refining vessel boundaries and recovering fine terminal structures. Therefore, the improvement of BRNet is not caused by a single dominant component alone or by a simple stacking of modules. Instead, VLDAM provides the main representation enhancement, while VSC-PFM and VE-DUSM further provide complementary benefits in structural continuity and boundary reconstruction. The three modules work collaboratively to address background interference, vascular discontinuity, and edge blurring in XCA coronary artery segmentation.

4.3. Limitations and Future Directions

BRNet demonstrates strong comprehensive performance compared with existing mainstream methods. As shown in the radar chart in Figure 4, BRNet covers a large area across indicators such as IoU and sensitivity, indicating its advantages in segmentation completeness and vessel detection. However, its precision is slightly lower than that of some competing models, suggesting that a small number of non-vascular regions may still be incorrectly classified as vessels. This result reflects a sensitivity–precision trade-off rather than an entirely one-sided improvement. From the perspective of coronary artery segmentation, higher sensitivity is useful for reducing missed detections of fine or low-contrast vascular branches and for maintaining the structural completeness of the vascular tree. However, false-positive predictions, such as background noise or artifact-like regions being misclassified as vessels, may increase the workload of expert review and reduce confidence in computer-assisted analysis. Therefore, although BRNet shows strong vessel detection capability, the remaining false-positive issue should be acknowledged as a limitation.

To provide a more balanced evaluation of the robustness limits of BRNet, we further analyze representative failure cases under complex imaging conditions, as illustrated in Figure 6. In Figure 6a, high-contrast line-like structures, such as suture-wire-like or catheter-like artifacts, appear with elongated tubular profiles and strong intensity responses. These structures may partially resemble contrast-filled coronary arteries in 2D XCA images, causing BRNet to occasionally misclassify non-vascular regions as vessels, as indicated by the red false-positive areas. Meanwhile, the strong responses from these artifacts may interfere with local feature enhancement and reduce the sensitivity to nearby weak vascular branches, leading to missed detections shown by the blue regions. Figure 6b presents another challenging case involving severe vessel overlap and extremely low-contrast background interference. In this situation, overlapping vascular projections and weak boundary contrast make it difficult to distinguish true vessel continuity from surrounding anatomical structures. As a result, BRNet may produce local connectivity errors and miss fine peripheral branches, especially in regions where small vessels are faint or partially obscured. These failure cases indicate that although BRNet improves overall vascular continuity and boundary reconstruction, it still has limitations when dealing with strong artifact interference, severe projection overlap, and extremely low-contrast vessel regions.

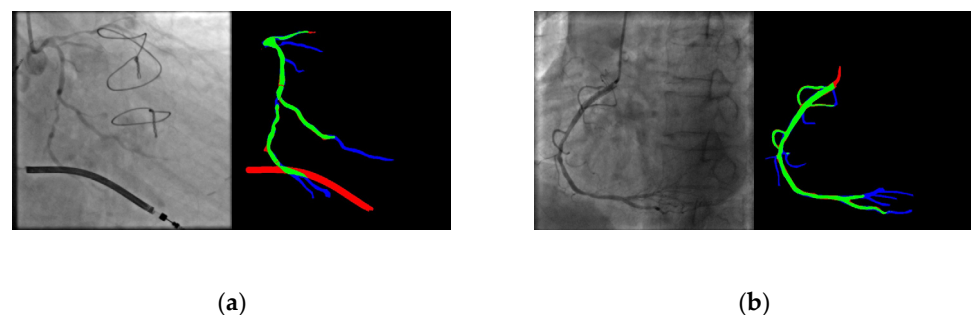


Figure 6. Representative failure cases. (a) Failure case with metallic artifact interference. (b) Failure case with severe vessel overlap and low-contrast background. The green, blue, and red areas represent true positive, false negative, and false positive pixels, respectively.

