UNCERTAINTY-GUIDED SPECTRAL ADAPTATION FOR CORONARY ARTERY SEGMENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Coronary artery segmentation in invasive coronary angiography (ICA) is hindered by changes in imaging conditions and ambiguous vessel boundaries. These challenges are particularly pronounced for thin, low-contrast branches, whose structural details are easily lost and whose uncertain pixels can compromise cross-domain alignment. We propose **FUP-Net**, a Frequency-aware Uncertainty-guided Prototype Network that couples spectral representation learning with uncertainty-guided adaptation. Its Frequency-aware Heterogeneous Mamba2 (FH-Mamba2) module captures complementary frequency information by combining long-range vascular modeling with convolutional processing of local details. Building on these features, an Uncertainty-guided Prototype Adaptation (UPA) mechanism reduces the influence of ambiguous pixels when updating semantic prototypes, enabling more reliable alignment across imaging domains. Experiments on multiple ICA benchmarks assess segmentation accuracy, vascular topology, and cross-domain generalization. FUP-Net improves Dice by 2.28, 7.40, and 3.65 percentage points over the strongest baselines on JSPH-ICA, DCA1, and ARCADE, respectively. Ablation studies further validate the contributions of frequency-aware modeling and uncertainty-guided prototype adaptation.

1 INTRODUCTION

Segmenting thin and elongated anatomical structures remains challenging in medical imaging. Low contrast and imaging noise can obscure fine structures, while small boundary errors may disrupt their connectivity. These difficulties are particularly evident in invasive coronary angiography (ICA), an important modality for assessing coronary anatomy (Li et al., 2015). A reliable segmentation model must delineate narrow vessel branches, preserve vascular connectivity, and remain effective across variations in imaging conditions.

Deep learning has advanced medical image segmentation through increasingly expressive feature modeling. CNN-based methods (LeCun et al., 1998; Ronneberger et al., 2015) are effective at extracting local patterns, but their ability to capture long-range vascular relationships can be limited. Transformer-based models (Vaswani et al., 2017) incorporate broader contextual information, often at a substantial computational cost. More recently, state space models and Mamba-based architectures (Gu et al., 2021; Gu & Dao, 2023; Ruan et al., 2024) have offered an efficient alternative for modeling long-range dependencies. However, improved contextual modeling alone does not resolve two persistent problems in ICA segmentation: the loss of fine vascular structures and the sensitivity of learned features to changes in image appearance.

These problems are closely related but require different forms of reasoning. Vessel connectivity and coarse anatomical organization benefit from broad spatial context, whereas thin branches and boundaries depend on localized detail. Processing both with the same feature operator may therefore favor one at the expense of the other. Meanwhile, differences in contrast, noise, and annotation characteristics can shift feature distributions across datasets. Under such shifts, pixels near poorly defined vessel boundaries are also more likely to be misclassified; allowing these uncertain features to contribute equally to semantic alignment may introduce errors into the adaptation process. As illustrated in Figure 1, existing methods can produce background interference, fragmented vessels, and missing thin branches under challenging imaging conditions.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

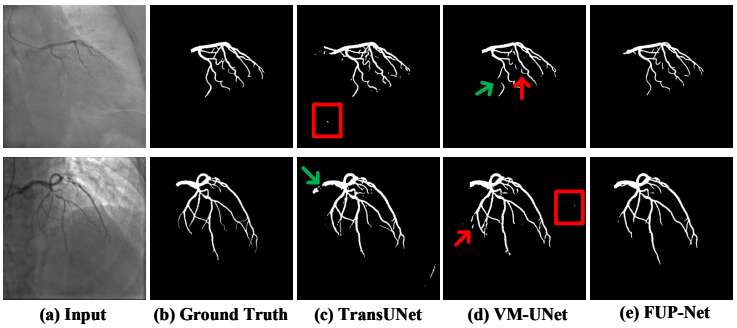


Figure 1: Typical failure cases of existing segmentation methods. (a) ICA images, (b) ground truths, and (c-e) segmentation results of TransUNet (Chen et al., 2024), VM-UNet (Ruan et al., 2024), and the proposed FUP-Net. Existing methods show background interference, discontinuous predictions, and missing fine vascular branches.

To address these issues, we propose **FUP-Net**, which combines **F**requency-aware representation learning with **U**ncertainty-guided **P**rototype adaptation. The central idea is to use complementary frequency information to represent vascular structure and detail, and then use prediction uncertainty to regulate how these representations contribute to cross-domain semantic alignment. Specifically, we introduce a Frequency-aware Heterogeneous Mamba2 (FH-Mamba2) module that processes low- and high-frequency feature components with different operators. A Mamba2 branch models long-range relationships in the low-frequency component, while a convolutional branch extracts localized patterns from the high-frequency component. On the resulting representation, an Uncertainty-guided Prototype Adaptation (UPA) mechanism updates semantic prototypes while reducing the influence of ambiguous features. This design aims to preserve vascular structure without allowing unreliable regions to dominate category-level adaptation.

Our contributions are summarized as follows:

- We propose FUP-Net, a framework that connects frequency-aware vascular representation learning with uncertainty-guided semantic adaptation for ICA segmentation.
- We develop FH-Mamba2 to model complementary low- and high-frequency features using Mamba2 and convolution, respectively, targeting long-range vessel relationships and fine local structures.
- We introduce UPA to regulate prototype updates using prediction uncertainty and promote more reliable category-level alignment across imaging domains.
- We evaluate FUP-Net on multiple ICA datasets using segmentation and topology metrics, including cross-domain experiments and ablations of the proposed components.

2 Related Work

Global-Local Representation Learning. Recent advances in representation learning have improved medical image segmentation by enhancing global contextual modeling and multi-scale feature interaction. CNN-based methods such as U-Net (Ronneberger et al., 2015) effectively capture local patterns but are limited in modeling long-range dependencies. Transformer-based approaches, including TransUNet (Chen et al., 2024), MT-UNet (Wang et al., 2021), and Swin-Unet (Cao et al., 2022), improve global dependency modeling through attention mechanisms. For vascular segmentation, Perspective+UNet (Hu et al., 2024) and ConDSeg (Lei et al., 2025) further exploit attention-based fusion strategies for boundary refinement. However, attention-based methods often introduce high computational costs and may still struggle with fine-grained vascular preservation, resulting in missing distal branches and topology degradation.

