HCF-UNet: Selective Skip Fusion and Calibrated Decoding for Coronary Vessel Segmentation

Xiu Yang[1,2], Li Li[1,2,*], and Xiqianlong Yuan[1,2],and Guoqiang Zheng[1,2]
[1]Southwest University of Science and Technology, Mianyang, China
[2]Sichuan Engineering Technology Research Center of Industrial Self-Supporting and Artificial Intelligence, Mianyang, China
[*]Corresponding author: llki2000@163.com

***Abstract*—Coronary vessel segmentation is essential for vascular modeling and quantitative analysis, yet remains challenging in coronary angiography due to elongated morphology, large scale variation, and heavy background interference. Many U-Net-based methods employ plain upsampling and skip-connection concatenation, which often fails to reconcile multi-scale semantics with high-resolution details, resulting in fragmented thin branches and inaccurate boundaries. We propose an efficient encoder–decoder network, termed HCF-UNet, to improve fine-structure reconstruction through redesigned skip connections and an enhanced decoder. Specifically, we introduce a Vessel-Dimension-Aware Selective Integration (V-DASI) module that combines channel-wise selection with spatial attention to adaptively fuse cross-level features and reduce redundancy. Moreover, we design a Residual–Channel–Spatial (RCS) decoder, where stacked residual blocks strengthen reconstruction and dual attention progressively calibrates vessel responses, enhancing sensitivity to tiny branches and boundary regions. Experiments on the public XCAD dataset demonstrate that HCF-UNet consistently outperforms representative baselines, yielding clearer delineation of fine vessels and more coherent vascular connectivity. These results indicate that selective fusion and attention-enhanced decoding offer an effective yet efficient solution for coronary angiography vessel segmentation.**

***Index Terms*—coronary angiography, vessel segmentation, U-Net, skip connections**

I. Introduction

Vessel segmentation is a cornerstone of medical image analysis, aiming to delineate elongated structures from cluttered backgrounds while preserving their topological connectivity as faithfully as possible [1] . In coronary angiography, this task becomes even more demanding: vessels appear thin and highly branched, exhibit substantial scale variation, and are often obscured by imaging noise, low contrast, superimposed anatomical structures, and artifact contamination. Under these conditions, tiny branches and boundary regions are particularly prone to missed detections or erroneous adhesion.

In recent years, encoder–decoder architectures represented by U-Net have become the dominant paradigm for medical image segmentation. Nevertheless, when a standard U-Net is directly applied to coronary angiography, two major bottlenecks remain. First, skip connections are typically implemented via simple concatenation, yet low-level details and high-level semantics differ markedly in both semantic granularity and noise characteristics; fusing them in a naïve manner can amplify redundant background responses and, in turn, hinder reliable decoding. Second, many decoders rely on basic stacks of convolutions, whose reconstruction and calibration capacity is often insufficient when confronted with multi-scale fused features, leading to broken thin branches, missed distal segments, and blurred vessel boundaries.

To tackle these limitations, prior studies have incorporated attention mechanisms, pyramid structures, or contextual modeling to enrich feature representations. However, many of these designs emphasize the encoder side or global context, and their more elaborate architectures often introduce extra parameters and computation, which makes deployment less practical in resource-constrained settings. Therefore, improving the discriminative fusion in skip connections while strengthening fine-grained reconstruction in the decoder—with practical efficiency trade-off—remains a central challenge for advancing coronary vessel segmentation performance [1] .

Motivated by these observations, we develop an efficient encoder–decoder framework for coronary angiography vessel segmentation, termed **HCF-UNet**. Centered on the principles of *reducing redundant interference* and *strengthening responses to fine structures*, the network revisits both the skip-connection pathway and the decoder design, enabling more reliable multi-level fusion and more precise structural recovery. The main contributions are summarized as follows:

- We propose **V-DASI**, a vessel-dimension-aware selective skip-connection module that jointly leverages channel selection and spatial refinement to adaptively fuse features across semantic levels, mitigating semantic mismatch and redundant background disturbance.
- We design an **RCS** (Residual–Channel–Spatial) decoder that enhances reconstruction via residual learning and progressively calibrates features through channel and spatial attention, improving the delineation of thin branches and boundary regions.
- Extensive experiments on the **XCAD** dataset demonstrate that the proposed approach achieves a favorable balance between accuracy and efficiency.

Data and Code Availability Note: The source code and implementation framework of HCF-UNet are not publicly available due to proprietary restrictions and strict confidentiality agreements with the supporting research institution. Furthermore, the intellectual property associated with the core modules, specifically V-DASI and the RCS decoder, is currently undergoing institutional review for technical protection.

Interested researchers may contact the corresponding author for academic inquiries regarding the implementation details.

II. Related Work

To clarify the development of existing approaches, we review the related literature from three perspectives: vessel segmentation methods, encoder–decoder architectures and their refinements, and the use of attention mechanisms in image segmentation.

A. Vessel Segmentation Methods

As a key research topic in medical image analysis, vessel segmentation has attracted sustained attention across a wide range of applications, including retinal vasculature, cerebral vessels, and coronary arteries. Early studies largely relied on rule-based image processing techniques—such as thresholding, region growing, and morphological operations [2]. These methods typically require handcrafted features and are highly sensitive to image quality and parameter tuning, which makes them difficult to generalize to complex backgrounds and multi-scale vascular structures [3].