In view of the above limitations, future research will focus on the following directions. First, temporal consistency constraints will be introduced. Since XCA is essentially a dynamic angiographic video sequence, coronary vessels move periodically with the cardiac cycle, whereas some high-contrast artifact-like structures, such as suture-wire-like or catheter-related regions, may remain relatively static or exhibit motion patterns different

from true vessels. Therefore, future work will no longer be limited to single-frame image segmentation. Instead, we will explore temporal attention mechanisms and multi-frame consistency modeling to exploit motion differences across adjacent frames, thereby distinguishing true vascular structures from static or weakly moving artifacts. This strategy is expected to reduce false-positive predictions caused by metal-like interference. Second, low-contrast enhancement and overlap-aware vascular modeling will be further investigated. In cases with severe vessel overlap and weak boundary contrast, the 2D projection of coronary angiography may cause different vascular branches to visually intersect or become partially obscured, making it difficult for the model to distinguish true vessel continuity from projection-induced overlap. Future work will explore contrast-adaptive feature enhancement, topology-aware loss functions, and branch-level structural constraints to improve the discrimination of faint vessels and preserve the continuity of small peripheral branches. In addition, multi-frame information from angiographic sequences will be incorporated to exploit the temporal displacement of vessels during the cardiac cycle, which may help separate overlapped vascular structures and reduce missed detections in low-contrast regions. In addition to model-level improvements, future work will also strengthen the experimental validation protocol. Specifically, we will conduct 5-fold cross-validation or repeated random split validation on larger multi-center datasets to provide a more rigorous assessment of model generalization. These improvements are expected to further enhance the robustness of BRNet under challenging coronary angiography segmentation conditions.

5. Conclusions

In this paper, a dual-backbone collaborative segmentation network, BRNet, is proposed to address the challenges of background noise interference, multi-scale feature fragmentation, and edge blur in X-ray coronary angiography. The method constructs a complementary feature space of local perception and global modeling through VLDAM, suppresses noise using SimAM, and combines the progressive fusion of VSC-PFM with the dynamic upsampling strategy of VE-DUSM to maintain vascular topological coherence and edge restoration accuracy. Experiments on the ARCADE standard dataset and the proposed heterogeneous benchmark, ZZ-CAHDS, demonstrate that BRNet achieves competitive performance across key indicators such as IoU, Dice, and Sensitivity, along with improvements in the topology-aware cIDice. Particularly on the heterogeneous benchmark, it exhibits better robustness compared to state-of-the-art models, mitigating performance degradation caused by domain shift while maintaining an effective balance between segmentation precision and computational efficiency. Overall, this study validates the effectiveness of the targeted dual-backbone architecture in complex medical image analysis and suggests its potential value for computer-aided coronary artery segmentation and quantitative analysis. In future work, we will incorporate temporal consistency constraints, artifact-aware modeling, and multi-frame information from angiographic sequences to reduce false positives caused by static artifacts and improve segmentation robustness in low-contrast and severely overlapped vascular regions.

Author Contributions: Conceptualization, W.C.; methodology, Z.Z.; software, Z.Z.; validation, Z.Z.; investigation, H.S.; resources, W.C.; data curation, H.S.; writing—original draft preparation, Z.Z. and W.C.; writing—review and editing, H.S.; supervision, H.S. and W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study uses a publicly available dataset with fully de-identified imaging data; therefore, ethical review and approval were not required.

Informed Consent Statement: Not applicable. The study does not involve human participants or identifiable patient data.

