RECONSTRUCTION FIRST: WHAT DRIVES RELIABLE CORONARY VESSEL RECOVERY?

Anonymous authors
Paper under double-blind review

ABSTRACT

Coronary vessel segmentation is structurally fragile: a few missed pixels can disconnect thin distal branches even when overall regional overlap is high. We ask whether a substantial part of this fragility can be addressed by better preserving and reconstructing fine local vessel evidence, and what additional benefit structure-aware modeling provides once reconstruction is strengthened. To disentangle these contributions, we develop and systematically decompose a reconstruction-first framework centered on a Context-Guided Vessel Reconstruction Decoder (CVRD), which preserves high-resolution local evidence, integrates multi-level contextual features to guide its interpretation, and progressively restores full-resolution vessel predictions. In controlled component ablations, weakening local detail extraction or replacing staged recovery produces the largest drops in both IoU and clDice among the tested reconstruction components. Across XACV, DCA1, and XCAD, the streamlined model yields consistent results over multiple seeds. Structural auxiliary objectives provide small but repeatable gains, but matched non-structural supervision and inference-time controls do not support a geometry-specific explanation for these improvements. Together, these results support reconstruction quality as a key contributor to coronary vessel recovery, while the evaluated structural supervision provides a smaller residual benefit once reconstruction is strong.

1 INTRODUCTION

Coronary vessel segmentation in X-ray coronary angiography (XCA) is a challenging thin-structure segmentation problem. Coronary arteries exhibit large variations in caliber, weak distal contrast, overlapping anatomy, foreshortening, and substantial background interference (Cervantes-Sanchez et al., 2019; Gao et al., 2022; He et al., 2025a). More importantly, errors that are small in area can have large structural consequences: a short false-negative gap may disconnect a distal branch, while a small false-positive bridge may create a spurious connection. High regional overlap therefore does not necessarily imply faithful recovery of the vascular tree. Reliable segmentation depends not only on regional accuracy, but also on preserving sufficiently fine local evidence to recover thin branches and their continuity.

Existing approaches address this fragility from complementary directions. High-resolution, temporal, and state-space models strengthen spatial recovery and contextual representation (Liu et al., 2022; He et al., 2025a; Zhang et al., 2026), whereas structure-aware methods introduce centerline, geometry, affinity, or topology information to encourage coherent vessel predictions (Shit et al., 2021; Shi et al., 2023; He et al., 2025b; Gao et al., 2026). However, these effects are often intertwined. An apparent structural improvement may result from better preservation of weak vessel evidence, generic benefits of auxiliary supervision, or the structural semantics themselves. Distinguishing these explanations is important for understanding what actually drives reliable vessel recovery.

We therefore study coronary segmentation from a *reconstruction-first* perspective. Our central question is how much structural fragility can be addressed by better preserving and reconstructing fine local vessel evidence, and what benefit remains from structure-aware learning once reconstruction is strong. To study this question, we introduce a Context-Guided Vessel Reconstruction Decoder (CVRD), which preserves high-resolution local evidence, integrates multi-level contextual features, and progressively restores full-resolution predictions. CVRD explicitly separates local-detail extrac-

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.

tion, feature fusion, contextual calibration, and staged recovery so that their contributions can be examined independently. Structural objectives are then introduced as auxiliary supervision on top of this reconstruction pathway rather than as inference-time conditioning.

Controlled experiments reveal an asymmetric picture. Weakening local-detail extraction or staged recovery produces the largest observed reductions in IoU and clDice among the tested reconstruction components. Across XACV, DCA1, and XCAD, structural auxiliary supervision provides smaller but repeatable improvements once reconstruction has been established. However, matched non-structural auxiliary supervision does not provide evidence that these gains require the intended geometric semantics. Together, these results support reconstruction quality as a key contributor to reliable coronary vessel recovery, with the evaluated structural objectives providing a more modest residual training benefit.

Our contributions are threefold:

- We formulate reliable coronary vessel recovery as a reconstruction-first problem and experimentally separate the contributions of fine-detail reconstruction and structural auxiliary supervision.
- We introduce CVRD, which preserves high-resolution local evidence, integrates multi-level contextual information, and progressively restores full-resolution predictions, with targeted controls for its major reconstruction functions.
- We provide a controlled empirical study across XACV, DCA1, and XCAD, showing that reconstruction components receive the clearest direct support, while structural auxiliary objectives yield smaller repeatable gains without a clear geometry-specific advantage under the tested matched control.