With the advancement of deep learning, convolutional neural networks have gradually become the mainstream approach for vascular segmentation [4]. Among these, the encoder-decoder architecture represented by U-Net has been widely applied to vascular segmentation tasks due to its excellent adaptability to pixel-level predictions [5]. Numerous studies have enhanced the segmentation performance of U-Net on complex vascular structures by incorporating multi-scale feature extraction, context modeling, or deep supervision mechanisms [6]. However, under challenging imaging conditions such as coronary angiography, where vessels are slender, densely branched, and exhibit low contrast against the background, existing methods still face limitations in modeling minute vascular branches and boundary regions [7].

B. Encoder–Decoder Architectures and Skip-Connection Refinements

Encoder–decoder architectures have become the central paradigm for modern image segmentation. The core idea is to progressively extract high-level semantic representations in the encoder, then restore spatial resolution in the decoder via upsampling, while leveraging skip connections to inject high-resolution information from the encoder into the reconstruction process. In the classic U-Net design, features from the encoder and decoder are directly concatenated, which effectively compensates for the spatial detail lost during downsampling.

Despite its simplicity and effectiveness, straightforward concatenation overlooks the fact that features from different stages can differ substantially in semantic abstraction and noise characteristics. Mixing them without discrimination often introduces redundant responses and amplifies interference during decoding. To mitigate this issue, some studies incorporate feature selection or weighting schemes into skip connections to make multi-level fusion more targeted, while others strengthen reconstruction by adopting deeper or more sophisticated decoder designs [8]. Although these strategies can improve segmentation accuracy to some extent, they frequently come with a noticeable increase in parameters and computational cost, which limits their practicality in resource-constrained deployments.

Therefore, designing more discriminative skip connections and a stronger decoder while keeping the overall architecture efficient remains an open and pressing challenge for encoder–decoder networks in complex vessel segmentation tasks.

C. Attention Mechanisms in Image Segmentation

Attention mechanisms endow models with explicit feature selection by adaptively reweighting responses, and they have been extensively used in image classification, object detection, and segmentation. In segmentation tasks, attention is commonly applied along the channel and/or spatial dimensions to recalibrate features, so that discriminative cues associated with the target region are emphasized.

Channel attention captures the relative importance across channels [9], which helps amplify task-relevant representations while suppressing redundant activations; spatial attention, in contrast, learns a spatial weighting map that guides the network toward key regions. For fine-structure modeling problems such as vessel segmentation, attention has been shown to strengthen responses to tiny targets and boundary areas, improving delineation quality [10].

Despite these benefits, most existing attention designs are inserted as standalone modules in the encoder or within the global representation, while the decoding stage—where feature reconstruction is crucial—often receives less targeted attention-oriented design. Moreover, introducing multiple attention components can noticeably increase model complexity, meaning the trade-off between accuracy gains and computational cost still leaves room for improvement.

D. Summary

In summary, although existing vessel segmentation methods have made steady progress in multi-scale feature fusion and fine-structure modeling, they still face notable challenges under complex imaging conditions. In particular, skip connections may introduce redundant information, decoders often lack sufficient reconstruction strength, and attention is not always applied in a way that directly supports fine-grained recovery. To address these limitations, we focus on both the skip-connection pathway and the decoder design, and propose an efficient yet more discriminative fusion-and-reconstruction strategy tailored to the structural complexity of coronary vessel segmentation.

III. Method

A. Overall Network Architecture

We propose an efficient encoder–decoder network for coronary vessel segmentation, termed HCF-UNet. As illustrated in Fig. 1, the network follows the classic U-Net paradigm, while