Data Availability Statement: The images and corresponding annotations used in this study are obtained from publicly accessible datasets. Additionally, the source code and dataset are available in our repository at: <https://github.com/ZZ-626/BRnet> (accessed on 1 July 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CAD	Coronary Artery Disease
CTA	Computed Tomographic Angiography
XCA	X-ray Coronary Angiography
CNN	Convolutional Neural Network
ViT	Vision Transformer
VLDAM	Vascular Local Detail Attention Module
VSC-PFM	Vascular Structure-Coherent Progressive Fusion Module
VE-DUSM	Vascular Enhancement Dynamic Upsampling Module
CARAFE	Content-Aware ReAssembly of FEatures
IoU	Intersection over Union
MCC	Matthews Correlation Coefficient
AUC	Area Under the Curve
clDice	Centerline Dice Coefficient
FLOPs	Floating-Point Operations
FPS	Frames per second
AUPRC	Area under the precision-recall curve

References

- Gaidai, O.; Cao, Y.; Loginov, S. Global cardiovascular diseases death rate prediction. *Curr. Probl. Cardiol.* **2023**, *48*, 101622. [[CrossRef](#)] [[PubMed](#)]
- McCullough, P.A. Coronary artery disease. *Clin. J. Am. Soc. Nephrol.* **2007**, *2*, 611–616. [[CrossRef](#)] [[PubMed](#)]
- Knuuti, J.; Wijns, W.; Saraste, A.; Capodanno, D.; Barbato, E.; Funck-Brentano, C.; Prescott, E.; Storey, R.F.; Deaton, C.; Cuisset, T.; et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes: The Task Force for the diagnosis and management of chronic coronary syndromes of the European Society of Cardiology (ESC). *Eur. Heart J.* **2020**, *41*, 407–477. [[CrossRef](#)] [[PubMed](#)]
- Malakar, A.K.; Choudhury, D.; Halder, B.; Paul, P.; Uddin, A.; Chakraborty, S. A review on coronary artery disease, its risk factors, and therapeutics. *J. Cell. Physiol.* **2019**, *234*, 16812–16823. [[CrossRef](#)] [[PubMed](#)]
- Wardziak, Ł.; Kruk, M.; Pleban, W.; Demkow, M.; Rużyłło, W.; Dzielińska, Z.; Kępka, C. Coronary CTA enhanced with CTA based FFR analysis provides higher diagnostic value than invasive coronary angiography in patients with intermediate coronary stenosis. *J. Cardiovasc. Comput. Tomogr.* **2019**, *13*, 62–67. [[CrossRef](#)] [[PubMed](#)]
- Raja, J.; Seitz, M.P.; Yedlapati, N.; Khouzam, R.N. Can computed fractional flow reserve coronary ct angiography (ffrct) offer an accurate noninvasive comparison to invasive coronary angiography (ICA)? “The noninvasive cath.” A comprehensive review. *Curr. Probl. Cardiol.* **2021**, *46*, 100642. [[CrossRef](#)] [[PubMed](#)]
- Knaapen, P. Computed tomography to replace invasive coronary angiography? Close, but not close enough. *Circ. Cardiovasc. Imaging* **2019**, *12*, e008710. [[CrossRef](#)] [[PubMed](#)]
- Bolognese, L.; Reccia, M.R. Computed tomography to replace invasive coronary angiography? The DISCHARGE trial. *Eur. Heart J. Suppl.* **2022**, *24*, I25–I28. [[CrossRef](#)] [[PubMed](#)]
- Zahergivar, A.; Kocher, M.; Waltz, J.; Kabakus, I.; Chamberlin, J.; Akkaya, S.; Agha, A.M.; Schoepf, U.J.; Burt, J.R. The diagnostic value of non-contrast magnetic resonance coronary angiography in the assessment of coronary artery disease: A systematic review and meta-analysis. *Heliyon* **2021**, *7*, e06386. [[CrossRef](#)] [[PubMed](#)]
- Basha, M.A.A.; Aly, S.A.; Ismail, A.A.A.; Bahaeldin, H.A.; Shehata, S.M. The validity and applicability of CAD-RADS in the management of patients with coronary artery disease. *Insights Imaging* **2019**, *10*, 117. [[CrossRef](#)] [[PubMed](#)]