2 Related Work

2.1 Reconstruction and Fine-Vessel Recovery

Encoder–decoder architectures such as U-Net and nnU-Net form the basis of biomedical image segmentation (Ronneberger et al., 2015; Isensee et al., 2021). For thin vessels, however, repeated downsampling can suppress weak branches that occupy only a few pixels. FR-UNet addresses this issue by maintaining full-resolution representations for vessel segmentation (Liu et al., 2022), while recent coronary-specific methods improve recovery through temporal information or persistent high-resolution representations (He et al., 2025a; Zhang et al., 2026). These approaches motivate preserving fine spatial evidence while incorporating broader context.

Our work treats reconstruction itself as an object of controlled study. Rather than using the decoder as a monolithic component, CVRD separates local evidence preservation, multi-level feature fusion, contextual calibration, and staged full-resolution recovery, allowing their contributions to vessel recovery to be examined individually.

2.2 Structure-Aware Learning for Thin and Branching Vessels

Regional overlap objectives do not explicitly capture the geometry or connectivity of thin tubular structures. Prior work therefore introduces boundary-, centerline-, and topology-sensitive objectives, including boundary loss, clDice, centerline-boundary Dice, and connected-component Dice (Kervadec et al., 2019; Shit et al., 2021; Shi et al., 2024; Rougé et al., 2024). Complementary vessel-specific methods incorporate affinity representations, structure-aware state-space modeling, or joint temporal–geometric learning (Shi et al., 2023; He et al., 2025b; Gao et al., 2026).

These studies establish the usefulness of structural information, but its contribution is often intertwined with improvements in representation and reconstruction. We instead evaluate structural auxiliary supervision on top of an already strengthened reconstruction pathway and compare it with matched non-structural supervision. Our goal is not to argue that structural modeling is unnecessary, but to distinguish its residual contribution from gains that can already be explained by reconstruction quality or generic auxiliary training.

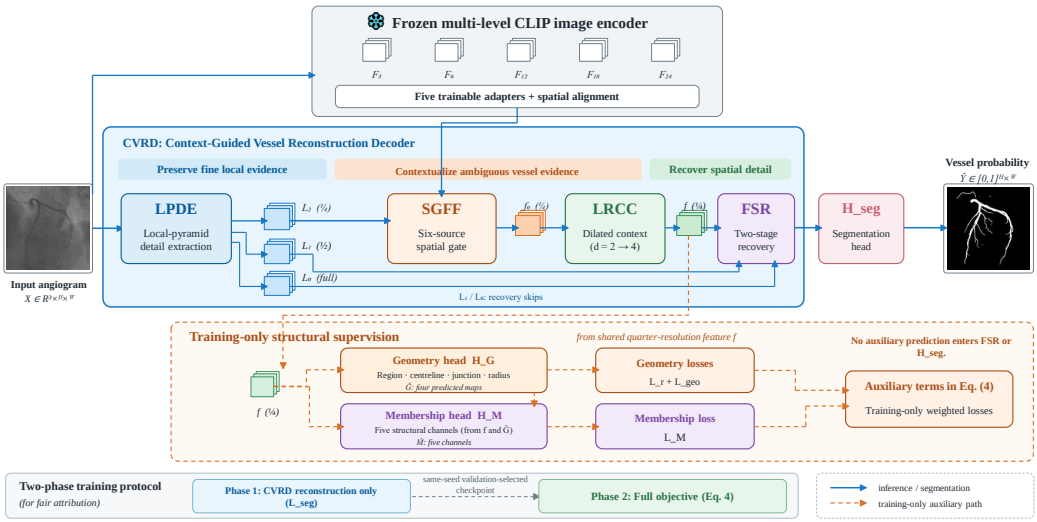


Figure 1: Overview of the reconstruction-first framework. CVRD preserves local evidence, integrates multi-level context, and progressively recovers the vessel prediction. Geometry and membership heads branch from the shared quarter-resolution representation f during training, but their predictions are not inputs to FSR or H_{seg}.

3 METHOD

3.1 RECONSTRUCTION-FIRST FORMULATION

Let $X \in \mathbb{R}^{3\times H\times W}$ denote the three-channel replicated version of a grayscale X-ray coronary angiogram, and let $Y \in \{0,1\}^{H\times W}$ denote its binary vessel mask. The model predicts a dense vessel probability map $\hat{Y} \in [0,1]^{H\times W}$.