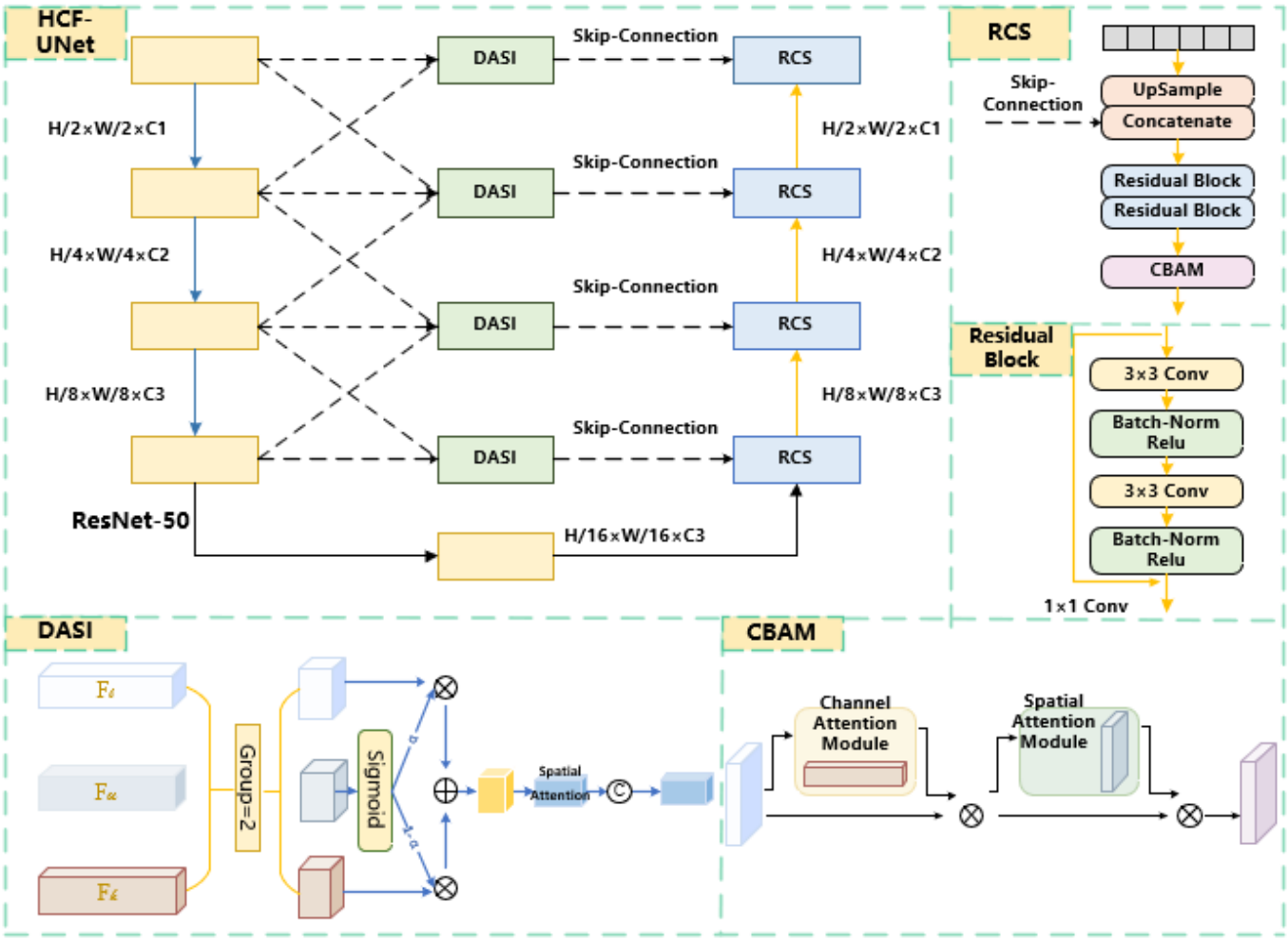

Fig. 1. Overall architecture of HCF-UNet. V-DASI is embedded into the skip connections for selective multi-level fusion, while the RCS decoder progressively refines vessel structures during reconstruction.

introducing dedicated modules in the skip-connection pathway and the decoder to better fuse and reconstruct features, thereby improving the representation of complex vascular structures.

In the encoder, multi-scale feature representations are extracted stage by stage through convolutional downsampling, so that vessel semantics at different scales can be captured. As the depth increases, the encoded features naturally evolve from local details toward higher-level semantics. Since direct skip connections tend to introduce semantic inconsistency and redundant responses, we insert the V-DASI module between the encoder and decoder to selectively integrate skip-connection features.

In the decoder, spatial resolution is progressively recovered via hierarchical upsampling, and the fused features are further refined by the improved RCS decoding module for deeper reconstruction and calibration. With this design, the decoder not only restores spatial details, but also leverages multi-scale semantics to more accurately recover thin vessel branches and boundary structures.

B. *Vessel-Dimension-Aware Selective Integration*

1) Design Motivation: To alleviate semantic misalignment across hierarchical features and suppress redundant background interference introduced by skip connections, we develop **V-DASI** (*Vessel-Dimension-Aware Selective Integration*). Guided by the feature representation at the current decoding scale, V-DASI aligns and selectively fuses neighboring low-resolution semantic features with high-resolution detail features. Given the current-scale feature F_u as well as its adjacent high-resolution feature F_h and low-resolution feature F_l, V-DASI first maps all three to a unified spatial resolution and then outputs an enhanced representation for subsequent decoding and reconstruction.

2) Module Structure: As illustrated in Fig. 2, V-DASI takes three inputs: the shallow encoder feature $F_l \in \mathbb{R}^{H_l \times W_l \times C_l}$, the deep encoder feature $F_h \in \mathbb{R}^{H_h \times W_h \times C_h}$, and the current decoder feature $F_u \in \mathbb{R}^{H \times W \times C}$. To facilitate effective fusion, the low-resolution branch F_l is first projected to the target channel dimension and then upsampled via bilinear interpolation, whereas the high-resolution branch F_h is downsampled