11. Li, Y.; Zhou, S.; Wu, J.; Ma, X.; Peng, K. A novel method of vessel segmentation for X-ray coronary angiography images. In *2012 Fourth International Conference on Computational and Information Sciences*; IEEE: Piscataway, NJ, USA, 2012; pp. 468–471. [\[CrossRef\]](#)
12. Vlachos, M.; Dermatas, E. Multi-scale retinal vessel segmentation using line tracking. *Comput. Med. Imaging Graph.* **2010**, *34*, 213–227. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Taghizadeh Dehkordi, M.; Doost Hoseini, A.M.; Sadri, S.; Soltanianzadeh, H. Local feature fitting active contour for segmenting vessels in angiograms. *IET Comput. Vis.* **2014**, *8*, 161–170. [\[CrossRef\]](#)
14. Eiho, S.; Sekiguchi, H.; Sugimoto, N.; Hanakawa, T.; Urayama, S. Branch-based region growing method for blood vessel segmentation. In *Proceedings of International Society for Photogram-metry and Remote Sensing Congress; Systems and Computers in Japan*; 2004; p. 801. Available online: <https://onlinelibrary.wiley.com/doi/10.1002/scj.20166> (accessed on 1 July 2026).
15. Paulinas, M.; Miniotas, D.; Meilūnas, M.; Ušinskas, A. An algorithm for segmentation of blood vessels in images. *Elektron. Elektrotech.* **2008**, *83*, 25–28. [\[CrossRef\]](#)
16. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241. [\[CrossRef\]](#)
17. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer International Publishing: Cham, Switzerland, 2018; pp. 3–11. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Zhao, C.; Vij, A.; Malhotra, S.; Tang, J.; Tang, H.; Pienta, D.; Xu, Z.; Zhou, W. Automatic extraction and stenosis evaluation of coronary arteries in invasive coronary angiograms. *Comput. Biol. Med.* **2021**, *136*, 104667. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Alirr, O.I. A Skip-Free Collaborative Residual U-Net for Secure Multi-Center Liver and Tumor Segmentation. *Eng* **2026**, *7*, 151. [\[CrossRef\]](#)
20. Li, Z.; Zhang, H.; Li, Z.; Ren, Z. Residual-Attention UNet++: A Nested Residual-Attention U-Net for Medical Image Segmentation. *Appl. Sci.* **2022**, *12*, 7149. [\[CrossRef\]](#)
21. Dosovitskiy, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929. [\[CrossRef\]](#)
22. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306. [\[CrossRef\]](#)
23. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. In *Computer Vision—ECCV 2022 Workshops*; Springer Nature: Cham, Switzerland, 2022; pp. 205–218. [\[CrossRef\]](#)
24. Zhu, L.; Wang, X.; Ke, Z.; Zhang, W.; Lau, R.W. Biformer: Vision transformer with bi-level routing attention. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2023; pp. 10323–10333. [\[CrossRef\]](#)
25. Guo, H.; Gao, R.; Ge, M.; Dong, W.; Liu, Y.; Zhao, X. Vessel segmentation method of coronary angiography image based on feature map pyramid. *Chin. Opt.* **2024**, *17*, 971–981. [\[CrossRef\]](#)
26. Xu, Y.; Zhai, N.; Ni, W.; Tian, Q.; Wang, J. Coronary angiography vessel segmentation based on deep convolution and multi-layer scale feature fusion. *Chin. J. Biomed. Eng.* **2025**, *44*, 34–42. [\[CrossRef\]](#)
27. Zhang, X.; Wang, S.; Lv, W.; Zhou, M.; Li, S.; Yang, J.; Huang, M.; Su, Y.; Shen, T. EAG-SAE-ESTF: An entropy-aware gated network with scale-adaptive enhancement and elastic spatial topology fusion for coronary artery segmentation. *Inf. Fusion* **2026**, *126*, 103597. [\[CrossRef\]](#)
28. Huang, B.; Luo, Y.; Wei, G.; He, S.; Shao, Y.; Zeng, X.; Zhang, Q. Deep learning model for coronary artery segmentation and quantitative stenosis detection in angiographic images. *Med. Phys.* **2025**, *52*, e17970. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Cheng, B.; Misra, I.; Schwing, A.G.; Kirillov, A.; Girdhar, R. Masked-attention mask transformer for universal image segmentation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2022; pp. 1290–1299. [\[CrossRef\]](#)
30. Ravi, N.; Gabeur, V.; Hu, Y.T.; Hu, R.; Ryali, C.; Ma, T.; Khedr, H.; Rädle, R.; Rolland, C.; Gustafson, L.; et al. Sam 2: Segment anything in images and videos. *arXiv* **2024**, arXiv:2408.00714. [\[CrossRef\]](#)
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778. [\[CrossRef\]](#)
32. Yang, L.; Zhang, R.Y.; Li, L.; Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In *Proceedings of the 38th International Conference on Machine Learning*; PMLR: Cambridge, MA, USA, 2021; pp. 11863–11874. Available online: <https://proceedings.mlr.press/v139/yang21o.html> (accessed on 1 July 2026).
33. Wang, J.; Chen, K.; Xu, R.; Liu, Z.; Loy, C.C.; Lin, D. Carafe: Content-aware reassembly of features. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: Piscataway, NJ, USA, 2019; pp. 3007–3016. [\[CrossRef\]](#)
34. Popov, M.; Amanturdieva, A.; Zhaksylyk, N.; Alkanov, A.; Saniyazbekov, A.; Aimyshev, T.; Ismailov, E.; Bulegenov, A.; Kuzhukeyev, A.; Kulanbayeva, A.; et al. Dataset for automatic region-based coronary artery disease diagnostics using X-ray angiography images. *Sci. Data* **2024**, *11*, 20. [\[CrossRef\]](#) [\[PubMed\]](#)