Recently, SSMs have emerged as efficient alternatives for long-range dependency modeling. S4 (Gu et al., 2021) employs structured state space representations, while Mamba (Gu & Dao, 2023) achieves linear-complexity sequence modeling through selective state mechanisms. Vision Mamba-based methods, such as VM-UNet (Ruan et al., 2024) and Mamba-UNet (Wang et al., 2024), inte-

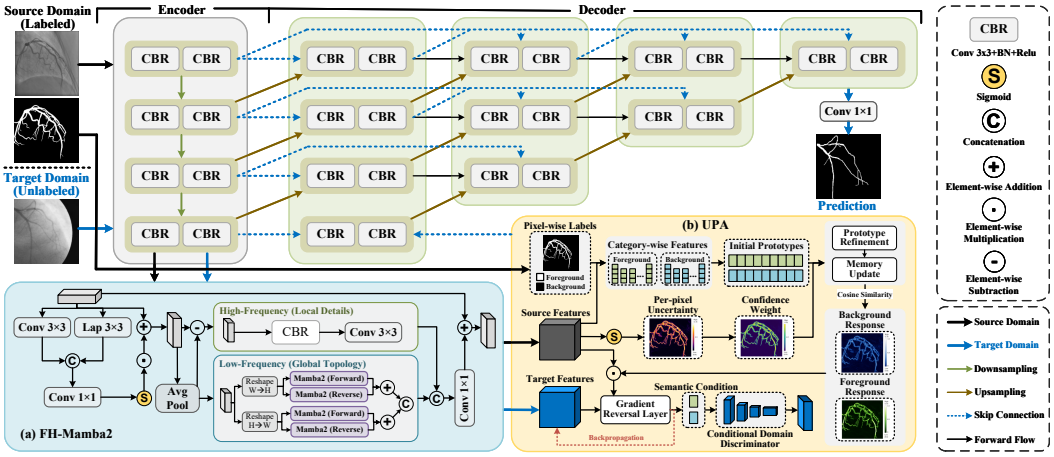


Figure 2: Overview of FUP-Net. FH-Mamba2 performs frequency-aware representation learning, while UPA enables uncertainty-guided semantic adaptation for robust coronary artery segmentation.

grate SSMs into medical segmentation frameworks and improve global feature representation. Nevertheless, existing Mamba-based approaches mainly emphasize global dependency modeling while applying homogeneous operations to diverse vascular structures, which may weaken the representation of high-frequency details such as thin branches and ambiguous boundaries.

Frequency-Aware Adaptation. Frequency-aware representation learning has been explored to enhance structural modeling in medical image segmentation. FRCU-Net (Azad et al., 2021) employs Discrete Cosine Transform (DCT) for high-frequency enhancement, Yan *et al.* Yan et al. (2025) introduce frequency-guided attention for multi-scale fusion, and FAT-Net (Mu et al., 2026) combines frequency attention with topology refinement for vascular segmentation. Despite their effectiveness, existing frequency-based methods mainly rely on predefined transformations or independent frequency enhancement, lacking adaptive modeling of heterogeneous structural patterns.

Meanwhile, domain shifts caused by imaging devices, acquisition protocols, and data sources remain a critical challenge in medical segmentation. Existing domain adaptation methods mainly alleviate domain discrepancies through adversarial learning or feature distribution alignment. Kamnitsas *et al.* Kamnitsas et al. (2017) introduced adversarial adaptation for medical image segmentation, while recent methods such as ASDA (Gou & Cui, 2026) and unified semantic anchoring (Wang et al., 2026) further improve adaptation through feature clustering and semantic constraints. However, these methods mainly focus on distribution-level alignment and often overlook category-specific semantic consistency and uncertainty in target-domain features. Therefore, jointly achieving adaptive structural representation and uncertainty-aware semantic adaptation remains challenging for robust coronary artery segmentation under domain shifts.

3 METHOD

Overview Architecture. As shown in Figure 2, FUP-Net integrates frequency-aware representation and uncertainty-guided prototype adaptation for coronary artery segmentation. We adopt UNet++ (Zhou et al., 2018) to extract hierarchical features and fuse vascular information across scales. At the bottleneck, FH-Mamba2 module decomposes features into complementary low- and high-frequency components. Mamba2 models long-range vascular relationships in the low-frequency branch, while convolution captures fine branches and boundary details in the high-frequency branch. The fused features are then passed to the decoder for segmentation. To improve robustness across heterogeneous ICA domains, the Uncertainty-guided Prototype Adaptation (UPA) mechanism establishes category-level semantic correspondence using the learned representations. It regulates prototype updates according to prediction uncertainty, reducing the contribution of ambiguous regions to cross-domain alignment. Thus, FH-Mamba2 strengthens the representation of vascular structure, while UPA improves the reliability of semantic adaptation.

3.1 FREQUENCY-AWARE HETEROGENEOUS MAMBA2

State space models (SSMs) aim to efficiently model long-range dependencies (Dao & Gu, 2024). However, applying the same operator to all visual features may be suboptimal for coronary artery segmentation: vessel connectivity requires broad contextual information, whereas thin branches and boundaries depend on localized detail. To accommodate these different structural characteristics, we propose Frequency-aware Heterogeneous Mamba2 (FH-Mamba2), which separates smoothly varying features from local residual details and processes them with complementary operators.

The motivation for this design is further discussed through an analysis of state transitions in **Appendix A**. In particular, state accumulation can support the modeling of contextual information across distant positions, while rapidly varying local patterns benefit from an explicit local feature extractor. This motivates the use of Mamba2 for the smooth component and convolution for the detail component, rather than treating Mamba2 as an inherently low-pass operator.

As shown in Figure 2(a), FH-Mamba2 first enhances local responses before frequency-aware separation. Given an input feature map $\mathbf{x}$, a fixed Laplacian operator extracts edge responses, while a learnable depthwise convolution captures local context:

$$\mathbf{e}_3 = \mathrm{Lap}_{3\times3}(\mathbf{x}), \qquad \mathbf{e}_5 = \mathrm{Conv}^{dw}_{5\times5}(\mathbf{x}). \tag{1}$$

Their responses are used to generate a spatially adaptive gate:

$$\mathbf{g} = \sigma\big(\mathrm{Conv}_{1\times1}([\mathbf{e}_3, \mathbf{e}_5])\big), \qquad \mathbf{x}_{\mathrm{enh}} = \mathbf{x} + \mathbf{e}_3 \odot \mathbf{g}, \tag{2}$$

where $[\cdot,\cdot]$ denotes concatenation and $\odot$ denotes element-wise multiplication. The depthwise response thus helps regulate the contribution of the explicit edge response. We then obtain a smooth component through average pooling and define the complementary detail component by subtraction:

$$\mathbf{x}_{\mathrm{low}} = \mathrm{AvgPool}(\mathbf{x}_{\mathrm{enh}}), \qquad \mathbf{x}_{\mathrm{high}} = \mathbf{x}_{\mathrm{enh}} - \mathbf{x}_{\mathrm{low}}. \tag{3}$$

Here, $\mathrm{AvgPool}(\cdot)$ preserves the spatial resolution of the input feature map. It acts as a lightweight smoothing operation, so $\mathbf{x}_{\mathrm{low}}$ and $\mathbf{x}_{\mathrm{high}}$ provide an approximate low-/high-frequency separation rather than an explicit spectral transform. This formulation retains spatial correspondence and avoids the additional processing required by FFT-, DCT-, or wavelet-based decomposition.