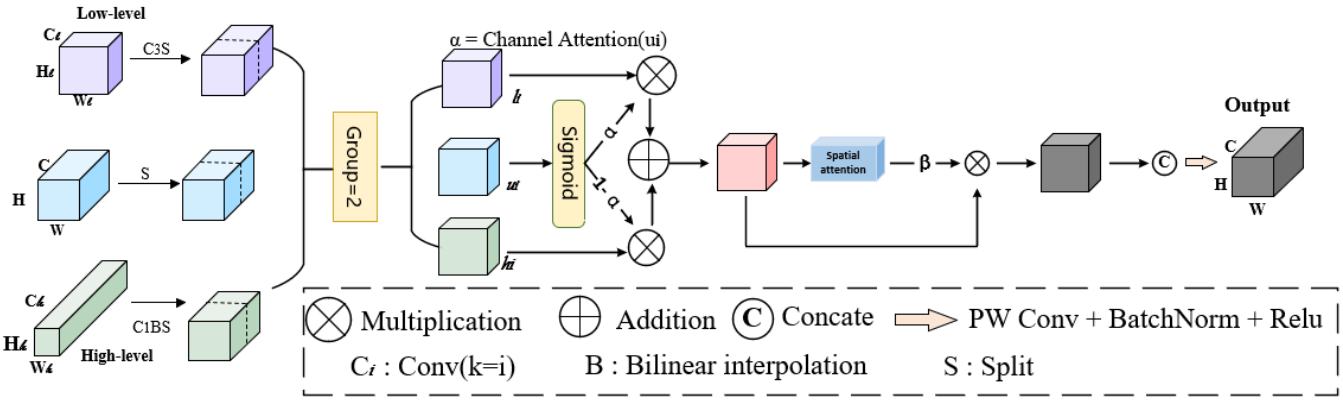

Fig. 2. Structure of the proposed V-DASI module.

with a learnable strided convolution for scale matching, followed by interpolation for precise spatial alignment. Finally, a 1×1 convolution is applied to map all three feature streams to the same channel dimension C.

After feature alignment, the three feature streams are split into two groups along the channel dimension. Taking the current decoder feature u as the reference, we employ a channel-attention mechanism to generate adaptive weights, which balance the contributions from shallow detail features and high-level semantic features. Specifically, for the i-th group, the channel-wise weight is computed from the current decoder feature u_i as

$$\alpha_i = \sigma(\mathrm{Conv}_2(\mathrm{ReLU}(\mathrm{Conv}_1(\mathrm{GAP}(u_i))))), \tag{1}$$

where $\mathrm{GAP}(\cdot)$ denotes global average pooling, Conv_1 and Conv_2 are shared channel-wise convolution layers, and $\sigma(\cdot)$ is the Sigmoid activation function. This weight reflects the extent to which the current decoding stage relies on semantic cues at different levels.

With α_i, the shallow feature l_i and the high-level feature h_i are then selectively fused as

$$u_i' = \alpha_i \odot l_i + (1 - \alpha_i) \odot h_i, \tag{2}$$

where $\odot$ denotes element-wise multiplication. In this manner, the fusion preserves spatial details while injecting the necessary high-level semantics, thereby mitigating semantic mismatch and suppressing redundant feature interference that often arises in conventional skip connections.

To further strengthen the responses along vessel trunks and tiny branch regions, we apply spatial-attention calibration to the fused feature u_i':

$$\beta = \sigma\Big(\mathrm{DWConv}_{7\times7}\Big([\mathrm{AvgPool}(u_i'), \mathrm{MaxPool}(u_i')]\Big)\Big), \tag{3}$$

$$u_i'' = \beta \odot u_i', \tag{4}$$

where $[\cdot,\cdot]$ denotes channel-wise concatenation, and $\mathrm{DWConv}_{7\times7}$ represents a depthwise separable convolution with a 7×7 kernel. This step suppresses spurious high responses in background regions while improving the spatial consistency of vessel structures.

a) Feature Reorganization and Output.: Finally, the spatially calibrated features $\{u_i''\}$ are concatenated along the channel dimension, and a lightweight **GhostConv + BN + ReLU** block is used for feature reorganization to produce the V-DASI output:

$$\hat{F}_u = \mathrm{ReLU}\big(\mathrm{BN}\big(\mathrm{GhostConv}\big([u_1'', u_2'']\big)\big)\big). \tag{5}$$

The resulting $\hat{F}_u$ serves as the skip-connection feature after selective integration and spatial calibration, and is fed into the subsequent decoding stages for feature reconstruction.

C. Residual–Channel–Spatial Decoder (RCS)

1) Design Motivation: The decoding stage is a critical component of encoder–decoder networks, where spatial resolution is progressively recovered and structural details are reconstructed. Conventional decoders typically follow a simple *upsample–concatenate–convolve* paradigm. Due to the limited nonlinear modeling capacity of such processing units, they often fail to fully exploit the complex relationships among multi-scale features, which can lead to missing thin vessel branches or blurred boundaries. This issue is particularly pronounced in fine-structure segmentation tasks such as coronary angiography.

To enhance reconstruction capability during decoding, we propose a **Residual–Channel–Spatial** decoding module (**RCS**). By jointly leveraging residual learning and dual attention mechanisms, RCS optimizes the decoding process in a coordinated manner. As a result, while spatial resolution is gradually restored, vessel-related features can be calibrated more finely and more stably.

2) Decoder Structure: As shown in Fig. 3, the RCS module first upsamples the feature map from the previous decoding stage by a factor of 2, and then fuses it with the encoder feature filtered by the V-DASI module via channel-wise concatenation, yielding the input feature of the decoding stage, denoted as F_{in}.