We adopt a *reconstruction-first* formulation in which reconstruction refers to preserving and recovering vessel evidence across spatial scales for segmentation, rather than reconstructing the input image. The prediction pathway is therefore designed to (i) preserve fine local evidence, (ii) contextualize ambiguous vessel patterns, and (iii) progressively recover full-resolution predictions. Structure-aware learning is added only as auxiliary supervision on a shared intermediate representation; its predictions are not consumed by the segmentation recovery path. Figure 1 summarizes this separation.

3.2 CONTEXT-GUIDED VESSEL RECONSTRUCTION DECODER

The Context-Guided Vessel Reconstruction Decoder (CVRD) contains four functional components: local-pyramid detail extraction (LPDE), spatially gated feature fusion (SGFF), low-resolution context calibration (LRCC), and full-resolution skip recovery (FSR). Together, they implement the sequence *preserve local evidence* $\rightarrow$ *contextualize evidence* $\rightarrow$ *recover spatial detail*, as illustrated in Figure 2.

Local-pyramid detail extraction. LPDE operates directly on the angiogram and produces full-, half-, and quarter-resolution local features L_0, L_1, and L_2. While L_2 enters contextual fusion, L_1 and L_0 are retained as recovery skips so that fine evidence need not be reconstructed solely from compressed features.

In parallel, a frozen multi-level image encoder provides contextual features from selected intermediate layers. Trainable layer-specific adapters project these features into spatial maps used by SGFF. The exact encoder instantiation, selected layers, and adapter details are given in Appendix A.

Spatially gated feature fusion. SGFF combines the quarter-resolution local feature with five adapted encoder features. After projection and spatial alignment to the $H/4 \times W/4$ grid, let $\{U_s\}_{s=0}^{5}$

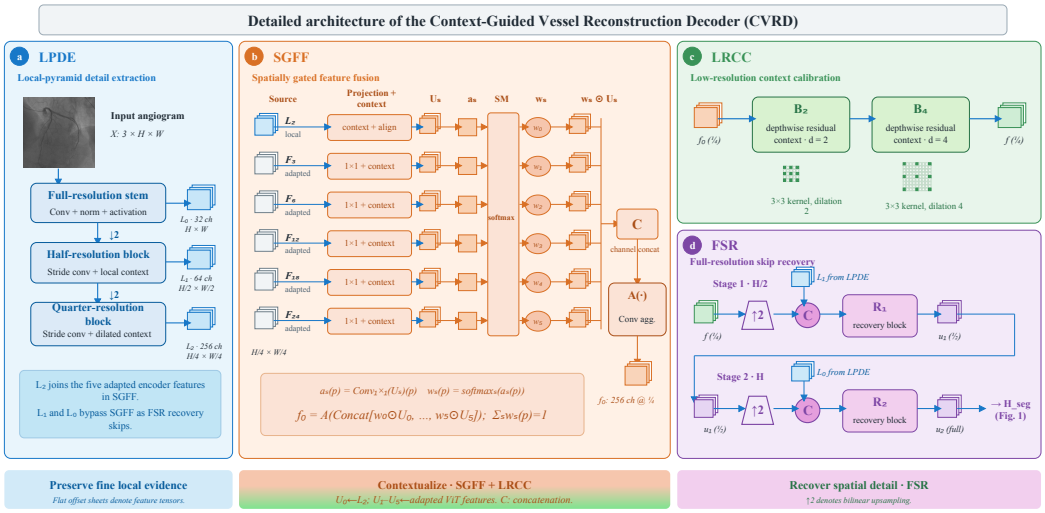


Figure 2: Detailed CVRD architecture. LPDE preserves multi-resolution local evidence; SGFF combines the quarter-resolution local feature with five adapted encoder features; LRCC enlarges contextual support; and FSR progressively restores spatial resolution using the retained L_1 and L_0 skips.

denote the six sources, where U_0 is local and $U_1, \ldots, U_5$ are encoder-derived. For each spatial location p, a lightweight scoring layer predicts source scores $a_s(p)$, which are normalized and used to aggregate the features:

$$w_s(p) = \frac{\exp a_s(p)}{\sum_{j=0}^{5} \exp a_j(p)}, \qquad f_0 = A(\mathrm{Concat}[w_0 \odot U_0, \ldots, w_5 \odot U_5]), \tag{1}$$

where $\odot$ denotes spatial weighting broadcast along channels and $A(\cdot)$ is the aggregation block. By construction, $\sum_{s=0}^{5} w_s(p) = 1$, yielding a location-dependent mixture of local and contextual evidence.