For the low-frequency branch, we serialize $\mathbf{x}_{\mathrm{low}}$ along the horizontal and vertical directions and apply bidirectional Mamba2 modeling. For each direction $i \in \{w, h\}$, let $\mathbf{s}_i = \mathcal{S}_i(\mathbf{x}_{\mathrm{low}})$ denote the serialized sequence. Its bidirectional response is computed as

$$\mathbf{z}_i = \mathrm{Mamba2}(\mathbf{s}_i) + \mathrm{reverse}\big(\mathrm{Mamba2}\big(\mathrm{reverse}(\mathbf{s}_i)\big)\big), \qquad i \in \{w, h\}. \tag{4}$$

The outer reverse operation restores the original sequence order before fusion. The responses are then mapped back to the spatial domain and fused:

$$\mathbf{y}_{\mathrm{low}} = \mathcal{F}_{\mathrm{dir}}\big(\mathcal{S}_w^{-1}(\mathbf{z}_w), \mathcal{S}_h^{-1}(\mathbf{z}_h)\big), \tag{5}$$

where $\mathcal{F}_{\mathrm{dir}}$ denotes directional fusion. In parallel, a lightweight convolutional branch $\mathbf{y}_{\mathrm{high}} = F_{\mathrm{CNN}}(\mathbf{x}_{\mathrm{high}})$ processes the high-frequency component to retain local vessel patterns. Finally, the two branches are combined with a residual mapping of the input:

$$\mathbf{y} = \mathcal{F}_{\mathrm{freq}}\left(\mathbf{y}_{\mathrm{low}}, \mathbf{y}_{\mathrm{high}}\right) + \mathcal{R}(\mathbf{x}), \tag{6}$$

where $\mathcal{F}_{\mathrm{freq}}$ denotes cross-branch fusion and $\mathcal{R}$ is the residual mapping. The resulting representation is used for segmentation and provides features for the subsequent uncertainty-guided prototype adaptation. By assigning different operators to the smooth and detail components, FH-Mamba2 combines long-range vascular modeling with local detail preservation. Figure 3 provides a qualitative comparison with the variant without frequency-aware separation: FH-Mamba2 shows more continuous responses along vessels and stronger responses around small branches. These visualizations complement, rather than replace, the quantitative evaluation of structural preservation.

3.2 UNCERTAINTY-GUIDED PROTOTYPE ADAPTATION

Feature distribution alignment alone may associate pixels with different semantic meanings across domains, particularly when vessel boundaries are ambiguous or image quality varies. Prototype-based adaptation introduces category-level anchors, but prototypes estimated from uncertain target

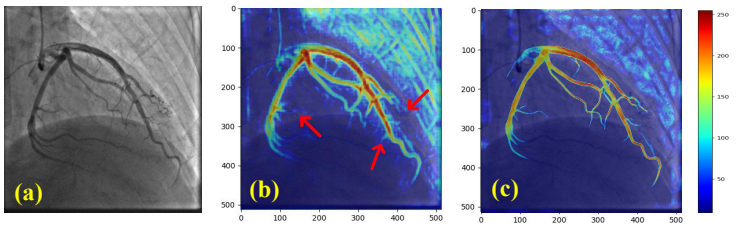


Figure 3: Grad-CAM (Selvaraju et al., 2020) visualization of FH-Mamba2 variants. (a) Input image, (b) FH-Mamba2 without frequency-aware decomposition, and (c) complete FH-Mamba2. The proposed design enhances responses around fine vascular structures and boundaries.

predictions can accumulate errors during training. To improve the reliability of semantic alignment, we propose Uncertainty-Guided Prototype Adaptation (UPA), which uses predictive uncertainty to regulate prototype updates and then performs prototype-guided conditional domain alignment, as shown in Figure 2(b).

Let $\mathbf{F} \in \mathbb{R}^{B\times C\times H\times W}$ denote the feature maps used for adaptation. For each category $c \in \{fg, bg\}$, source-domain annotations identify the corresponding pixel features. We initialize the category prototype by averaging these source features:

$$P_c^{(0)} = \frac{1}{|\mathcal{I}_c^s|} \sum_{i\in\mathcal{I}_c^s} f_i^s, \tag{7}$$

where $\mathcal{I}_c^s$ is the set of source pixels labeled as category c, and f_i^s is the feature vector at pixel i. For the target domain, category assignments are obtained from the segmentation predictions. Because these assignments may be unreliable, their contributions to subsequent prototype updates are weighted by predictive uncertainty.

For a target pixel i, uncertainty is defined as the entropy of its predicted foreground probability p_i^t:

$$U_i^t = -\left[p_i^t \log p_i^t + (1-p_i^t)\log(1-p_i^t)\right]. \tag{8}$$

The entropy is converted into a confidence weight:

$$w_i^t = 1 - \lambda_u \frac{U_i^t - U_{\min}^t}{U_{\max}^t - U_{\min}^t + \epsilon}, \tag{9}$$

where $U_{\min}^t$ and $U_{\max}^t$ are the minimum and maximum target-domain entropy values within the current batch, $\lambda_u \in [0,1]$ controls the strength of uncertainty suppression, and ϵ ensures numerical stability. Thus, more uncertain target pixels receive lower weights.

At each iteration, source features and uncertainty-weighted target features are used to estimate the category prototypes:

$$P_c^{\text{new}} = \frac{\sum_{i\in\mathcal{I}_c^s} f_i^s + \sum_{j\in\mathcal{I}_c^t} w_j^t f_j^t}{|\mathcal{I}_c^s| + \sum_{j\in\mathcal{I}_c^t} w_j^t + \epsilon}, \tag{10}$$

where $\mathcal{I}_c^t$ denotes the set of target pixels assigned to category c by the segmentation model. The prototypes are maintained using a momentum update:

$$P_c \leftarrow \mu P_c + (1-\mu) P_c^{\text{new}}, \tag{11}$$

where μ controls the contribution of the existing prototype. This update reduces the effect of uncertain target features while allowing the semantic anchors to evolve during training.

The updated prototypes are then used to recalibrate pixel features. For each feature vector f_i, we compute its similarity to the foreground and background prototypes and normalize them:

$$[\alpha_{i,bg}, \alpha_{i,fg}] = \text{Softmax}\left([\cos(f_i, P_{bg}), \cos(f_i, P_{fg})]\right). \tag{12}$$

The foreground response provides a residual enhancement of the corresponding feature:

$$\hat{f}_i = f_i\left(1 + \alpha_{i,fg}\right). \tag{13}$$

Accordingly, features with stronger foreground-prototype affinity receive greater amplification. Applying this operation to all pixels yields the enhanced feature map $\hat{\mathbf{F}}$.

Finally, UPA applies conditional domain alignment to the enhanced features. A Gradient Reversal Layer (GRL) passes $\hat{\mathbf{F}}$ to the domain discriminator while reversing its gradients:

$$D_{\text{out}} = D\Big(\text{GRL}(\hat{\mathbf{F}}, \lambda), \mathbf{c}\Big), \tag{14}$$

where λ is the gradient reversal coefficient and $\mathbf{c}$ provides category-level semantic conditions derived from source annotations or target predictions. Conditioning the discriminator on semantic information encourages alignment between corresponding categories rather than only between marginal feature distributions. Together, uncertainty-weighted prototype updates and conditional alignment make the adaptation process less dependent on ambiguous target regions.

3.3 Training Objectives

FUP-Net is trained with a joint objective that combines supervised segmentation and domain adaptation. The source-domain segmentation loss consists of Dice loss and binary cross-entropy (BCE) loss, which encourage accurate pixel-wise predictions and vessel-region overlap. Meanwhile, the UPA objective $\mathcal{L}_{\text{UPA}}$ promotes uncertainty-guided, prototype-based alignment between source and target domains. The overall objective is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seg}}^{\text{src}} + \lambda_{\text{upa}} \mathcal{L}_{\text{UPA}}, \tag{15}$$

where $\mathcal{L}_{\text{seg}}^{\text{src}}$ denotes the source-domain segmentation loss and λ_{upa} balances supervised segmentation and domain adaptation. The target domain is used without segmentation labels during adaptation. Detailed formulations are provided in **Appendix B**.