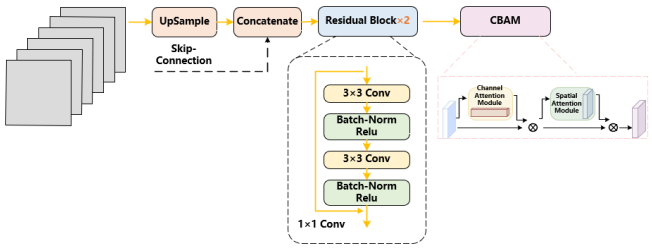

Fig. 3. Structure of the proposed RCS decoder. The decoder upsamples features, concatenates skip-connected features, applies two residual blocks, and refines representations with CBAM.

a) Residual Feature Enhancement.: The fused feature is subsequently transformed by two stacked residual blocks [11]. Each residual block adopts a shortcut connection (*Shortcut*) to mitigate potential gradient vanishing during deep decoding and to improve the stability of feature representation. The forward propagation can be expressed as

$$F_{\text{out}} = \text{ReLU}\Big(\text{ConvBlock}(F_{\text{in}}) + \text{Shortcut}(F_{\text{in}})\Big), \quad (6)$$

where $\text{ConvBlock}(\cdot)$ denotes a nonlinear transformation path composed of two 3×3 convolutions, batch normalization, and ReLU activations, and $\text{Shortcut}(\cdot)$ denotes either an identity mapping or a 1×1 convolution for dimension matching. By stacking residual blocks, the RCS module can more thoroughly integrate high-level semantic information with low-level spatial details, providing a more discriminative representation for subsequent attention calibration.

b) Channel and Spatial Attention Calibration.: After residual enhancement, the RCS module introduces a convolutional block attention module (CBAM) to recalibrate the features in a progressive manner. First, channel attention models the importance of different channels, formulated as

$$X' = X \odot \sigma\Big(\text{MLP}(\text{GAP}(X)) + \text{MLP}(\text{GMP}(X))\Big), \quad (7)$$

where X denotes the output feature of the residual blocks, $\text{GAP}(\cdot)$ and $\text{GMP}(\cdot)$ indicate global average pooling and global max pooling, respectively, $\text{MLP}(\cdot)$ is a shared multilayer perceptron, and $\sigma(\cdot)$ is the Sigmoid function. This mechanism adaptively strengthens discriminative channels that are highly relevant to vessel segmentation while suppressing redundant or irrelevant responses.

On top of channel recalibration, a spatial attention mechanism is further applied to refine the spatial distribution of features:

$$\text{CBAM}(X) = X' \odot \sigma\Big(f_{7\times7}\big([\text{GAP}_c(X'), \text{GMP}_c(X')]\big)\Big), \quad (8)$$

where $[\cdot]$ denotes feature concatenation, $\text{GAP}_c(\cdot)$ and $\text{GMP}_c(\cdot)$ represent average pooling and max pooling along the channel dimension, respectively, and $f_{7\times7}(\cdot)$ denotes a 7×7 convolution. Spatial attention highlights vessel trunks and boundary regions, thereby further improving responsiveness to tiny vessel branches.

Finally, the RCS module outputs the reconstructed feature after residual enhancement and dual-attention calibration, which is used in subsequent decoding stages and for the final segmentation prediction.

D. Discussion on Complexity

Although additional feature processing modules are introduced in the skip-connection pathway and the decoding stage, HCF-UNet still maintains a low overall model complexity. On the one hand, V-DASI adopts lightweight convolutions together with a grouping strategy, which avoids the parameter inflation caused by large-scale feature concatenation [12]. On the other hand, RCS improves feature representation through shared attention structures and residual connections, without noticeably increasing computational overhead. In this way, the proposed design enhances segmentation accuracy while preserving a favorable balance between performance and efficiency, making it well suited for practical deployments in resource-constrained scenarios.

IV. Experiments

A. Dataset and Implementation Details

To verify the effectiveness of the proposed method, we conducted experiments on the public coronary angiography vessel segmentation dataset **XCAD**. The XCAD images are single-channel with a resolution of 512×512. Considering the limited number of pixel-level annotated samples, we evaluated our method using **3-fold cross-validation** on 126 labeled images [13]. Specifically, all annotated data were randomly partitioned into three non-overlapping subsets. In each run, one fold was used as the test set and the remaining two folds were used for training. This procedure was repeated three times, and we finally report the results as *mean*$\pm$*std* over the three test folds.

Within each training run, we further split a portion of the training folds into a validation set, which was used only for model selection and early-stage hyperparameter tuning, while the test fold was strictly reserved for the final evaluation.

During training, we applied online data augmentation to improve generalization, including random rotations, horizontal flipping, and brightness/contrast perturbations. All experiments were implemented in PyTorch and trained on a single NVIDIA RTX 3090 GPU (24 GB). The input images were uniformly resized to 512×512. We used Adam [14] as the optimizer with an initial learning rate of 1×10^{-4} [15], and adopted cosine annealing for learning-rate decay. The batch size was set to 8, and the model was trained for 100 epochs.

Following [16], the training objective was defined as a weighted sum of the BCE loss and the Dice loss:

$$\mathcal{L}_{\text{total}} = \lambda_1\mathcal{L}_{\text{BCE}} + \lambda_2\mathcal{L}_{\text{Dice}}, \quad (9)$$

where the weighting factors were set to $\lambda_1 = 0.5$ and $\lambda_2 = 0.5$. During inference, the predicted probability map was binarized using a fixed threshold, and no additional morphological post-processing was applied.

TABLE I
PERFORMANCE COMPARISON ON XCAD (3-FOLD, MEAN±STD).