35. Cervantes-Sanchez, F.; Cruz-Aceves, I.; Hernandez-Aguirre, A.; Hernandez-Gonzalez, M.A.; Solorio-Meza, S.E. Automatic segmentation of coronary arteries in X-ray angiograms using multiscale analysis and artificial neural networks. *Appl. Sci.* **2019**, *9*, 5507. [\[CrossRef\]](#)
36. Ma, Y.; Hua, Y.; Deng, H.; Song, T.; Wang, H.; Xue, Z.; Cao, H.; Ma, R.; Guan, H. Self-supervised vessel segmentation via adversarial learning. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: Piscataway, NJ, USA, 2021; pp. 7536–7545. [\[CrossRef\]](#)
37. Carballal, A.; Novoa, F.J.; Fernandez-Lozano, C.; García-Guimaraes, M.; Aldama-López, G.; Calviño-Santos, R.; Vazquez-Rodriguez, J.M.; Pazos, A. Automatic multiscale vascular image segmentation algorithm for coronary angiography. *Biomed. Signal Process. Control* **2018**, *46*, 1–9. [\[CrossRef\]](#)
38. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Salt Lake City, UT, USA, 2018; pp. 7132–7141. [\[CrossRef\]](#)
39. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In *Computer Vision—ECCV 2018*; Springer International Publishing: Cham, Switzerland, 2018; pp. 3–19. [\[CrossRef\]](#)
40. Gao, Y.; Zhou, M.; Metaxas, D.N. UTNet: A hybrid transformer architecture for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021*; Springer International Publishing: Cham, Switzerland, 2021; pp. 61–71. [\[CrossRef\]](#)
41. Ruan, J.; Li, J.; Xiang, S. Vm-unet: Vision mamba unet for medical image segmentation. *ACM Trans. Multimed. Comput. Commun. Appl.* **2024**. [\[CrossRef\]](#)
42. Shit, S.; Paetzold, J.C.; Sekuboyina, A.; Ezhov, I.; Unger, A.; Zhylka, A.; Pluim, J.P.; Bauer, U.; Menze, B.H. clDice-a novel topology-preserving loss function for tubular structure segmentation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2021; pp. 16560–16569. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.