4 Experiments

4.1 Experimental Setup

FUP-Net is implemented in PyTorch and trained using the AdamW optimizer (Loshchilov & Hutter, 2017). All compared methods are evaluated under identical training settings, data splits, and evaluation protocols to ensure fair comparison. To assess both segmentation accuracy and cross-domain robustness, experiments are conducted on three invasive coronary angiography (ICA) datasets: JSPH-ICA (Zhao et al., 2021), DCA1 (Cervantes-Sanchez et al., 2019), and ARCADE (Popov et al., 2024). JSPH-ICA and DCA1 use fixed test sets with 5-fold cross-validation on the remaining data, with each fold model evaluated on the fixed test set, while ARCADE follows the official 1000/200/300 training-validation-test split. Segmentation performance is evaluated using Dice, Sensitivity, Specificity, HD95, and average symmetric surface distance (ASSD), while computational efficiency is analyzed using model complexity metrics. Detailed training configurations, dataset descriptions, and metric definitions are provided in **Appendix C**.

4.2 Segmentation Performance under Domain Shift

We compare FUP-Net with seven state-of-the-art segmentation models, including Swin-Unet (Cao et al., 2022), U-Net (Ronneberger et al., 2015), U-Net++ (Zhou et al., 2018), DSCNet (Qi et al., 2023), VM-UNet (Ruan et al., 2024), ConDSeg (Lei et al., 2025), and FAT-Net (Mu et al., 2026).

For JSPH-ICA, all methods are evaluated under supervised segmentation settings without domain adaptation. For DCA1 and ARCADE, JSPH-ICA serves as the source domain, and UPA is jointly optimized using labeled source-domain data and unlabeled target-domain data to improve cross-domain robustness. FUP-Net without UPA is trained using source-domain supervision only and evaluated on the target domains to isolate the contribution of prototype-based adaptation. Wilcoxon signed-rank tests on Dice scores are conducted to assess the statistical significance of the observed improvements, with detailed results reported in **Appendix D**.

Quantitative Evaluation. Table 1 summarizes the quantitative comparison among eight models across three ICA datasets. On JSPH-ICA, all methods are trained and evaluated within the same

Table 1: Quantitative comparison on three ICA datasets. For FUP-Net, JSPH-ICA is used as the source domain, while DCA1 and ARCADE are evaluated for cross-domain generalization. Each entry reports Dice/Sensitivity/Specificity (%) (first row) and HD95/ASSD (pixel) (second row). Higher Dice, Sensitivity, and Specificity, and lower HD95 and ASSD indicate better performance.

Model	JSPH-ICA	DCA1	ARCADE
Swin-Unet	72.07/73.00/98.35/21.34/3.37	52.05/52.44/97.06/30.24/9.83	60.93/65.87/97.93/70.98/14.26
U-Net	85.03/83.49/99.38/11.62/2.46	61.02/54.28/99.42/28.90/6.69	71.33/70.80/99.13/60.76/10.22
U-Net++	85.57/83.53/99.41/9.36/2.03	69.95/66.73/**99.55**/27.95/4.95	66.83/60.53/99.04/57.76/10.35
DSCNet	86.25/**91.84**/99.12/10.91/2.00	67.80/78.79/98.46/24.43/5.34	72.32/**79.04**/98.96/55.74/9.93
VM-UNet	87.30/89.11/99.38/9.64/1.96	73.24/77.55/99.32/16.81/4.43	73.18/71.55/99.26/54.82/10.08
CondSeg	87.02/85.46/99.48/9.93/2.12	72.03/77.46/99.39/17.56/4.52	73.94/69.83/99.06/55.25/9.61
FAT-Net	87.25/85.75/**99.53**/9.52/2.35	73.11/78.95/99.26/**16.54**/4.61	72.98/76.67/99.32/54.53/10.99
FUP-Net w/o UPA	**89.58**/90.42/99.44/**9.19**/**1.89**	77.37/77.30/99.13/18.13/4.71	75.33/76.27/99.33/56.77/10.38
FUP-Net	*Source*	**80.64**/**81.93**/99.50/17.14/**3.35**	**77.59**/77.98/**99.37**/**53.63**/**9.24**

domain. FUP-Net achieves the best segmentation performance with a Dice of 89.58%, together with the lowest HD95 (9.19 pixels) and ASSD (1.89 pixels). These results demonstrate the effectiveness of FH-Mamba2 in preserving fine vascular structures and maintaining topology consistency. For cross-domain evaluation on DCA1 and ARCADE, FUP-Net without UPA achieves competitive performance compared with existing methods, while incorporating UPA further improves domain adaptation. On DCA1, FUP-Net obtains a Dice score of 80.64% and reduces boundary errors, indicating improved robustness under domain variations and image degradation. On ARCADE, FUP-Net improves Dice from 75.33% to 77.59% compared with the variant without UPA, while achieving the best HD95 (53.63 pixels), ASSD (9.24 pixels), and specificity (99.37%). These results demonstrate that uncertainty-guided prototype adaptation enhances semantic consistency and cross-domain generalization under heterogeneous ICA conditions.

Table 2 compares the computational complexity of different methods. Although U-Net and Swin-Unet require fewer computational resources, they generally provide lower segmentation performance. FAT-Net achieves a relatively compact design, while FUP-Net obtains superior segmentation accuracy with 9.14M parameters, 277.77G FLOPs, 23.27 ms inference time, and 689.98 MB memory consumption, demonstrating a favorable trade-off between accuracy and computational cost.

Table 2: Comparisons of Computational complexity.

Model	Params(M)↓	FLOPs(G)↓	Inference Time(ms)↓	Peak Memory(MB)↓
Swin-Unet	27.20	22.10	9.20	980.4
U-Net	1.94	84.00	6.50	612.3
U-Net++	7.30	97.20	8.10	1350.7
DSCNet	2.10	81.00	15.80	1120.4
VM-UNet	29.95	402.50	108.50	5870.2
CondSeg	45.60	812.10	244.10	6188.7
FAT-Net	9.06	101.74	25.94	1411.0
FUP-Net	9.14	277.77	23.27	689.98

Qualitative Evaluation. Figure 4 presents qualitative comparisons across three ICA datasets. FUP-Net achieves more continuous vessel reconstruction and better boundary preservation, particularly in tortuous vessels, bifurcation regions, and distal branches (red arrows). In contrast, existing methods often suffer from missing thin branches, discontinuous predictions, background interference, and inaccurate boundaries under low-contrast conditions. Compared with U-Net++ and DSCNet, FUP-Net demonstrates improved stability in bifurcation regions and weak-texture areas. Transformer-based Swin-Unet and CNN-based U-Net can recover major vascular structures but tend to miss fine branches. Meanwhile, VM-UNet and ConDSeg occasionally produce discontinuous predictions in distal vessel regions. Although FAT-Net achieves competitive segmentation performance, FUP-Net further improves fine-grained vessel preservation and boundary recovery, benefiting from frequency-aware global-local representation learning and uncertainty-guided semantic adaptation.