Method	**Dice↑**	**IoU↑**	**Sn.↑**	**Sp.↑**	**MCC↑**
U-Net	0.724±0.010	0.571±0.020	0.868±0.011	0.994±0.004	0.661±0.010
U-Net++	0.731±0.009	0.570±0.020	0.870±0.020	0.996±0.002	0.700±0.040
DeepLabv3+	0.730±0.020	0.570±0.020	0.860±0.020	0.995±0.002	0.690±0.030
CE-Net	0.740±0.020	0.580±0.030	0.870±0.010	0.995±0.003	0.700±0.040
R2U-Net	0.740±0.030	0.590±0.030	0.880±0.020	0.995±0.002	0.700±0.040
SE-RegUNet	0.740±0.030	0.590±0.040	0.890±0.020	0.995±0.002	0.730±0.050
TransUNet	0.750±0.020	0.600±0.030	0.890±0.010	0.994±0.004	0.740±0.040
Ours (HCF-UNet)	**0.790±0.020**	**0.650±0.010**	**0.920±0.010**	**0.996±0.003**	**0.780±0.040**

B. Compared Methods, Metrics, and Fair Comparison

We compared **HCF-UNet** with seven representative segmentation models, including U-Net [17], U-Net++ [18], DeepLabv3+ [19], CE-Net [20], R2U-Net [21], SE-RegUNet, and TransUNet [22]. These baselines cover a broad spectrum of design paradigms, ranging from classic encoder–decoder architectures and dense skip connections to multi-scale context modeling, residual recurrent enhancement, and Transformer-based global representation learning.

We adopted Dice [23], IoU, Sensitivity (Sn.), Specificity (Sp.), and MCC [24] as evaluation metrics, so as to assess performance from multiple aspects, including region overlap, detection capability, false-positive suppression, and robustness to class imbalance.

To ensure a fair comparison, all methods were trained from scratch on the same 3-fold splits, using identical input resolution, data augmentation, optimizer, learning-rate schedule, training epochs, and loss function. Moreover, we employed a consistent model selection strategy: for each fold, the best checkpoint was saved according to the validation Dice score, and the final results were obtained by evaluating the selected checkpoint on the corresponding test fold and reporting *mean*±*std* over the three folds.

C. Main Results

Table 1 summarizes the quantitative comparison results on XCAD under the 3-fold evaluation protocol. HCF-UNet achieves the best performance across the key metrics, including Dice, IoU, Sn., and MCC, indicating that the proposed selective fusion in skip connections and the decoder calibration mechanism effectively enhance fine-structure modeling capability and improve overall robustness in class-imbalanced settings.

As shown in Fig. 4, qualitative segmentation results of different methods on representative test samples. Overall, the baseline models are more prone to discontinuities or missed distal segments in low-contrast thin branches, often accompanied by background false positives. In contrast, HCF-UNet preserves the connectivity between vessel trunks and fine branches more reliably and produces sharper, more converged boundaries. More specifically, U-Net and U-Net++ tend to break small branches, while DeepLabv3+ and CE-Net show insufficient responses to fine structures in some cases, leading to missing branches or degraded connectivity.Overall, these observations are consistent with the quantitative results in Table I, indicating that the proposed method enables more robust vessel reconstruction under complex backgrounds and fine-structure scenarios.

TABLE II
ABLATION STUDY ON XCAD (3-FOLD, MEAN±STD).

Variant	**V-DASI**	**RCS**	**Dice↑**	**IoU↑**	**Sn↑**	**MCC↑**
Baseline	×	×	0.724±0.010	0.571±0.020	0.868±0.011	0.661±0.010
+V-DASI	✓	×	0.760±0.018	0.615±0.016	0.895±0.012	0.735±0.035
+RCS	×	✓	0.752±0.020	0.602±0.018	0.905±0.011	0.742±0.032
Full	✓	✓	**0.790±0.020**	**0.650±0.010**	**0.920±0.010**	**0.780±0.040**

To facilitate an intuitive comparison of response intensity across models in vessel regions, we normalize the vessel probability maps produced by each method and overlay them on the original angiography images using a unified colormap, where red denotes high response and blue denotes low response. As shown in Fig. 5, the response distributions over coronary trunks and branches differ markedly: for some baselines, strong responses are mainly concentrated on the main trunk, while coverage on tiny branches is insufficient or relatively scattered; CE-Net exhibits comparatively sparse effective responses and unstable continuous coverage along branches. In contrast, our method forms more continuous and more complete high-response patterns over both the main trunk and multi-level fine branches, while producing weaker responses in background regions. These observations indicate that the proposed approach more effectively suppresses background interference and enhances attention to fine structures, thereby improving tiny-branch detection and connectivity reconstruction.

D. Ablation Studies

To verify the effectiveness of the key components in HCF-UNet, namely V-DASI and RCS, we used the standard U-Net as the baseline model (*Baseline*). We then introduced V-DASI (*+V-DASI*) or RCS (*+RCS*) individually on top of the baseline, and further combined both to obtain the full model (*Full*, i.e., HCF-UNet). All variants followed the same 3-fold cross-validation splits and the same training/model selection protocol as in the main experiments (saving the best checkpoint based on validation Dice and evaluating it on the corresponding test