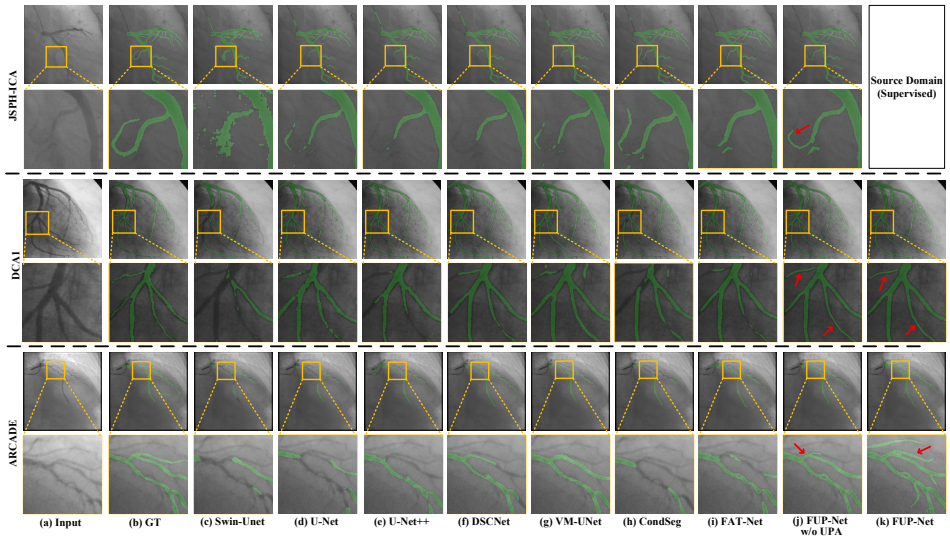


Figure 4: Qualitative comparison of coronary artery segmentation results on three ICA datasets. (a) ICA images; (b) ground truth; (c-k) segmentation results of different methods with zoom-in views. For FUP-Net, JSPH-ICA is used as the source domain, where domain adaptation is not applied.

These qualitative observations are consistent with the quantitative results in Table 1, confirming the capability of FUP-Net in preserving vascular topology and enhancing cross-domain robustness. Additional feature response visualizations in **Appendix E** further illustrate that FUP-Net focuses more accurately on discriminative vascular regions. Overall, Figure 4 and Appendix Figure 1 provide visual evidence of the effectiveness of FUP-Net for reliable coronary artery segmentation under heterogeneous ICA conditions.

4.3 Ablation Studies

To investigate the contribution of FH-Mamba2 and UPA, we conduct two ablation studies. First, for FH-Mamba2, three variants are compared: UPP, UPP+FH-M2 w/o FA, and UPP+FH-M2, where UPP denotes the UNet++ baseline, FA denotes the frequency-aware design, and FH-M2 denotes the proposed FH-Mamba2 module. These variants evaluate the effect of frequency-aware heterogeneous modeling on vascular representation learning. Second, for UPA, four variants are evaluated: UPP, UPP+DA, UPP+ProtoDA, and UPP+UPA. UPP+DA introduces adversarial domain adaptation, UPP+ProtoDA incorporates prototype-guided semantic alignment, and UPP+UPA further introduces uncertainty-guided prototype adaptation. These progressive variants validate the contributions of adversarial alignment, prototype-based adaptation, and uncertainty-guided prototype refinement for cross-domain robustness.

FH-Mamba2 ablation experiments are conducted on all three datasets, while UPA ablation experiments are evaluated on the two target-domain datasets. The results demonstrate the effectiveness of frequency-aware representation learning and uncertainty-guided prototype adaptation.

Quantitative Evaluation. Table 3 presents component-wise ablation studies of FH-Mamba2 and UPA. For FH-Mamba2, the variant without frequency-aware design improves performance over UPP, demonstrating the benefit of Mamba2-based long-range dependency modeling. Incorporating frequency-aware design further enhances fine-grained vascular representation and structural preservation. On JSPH-ICA, UPP+FH-M2 improves Dice from 85.57% to 89.58%, achieving a Sensitivity of 90.42% and an ASSD of 1.89 pixels. Similar improvements are observed on DCA1 and ARCADE, where UPP+FH-M2 obtains Dice scores of 77.37% and 75.33%, respectively. For UPA, progressive introduction of adversarial alignment, prototype-based semantic alignment, and uncertainty-guided refinement consistently improves cross-domain segmentation performance. On DCA1, UPP+UPA increases Dice from 71.17% to 76.25% and reduces HD95 to 19.05 pixels. On ARCADE, UPP+UPA achieves a Dice of 76.19% and a Specificity of 99.35%, demonstrating improved domain adaptation and false-positive suppression. Overall, the ablation results validate the

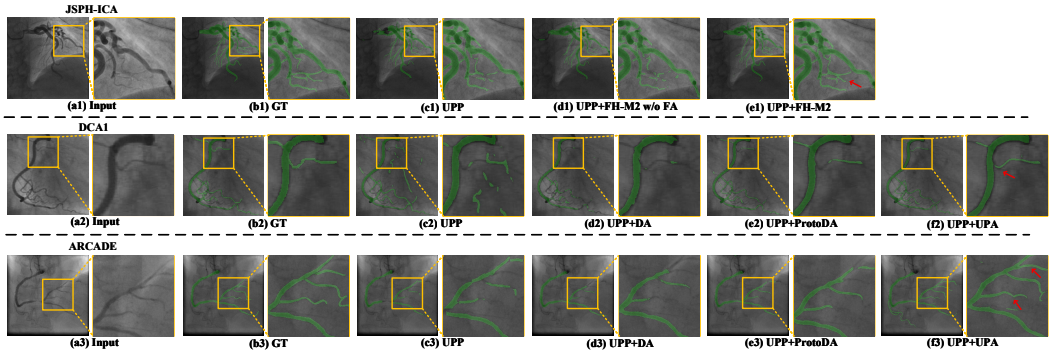


Figure 5: Visual comparison of FH-Mamba2 and UPA ablation results. The FH-Mamba2 ablation results are presented on JSPH-ICA, while the UPA ablation results are presented on DCA1 and ARCADE. Each row presents the input image, ground truth, and segmentation prediction with corresponding zoom-in regions.

Table 3: Ablation studies of FH-Mamba2 and UPA in FUP-Net. For UPA ablation, JSPH-ICA is used as the source domain and DCA1/ARCADE are target domains. Each entry reports Dice/Sensitivity/Specificity (%) (first row) and HD95/ASSD (pixel) (second row).

Model	JSPH-ICA	DCA1	ARCADE
FH-Mamba2 Ablation			
UPP	85.57/83.53/99.41/9.36/2.03	69.95/66.73/**99.55**/27.95/4.95	66.83/60.53/99.04/57.76/**10.35**
UPP+FH-M2 w/o FA	87.39/87.30/99.40/**8.90**/1.98	74.13/74.52/98.42/19.76/4.86	73.22/72.58/99.11/57.48/10.74
UPP+FH-M2	**89.58**/**90.42**/**99.44**/9.19/**1.89**	**77.37**/**77.30**/98.13/**18.13**/**4.71**	**75.33**/**76.27**/**99.33**/**56.77**/10.38
UPA Ablation			
UPP	*Source*	69.95/66.73/**99.55**/27.95/4.95	66.83/60.53/99.04/57.76/10.35
UPP+DA	*Source*	71.17/70.43/99.21/21.39/5.02	70.22/69.78/99.21/56.75/**10.05**
UPP+ProtoDA	*Source*	75.63/76.32/99.33/21.58/4.89	75.02/77.01/99.03/57.62/10.57
UPP+UPA	*Source*	**76.25**/**77.38**/99.36/**19.05**/**4.77**	**76.19**/**76.27**/**99.35**/**55.74**/10.21

benefits of FH-Mamba2 and UPA for segmentation accuracy, boundary stability, and cross-domain robustness.

Qualitative Evaluation. Figure 5 presents qualitative comparisons of FH-Mamba2 and UPA ablations. For FH-Mamba2, UPP+FH-M2 produces more continuous vessel structures and smoother boundaries on JSPH-ICA, especially around thin branches and bifurcations. Compared with other variants, it reduces fragmentation and boundary ambiguity, demonstrating the benefit of frequency-aware heterogeneous modeling. For UPA, UPP+UPA achieves more reliable segmentation on DCA1 and ARCADE under domain shifts, preserving fine branches and curved vessels while suppressing background noise and false positives. The red-arrow regions highlight recovered vascular structures, validating the effectiveness of prototype-based alignment and uncertainty-guided adaptation.

5 Conclusion

This paper proposes FUP-Net, a UNet++-based framework integrating FH-Mamba2 and UPA for robust coronary artery segmentation from invasive coronary angiography. FH-Mamba2 enables frequency-aware heterogeneous representation learning by modeling global topology and fine-grained vascular details with complementary operators. UPA improves cross-domain generalization through prototype-based semantic alignment and uncertainty-guided adaptation. Extensive experiments on multiple ICA datasets demonstrate superior segmentation performance with improved vessel continuity, boundary preservation, and robustness under domain shifts. These results highlight the effectiveness of combining frequency-aware modeling with uncertainty-guided semantic adaptation for coronary artery segmentation.

AI USE STATEMENT

In this work, we used generative AI tools for language proofreading and improving the readability of the manuscript. We have not used generative AI tools for designing the proposed method, conducting experiments, generating experimental results, or analyzing data. The remaining disclosure tasks are not applicable to this work. Additionally, we used generative AI tools for formatting assistance during manuscript preparation. We have reviewed all AI-assisted content and verified its correctness. We take responsibility for the final content of this work, including all text, claims, and artifacts presented in the manuscript.

ETHICS STATEMENT

This study uses publicly available invasive coronary angiography datasets for medical image segmentation research. No new data collection involving human subjects was conducted in this work. The experiments were performed on de-identified datasets following the original dataset release protocols. This study does not involve clinical intervention, patient identification, or deployment of an automated diagnostic system.

REPRODUCIBILITY STATEMENT

To facilitate reproducibility, we provide detailed descriptions of the proposed architecture, training settings, datasets, evaluation metrics, and implementation details in the main paper and Appendix materials. The experimental protocols, including data preprocessing, optimization settings, and evaluation procedures, are also reported to enable independent verification of the results. The source code and trained models will be made available upon acceptance.

REFERENCES

Reza Azad, Afshin Bozorgpour, Maryam Asadi-Aghbolaghi, Dorit Merhof, and Sergio Escalera. Deep frequency re-calibration u-net for medical image segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3274–3283, 2021.

Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pp. 205–218. Springer, 2022.

Fernando Cervantes-Sanchez, Ivan Cruz-Aceves, Arturo Hernandez-Aguirre, Martha Alicia Hernandez-Gonzalez, and Sergio Eduardo Solorio-Meza. Automatic segmentation of coronary arteries in x-ray angiograms using multiscale analysis and artificial neural networks. *Applied Sciences*, 9(24):5507, 2019.

Jieneng Chen, Jieru Mei, Xianhang Li, Yongyi Lu, Qihang Yu, Qingyue Wei, Xiangde Luo, Yutong Xie, Ehsan Adeli, Yan Wang, et al. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis*, 97:103280, 2024.

Tri Dao and Albert Gu. Transformers are ssms: Generalized models and efficient algorithms through structured state space duality. *arXiv preprint arXiv:2405.21060*, 2024.

Qiyu Gou and Guanxun Cui. Adversarial selective domain adaptation with feature cluster for skin cancer diagnosis. *Scientific Reports*, 2026.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021.

Jintong Hu, Siyan Chen, Zhiyi Pan, Sen Zeng, and Wenming Yang. Perspective+ unet: Enhancing segmentation with bi-path fusion and efficient non-local attention for superior receptive fields. In *International conference on medical image computing and computer-assisted intervention*, pp. 499–509. Springer, 2024.

Konstantinos Kamnitsas, Christian Baumgartner, Christian Ledig, Virginia Newcombe, Joanna Simpson, Andrew Kane, David Menon, Aditya Nori, Antonio Criminisi, Daniel Rueckert, et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In *International conference on information processing in medical imaging*, pp. 597–609. Springer, 2017.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Mengqi Lei, Haochen Wu, Xinhua Lv, and Xin Wang. Condseg: A general medical image segmentation framework via contrast-driven feature enhancement. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pp. 4571–4579, 2025.

Zhixun Li, Yingtao Zhang, Guangzhong Liu, Haoyang Shao, Weimin Li, and Xianglong Tang. A robust coronary artery identification and centerline extraction method in angiographies. *Biomedical Signal Processing and Control*, 16:1–8, 2015.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Nan Mu, Ruiqi Song, Xiaoning Li, Zhihui Xu, Jingfeng Jiang, and Chen Zhao. Fat-net: Frequency-domain attention-guided topology-refinement network for coronary artery segmentation in invasive coronary angiography. *IEEE Journal of Biomedical and Health Informatics*, 2026.

Maxim Popov, Akmaral Amanturdieva, Nuren Zhaksylyk, Alsabir Alkanov, Adilbek Saniyazbekov, Temirgali Aimyshev, Eldar Ismailov, Ablay Bulegenov, Arystan Kuzhukeyev, Aizhan Kulanbayeva, et al. Dataset for automatic region-based coronary artery disease diagnostics using x-ray angiography images. *Scientific data*, 11(1):20, 2024.

Yaolei Qi, Yuting He, Xiaoming Qi, Yuan Zhang, and Guanyu Yang. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6070–6079, 2023.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Jiacheng Ruan, Jincheng Li, and Suncheng Xiang. Vm-unet: Vision mamba unet for medical image segmentation. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2024.

Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: visual explanations from deep networks via gradient-based localization. *International journal of computer vision*, 128(2):336–359, 2020.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Bowen Wang, Liangzhi Li, Manisha Verma, Yuta Nakashima, Ryo Kawasaki, and Hajime Nagahara. Mtunet: Few-shot image classification with visual explanations. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pp. 2294–2298, 2021.

Xin Wang, Yin Guo, Jiamin Xia, Kaiyu Zhang, Niranjan Balu, Mahmud Mossa-Basha, Linda Shapiro, and Chun Yuan. Unified and semantically grounded domain adaptation for medical image segmentation. *IEEE Transactions on Medical Imaging*, 2026.

Ziyang Wang, Jian-Qing Zheng, Yichi Zhang, Ge Cui, and Lei Li. Mamba-unet: Unet-like pure visual mamba for medical image segmentation. *arXiv preprint arXiv:2402.05079*, 2024.

Shun Yan, Benquan Yang, Aihua Chen, Xiaoming Zhao, and Shiqing Zhang. Multi-scale convolutional attention frequency-enhanced transformer network for medical image segmentation. *Information Fusion*, 119:103019, 2025.