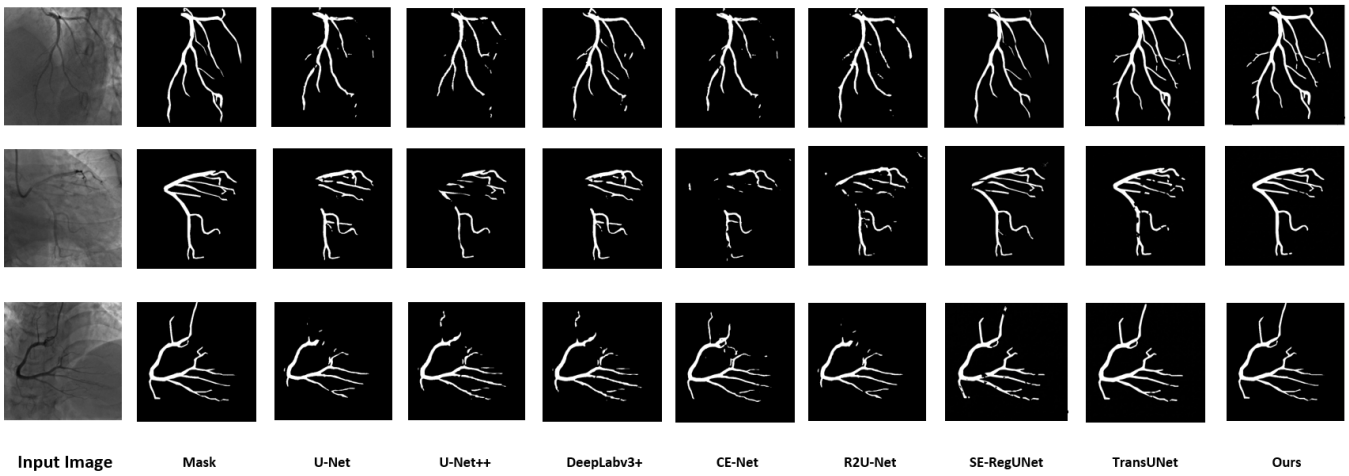

Fig. 4. Qualitative comparison on representative XCAD test samples. From left to right: input image, ground-truth mask, U-Net, U-Net++, DeepLabv3+, CE-Net, R2U-Net, SE-RegUNet, TransUNet, and our HCF-UNet.

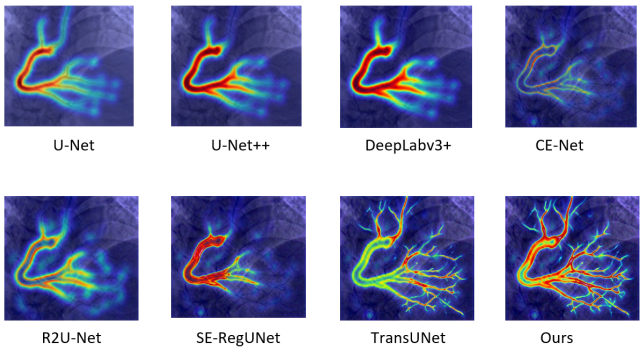

Fig. 5. Normalized vessel-class probability heatmaps overlaid on the original angiography images. Red indicates high probability/high response and blue indicates low probability/low response.

fold) to ensure a fair comparison. The ablation results are summarized in Table 2.

a) Effectiveness of V-DASI.: Adding V-DASI to the baseline (+*V-DASI*) leads to consistent improvements across all metrics: Dice/IoU increases from $0.724 \pm 0.010/0.571 \pm 0.020$ to $0.760 \pm 0.018/0.615 \pm 0.016$ (ΔDice= $+0.036$, ΔIoU= $+0.044$), and Sn/MCC rises from $0.868 \pm 0.011/0.661 \pm 0.010$ to $0.895 \pm 0.012/0.735 \pm 0.035$ (ΔSn= $+0.027$, ΔMCC= $+0.074$). These results indicate that V-DASI enhances feature effectiveness and structural integrity by selectively fusing skip-connection features to suppress redundant background responses and alleviate cross-level semantic mismatch.

b) Effectiveness of RCS.: Introducing RCS only in the decoder (+*RCS*) also yields substantial gains, with a more pronounced improvement in detection capability: Sn/MCC increases to $0.905 \pm 0.011/0.742 \pm 0.032$ (ΔSn= $+0.037$, ΔMCC= $+0.081$), while Dice/IoU improves to $0.752 \pm 0.020/0.602 \pm 0.018$ (ΔDice= $+0.028$, ΔIoU= $+0.031$). This suggests that the residual enhancement and attention-based

TABLE III
MODEL COMPLEXITY COMPARISON UNDER THE SAME SETTING (INPUT 512×512, BATCH=1).

Method	Params (M) ↓	FLOPs (G) ↓	Latency (ms) ↓	FPS ↑
U-Net	31.2	62.8	18.9	52.9
SPIRONet	69.1	152.0	20.0	50.0
Ours (HCF-UNet)	**34.6**	**74.5**	**16.0**	**62.5**

Note: Input size 512×512, batch size $= 1$, FP32, same GPU and PyTorch/CUDA settings. FLOPs are computed with *¡your-tool¿*. Latency/FPS are averaged over *¡N¿* runs after *¡warm-up¿* iterations.

calibration in RCS facilitate the recovery of thin vessels and boundary structures, improving recall for tiny branches and reducing misclassification under class-imbalanced settings.

c) Complementary Effects.: Jointly incorporating V-DASI and RCS (Full, i.e., HCF-UNet) achieves the best results: Dice= 0.790 ± 0.020, IoU= 0.650 ± 0.010, Sn= 0.920 ± 0.010, and MCC= 0.780 ± 0.040. Compared with the baseline, the improvements are $+0.066/+0.079/+0.052/+0.119$ on Dice/IoU/Sn/MCC, respectively. Moreover, Full still outperforms +*V-DASI* (ΔDice= $+0.030$, ΔIoU= $+0.035$) and +*RCS* (ΔDice= $+0.038$, ΔIoU= $+0.048$), demonstrating that the cleaner and better-aligned fused features provided by V-DASI complement the fine-grained reconstruction capability of RCS. Their synergy further strengthens fine-structure representation and preserves topological integrity, resulting in the best overall performance.