Chen Zhao, Aviral Vij, Saurabh Malhotra, Jinshan Tang, Haipeng Tang, Drew Pienta, Zhihui Xu, and Weihua Zhou. Automatic extraction and stenosis evaluation of coronary arteries in invasive coronary angiograms. *Computers in biology and medicine*, 136:104667, 2021.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *International workshop on deep learning in medical image analysis*, pp. 3–11. Springer, 2018.

A Frequency-aware Motivation of Mamba2-based Representation Learning

To investigate the intrinsic frequency characteristics of Mamba2, we analyze its state propagation behavior from the perspective of signal processing. Recent state space models (SSMs), such as Mamba2 (Dao & Gu, 2024), achieve efficient long-range dependency modeling through selective state propagation. Given an input sequence x_k, the state transition process can be formulated as:

$$h_k = \sum_i A^i B x_{k-i}, \tag{16}$$

where A denotes the state transition matrix and B represents the input projection matrix. This formulation indicates that the current hidden state is obtained by aggregating historical inputs with different propagation weights. Specifically, the term A^i determines the contribution of previous positions to the current state, corresponding to the memory behavior of the SSM.

Due to the stability constraint of SSMs, the state transition matrix generally satisfies:

$$\|A^i\| \to 0, \tag{17}$$

which indicates an exponentially decaying memory mechanism. Therefore, historical information is gradually attenuated during state propagation, while neighboring and long-range features can still be integrated through recursive state propagation. This property enables Mamba2 to effectively capture global structural dependencies and smooth spatial variations, which are important for modeling large-scale anatomical patterns.

To further analyze the frequency preference of this propagation mechanism, the state response can be transformed into the frequency domain as:

$$H(\omega) = \sum_i A^i B e^{-j\omega i}, \tag{18}$$

where ω denotes the frequency component along the scanning direction and $e^{-j\omega i}$ represents the phase variation introduced by different frequency signals. This formulation reveals that the response of Mamba2 depends on the interaction between the state propagation kernel A^i and the frequency component of the input signal.

For low-frequency components, which correspond to slowly varying spatial patterns, the frequency approaches zero ($\omega \to 0$). Therefore, the phase term can be approximated as:

$$e^{-j\omega i} \approx 1. \tag{19}$$

In this case, historical responses maintain similar phases and can be accumulated coherently. The state propagation mechanism therefore reinforces consistent structural patterns over long distances. In vascular images, such low-frequency information corresponds to global anatomical characteristics, including vessel trajectories, overall topology, and large-scale connectivity. These properties explain why Mamba2 is effective for capturing long-range vascular structures.

However, for high-frequency components, where the signal varies rapidly along the scanning direction, ω approaches π, and the phase term becomes:

$$e^{-j\omega i} = (-1)^i. \tag{20}$$

Under this condition, adjacent historical responses exhibit alternating signs, causing partial cancellation during accumulation. As a result, rapidly changing signals may be weakened by the state propagation process. In medical images, high-frequency information mainly represents local intensity variations, sharp boundaries, thin vessel walls, and small bifurcation branches. Therefore, when all visual features are directly processed by Mamba2, these fine-grained details may be insufficiently preserved, potentially resulting in boundary smoothing or loss of small vascular structures.

Based on the above analysis, vascular features can be considered as the combination of complementary frequency components:

$$x = x_{\text{low}} + x_{\text{high}}, \tag{21}$$

where x_{low} represents slowly varying structural information, including global vessel topology and long-range connectivity, while x_{high} contains rapidly changing local details, such as vessel boundaries and fine branches.

Therefore, instead of applying Mamba2 uniformly to all feature components, we design FH-Mamba2 with a frequency-aware heterogeneous strategy. Specifically, Mamba2 is assigned to model low-frequency structural representations by exploiting its long-range dependency capability, while convolution-based enhancement is introduced for high-frequency components to strengthen local spatial details. This complementary design allows the network to simultaneously preserve global vascular topology and recover fine-grained vessel structures.

B Loss Function Design

To jointly optimize coronary artery segmentation and cross-domain adaptation, FUP-Net adopts a multi-objective optimization strategy consisting of a segmentation loss and an adversarial domain adaptation loss. The segmentation objective aims to preserve accurate vascular structures, while the domain adaptation objective encourages the learning of domain-invariant representations across heterogeneous ICA distributions.

For coronary artery segmentation, pixel-level class imbalance is a critical challenge because vessel regions usually occupy only a small proportion of the entire angiographic image, especially for thin distal branches. Meanwhile, accurate segmentation requires not only pixel-wise classification but also preservation of continuous vascular topology and boundary integrity. Therefore, we employ a hybrid loss combining Binary Cross Entropy (BCE) loss and Dice loss:

$$\mathcal{L}_{seg} = \lambda \cdot \mathcal{L}_{BCE} + (1 - \lambda) \cdot (1 - \text{Dice}), \tag{22}$$

where $\mathcal{L}_{BCE}$ provides pixel-level supervision by measuring the classification discrepancy between prediction and ground truth. However, BCE alone may bias the optimization toward dominant background regions due to the severe foreground-background imbalance in ICA images. Therefore, Dice loss is introduced to directly optimize region-level overlap and improve the continuity of small and fragmented vascular structures. The weighting coefficient λ balances pixel-wise classification accuracy and structural overlap constraints, enabling more stable optimization for coronary artery segmentation.

To improve the generalization capability under cross-domain scenarios, an adversarial domain adaptation loss $\mathcal{L}_{UPA}$ is introduced to reduce distribution discrepancies between different ICA datasets. In clinical angiography, domain shifts may arise from variations in acquisition devices, imaging protocols, contrast injection conditions, image resolution, and annotation styles. These variations can lead to inconsistent vessel appearance, boundary ambiguity, and intensity distribution changes across datasets. Therefore, FUP-Net employs a domain discriminator to distinguish source-domain and target-domain features, while the feature extractor learns domain-invariant representations through adversarial optimization.

Specifically, source and target domain features are assigned domain labels of 0 and 1, respectively. The domain classification objective is formulated as:

$$\mathcal{L}_{UPA} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{BCE}(p_i, \hat{p}_i), \tag{23}$$

where p_i denotes the ground-truth domain label and $\hat{p}_i$ represents the predicted probability from the domain discriminator. By minimizing this adversarial objective through the discriminator and reversing gradients through the feature extractor, the learned vascular representations become less sensitive to domain-specific appearance variations while maintaining discriminative anatomical information.

Finally, the overall training objective is defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seg}}^{\text{src}} + \lambda_{\text{upa}}\mathcal{L}_{\text{UPA}}, \tag{24}$$

where $\mathcal{L}_{\text{seg}}^{\text{src}}$ denotes the segmentation loss on the source domain, and λ_{upa} controls the contribution of domain adaptation. The target domain is used without segmentation labels during adaptation. The joint optimization enables FUP-Net to maintain accurate vessel reconstruction while learning domain-robust representations. In particular, the source-domain segmentation loss preserves ICA-specific anatomical structures, while the UPA loss reduces the impact of imaging variations across datasets, improving robustness for heterogeneous coronary artery segmentation.