E. Model Complexity Analysis

Under the same setting (input 512×512, batch=1), we compare the parameter count, FLOPs, and inference efficiency of U-Net, SPIRONet [25], and HCF-UNet (Table III). The results show that HCF-UNet introduces only a slight increase in parameters and computational cost over U-Net, while achieving lower inference latency. Moreover, compared with SPIRONet, HCF-UNet roughly halves both parameters and FLOPs, and reduces the per-image latency from 20.0 ms to

16.0 ms (FPS increases from 50.0 to 62.5), indicating a more favorable accuracy–complexity trade-off and better suitability for clinical deployment.

V. Conclusion

This paper presents an efficient encoder–decoder network, termed **HCF-UNet**, for coronary angiography vessel segmentation. By introducing the vessel-dimension-aware selective integration module **V-DASI** in the skip-connection pathway and designing the **RCS** (Residual–Channel–Spatial) decoding module, HCF-UNet enables efficient fusion of multi-scale semantics with high-resolution details, leading to fine-grained reconstruction of vascular structures. Experimental results on the XCAD dataset demonstrate that HCF-UNet consistently outperforms a range of mainstream segmentation methods in terms of Dice, IoU, Sn, and MCC. Ablation studies further confirm the individual effectiveness of V-DASI and RCS as well as their complementary gains when combined. Moreover, the complexity analysis indicates that HCF-UNet achieves a more favorable efficiency–accuracy trade-off than the compared model in terms of parameter size and inference speed, highlighting its potential for practical deployment.

Acknowledgment

This research was partially supported by grants from the Key Research of the Central Guidance Fund for Local Science and Technology Development (No. 2023ZYDF078).

References

[1] S. Shit, J. C. Paetzold, A. Sekuboyina *et al.*, "cldice—a novel topology-preserving loss function for tubular structure segmentation," in *Proc. CVPR*, 2021, pp. 16 560–16 569.

[2] S. Chang, C. Lin, W. Wang *et al.*, "Optimizing ensemble u-net architectures for robust coronary vessel segmentation in angiographic images," *Scientific Reports*, vol. 14, p. 6640, 2024.

[3] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Proc. MICCAI*, 1998, pp. 130–137.

[4] Z. Gao, L. Wang, S. M. R. Soroushmehr *et al.*, "Vessel segmentation for x-ray coronary angiography using ensemble methods with deep learning and filter-based features," *BMC Medical Imaging*, vol. 22, p. 10, 2022.

[5] K. Iyer, C. P. Najarian, A. A. Fattah *et al.*, "Angionet: a convolutional neural network for vessel segmentation in x-ray angiography," *Scientific Reports*, vol. 11, p. 18066, 2021.

[6] Anonymous, "Self-supervised vessel segmentation via adversarial learning," in *Proc. ICCV*, 2021, pp. 7516–7525.

[7] Y. Zeng, H. Liu, J. Hu *et al.*, "Pretrained subtraction and segmentation model for coronary angiograms," *Scientific Reports*, vol. 14, p. 19888, 2024.

[8] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, pp. 203–211, 2021.

[9] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. CVPR*, 2018, pp. 7132–7141.

[10] S. Woo, J. Park, J. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proc. ECCV*, 2018, pp. 3–19.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, 2016, pp. 770–778.

[12] K. Han, Y. Wang, Q. Tian *et al.*, "Ghostnet: More features from cheap operations," in *Proc. CVPR*, 2020, pp. 1577–1586.

[13] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. IJCAI*, 1995, pp. 1137–1145.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015.

[15] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," in *Proc. ICLR*, 2017.

[16] F. Milletari, N. Navab, and S. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 3DV*, 2016, pp. 565–571.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[18] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Proc. DLMIA*, 2018, pp. 3–11.

[19] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. ECCV*, 2018, pp. 801–818.

[20] Z. Gu, J. Cheng, H. Fu *et al.*, "Ce-net: Context encoder network for 2d medical image segmentation," arXiv preprint, 2019.

[21] M. Z. Alom, M. Hasan, C. Yakopcic *et al.*, "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," *Journal of Medical Imaging*, vol. 6, no. 1, p. 014006, 2019.

[22] J. Chen, Y. Lu, Q. Yu *et al.*, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv preprint, 2021.

[23] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[24] B. W. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme," *Biochimica et Biophysica Acta (BBA) - Protein Structure*, vol. 405, no. 2, pp. 442–451, 1975.

[25] D.-X. Huang, X.-H. Zhou, X.-L. Xie *et al.*, "Spironet: Spatial-frequency learning and topological channel interaction network for vessel segmentation," arXiv preprint, 2024.