C Experimental Setup

Training Details. FUP-Net is implemented in PyTorch and trained on two NVIDIA RTX 3090 GPUs with 48 GB of total memory. The model is optimized using AdamW (Loshchilov & Hutter, 2017) with an initial learning rate of 1×10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a weight decay of 1×10^{-5}. A StepLR scheduler is adopted, with the learning rate decayed according to the predefined schedule. The batch size is set to 2.

Online data augmentation includes random flipping, rotation within $\pm10°$, scaling within 0.9–1.1, and random cropping to 512×512. All comparison methods use identical training settings, data splits, and cross-validation protocols for fair comparison. For experiments without UPA, FUP-Net is trained using standard supervised learning on each dataset. No early stopping is applied, and the model with the highest validation Dice score is selected for evaluation on the held-out test set.

Image Datasets. FUP-Net is evaluated on three ICA datasets with diverse imaging conditions:

1) JSPH-ICA (Zhao et al., 2021) contains 314 ICA images from 99 patients collected at Jiangsu Province People's Hospital, China. The images were acquired using a Siemens AXIOM-Artis system at 15 fps, with a resolution of 512×512 and a spatial resolution of 0.258–0.390 mm/pixel. The dataset includes 187 left coronary artery (LCA) and 127 right coronary artery (RCA) images, all manually annotated by experienced cardiologists.

2) DCA1 (Cervantes-Sanchez et al., 2019) contains 134 expert-annotated ICA images with a resolution of 300×300 pixels, collected from the Cardiology Department of UMAE T1-León (IMSS). The dataset provides a benchmark for coronary artery segmentation under different imaging conditions.

3) ARCADE (Popov et al., 2024) contains 1500 ICA images designed for region-based coronary artery disease analysis, covering diverse anatomical variations and imaging characteristics.

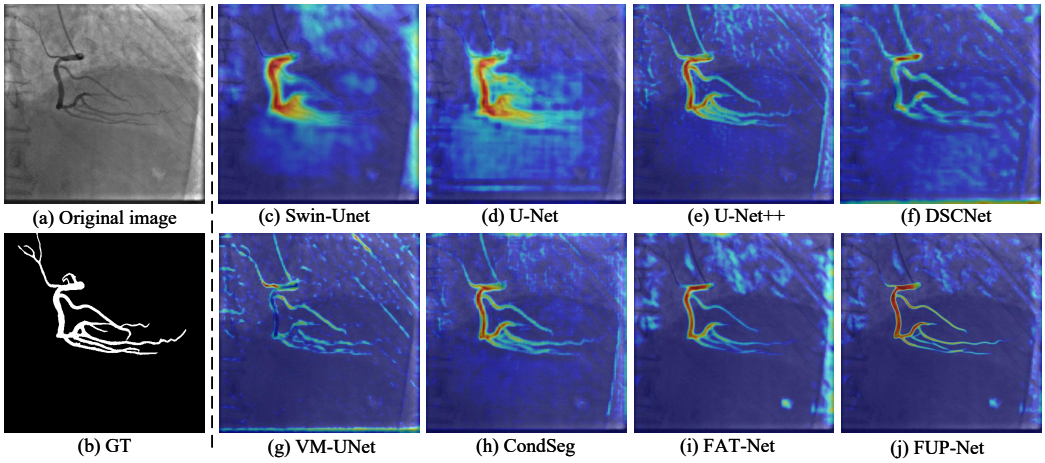
JSPH-ICA and DCA1 use fixed held-out test sets with 5-fold cross-validation on the remaining data, while ARCADE follows its predefined split of 1000/200/300 samples for training, validation, and testing.

Evaluation Metrics. Segmentation performance is evaluated using five commonly used metrics: Dice Similarity Coefficient (Dice), Sensitivity (Sens), Specificity (Spec), 95% Hausdorff Distance (HD95), and Average Symmetric Surface Distance (ASSD). Their formal definitions are provided in (Zhao et al., 2021).

We also evaluate computational characteristics, including parameter count, FLOPs, inference time, and peak memory usage, to assess model complexity and efficiency.

Appendix Table 1: Wilcoxon one-tailed signed-rank test results of FUP-Net against comparison models across three ICA datasets based on per-image Dice scores.

Model	JSPH-ICA	DCA1	ARCADE
Swin-Unet	$p = 1.2 \times 10^{-6}$ (significant)	$p = 3.5 \times 10^{-8}$ (significant)	$p = 2.1 \times 10^{-7}$ (significant)
U-Net	$p = 8.6 \times 10^{-7}$ (significant)	$p = 1.7 \times 10^{-7}$ (significant)	$p = 4.3 \times 10^{-5}$ (significant)
U-Net++	$p = 2.4 \times 10^{-5}$ (significant)	$p = 5.6 \times 10^{-8}$ (significant)	$p = 7.9 \times 10^{-6}$ (significant)
DSCNet	$p = 3.1 \times 10^{-3}$ (significant)	$p = 2.8 \times 10^{-5}$ (significant)	$p = 1.6 \times 10^{-3}$ (significant)
VM-UNet	$p = 7.5 \times 10^{-4}$ (significant)	$p = 8.9 \times 10^{-4}$ (significant)	$p = 4.2 \times 10^{-3}$ (significant)
CondSeg	$p = 1.6 \times 10^{-3}$ (significant)	$p = 2.1 \times 10^{-3}$ (significant)	$p = 6.8 \times 10^{-3}$ (significant)
FAT-Net	$p = 4.8 \times 10^{-4}$ (significant)	$p = 8.3 \times 10^{-4}$ (significant)	$p = 9.7 \times 10^{-4}$ (significant)



Appendix Figure 1: Attention heatmap visualization of different segmentation methods. (a-b) Original image and GT; (c-j) attention maps showing activation regions, where red indicates high responses and blue indicates low responses.

D Statistical Significance Analysis

Appendix Table S1 reports the results of one-sided Wilcoxon signed-rank tests comparing FUP-Net with the competing methods based on per-image Dice scores across the three ICA datasets. For each comparison, the per-image Dice scores of FUP-Net and the corresponding competing method are paired and evaluated using a one-sided Wilcoxon signed-rank test, with the alternative hypothesis that FUP-Net achieves higher Dice scores. As shown in Table S1, FUP-Net achieves statistically significant improvements over all comparison methods on JSPH-ICA, DCA1, and ARCADE, with all comparisons yielding $p < 0.05$. These results provide statistical evidence supporting the observed performance improvements of FUP-Net across the evaluated datasets.

E Activation Visualization Analysis

Appendix Figure 1 presents the attention heatmaps generated by different segmentation models on the JSPH-ICA dataset. Red regions indicate highly activated areas contributing to the segmentation prediction, while blue regions represent weak responses. The visualization provides an intuitive analysis of whether the models can effectively focus on anatomically relevant vascular structures.

Compared with other methods, FUP-Net produces more concentrated and continuous responses along coronary artery regions, especially around thin vessels, curved structures, and bifurcation areas. This indicates that FUP-Net can better capture fine-grained vascular patterns while maintaining structural continuity. Such improved attention localization benefits from the frequency-aware heterogeneous modeling of FH-Mamba2, where global topology and local details are jointly preserved. Meanwhile, UPA further enhances semantic discrimination by reducing irrelevant activations and suppressing uncertain background responses. Overall, the visualization results demonstrate that FUP-Net learns more reliable vascular representations, which is consistent with the quantitative improvements reported in the experimental results.