Low-resolution context calibration and recovery. LRCC processes f_0 using residual depthwise context blocks with dilation rates 2 and 4. FSR then restores spatial resolution in two stages, reintroducing the retained half- and full-resolution local features:

$$\begin{aligned} f &= \mathcal{B}_4(\mathcal{B}_2(f_0)), \\ u_1 &= \mathcal{R}_1(\mathrm{Concat}[\mathcal{U}_2(f), L_1]), \\ u_2 &= \mathcal{R}_2(\mathrm{Concat}[\mathcal{U}_2(u_1), L_0]), \\ \hat{Y} &= H_{\mathrm{seg}}(u_2). \end{aligned} \tag{2}$$

Here, $\mathcal{B}_d$ denotes a residual context block with dilation d, $\mathcal{U}_2(\cdot)$ denotes $2\times$ bilinear upsampling, and $\mathcal{R}_1, \mathcal{R}_2$ are the half- and full-resolution recovery blocks. The contextualized feature f is also shared with the auxiliary heads described next.

3.3 Structure-Aware Auxiliary Learning

Once the reconstruction pathway has been established, we introduce structural supervision through auxiliary heads operating on the shared quarter-resolution representation f. These heads affect representation learning during training but do not alter the recovery computation in Eq. (2). Figure 3 summarizes the deterministic target construction and the training-only auxiliary pathway.

Mask-derived structural targets. All auxiliary targets are generated deterministically from the binary vessel mask and require no additional annotations. The vessel foreground defines the region

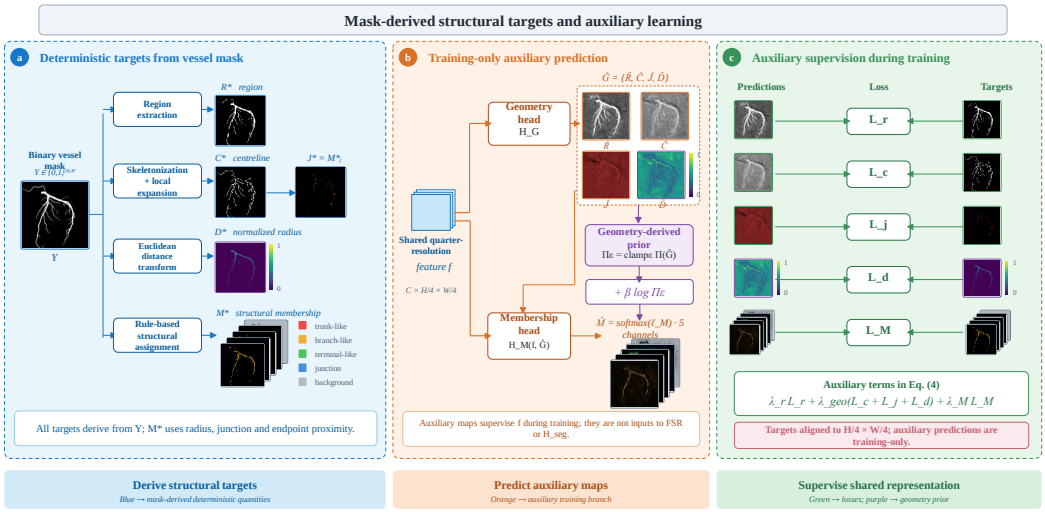


Figure 3: Mask-derived structural targets and training-only auxiliary learning. Region, centreline, junction, normalized-radius, and five-channel membership targets are derived deterministically from the binary vessel mask without additional manual annotation. The geometry and membership heads supervise the shared quarter-resolution representation f during training, but their predictions are not consumed by FSR or H_{seg}.

target R^*. Skeletonization provides the centreline target C^*, with a small local expansion used to reduce sensitivity to single-pixel displacement, while a normalized foreground Euclidean distance transform provides the radius target D^*. Junctions and endpoints are identified from local skeleton connectivity.

These quantities are further combined to construct a five-class categorical membership target

$$M^*(p) \in \{0,1\}^5, \qquad \sum_{k=1}^{5} M_k^*(p) = 1,$$

whose channels correspond to *trunk-like*, *branch-like*, *terminal-like*, *junction*, and *background*. The junction target is the corresponding membership channel,

$$J^* = M^*_{\text{junction}}.$$

Non-junction vessel pixels are separated using normalized radius and endpoint proximity. All targets are aligned to the quarter-resolution prediction grid before supervision. Exact thresholds, connectivity definitions, and morphological operations are reported in Appendix B.1.

Auxiliary geometry and membership prediction. A geometry head H_G predicts four quarter-resolution structural fields from f:

$$\hat{G} = \{\hat{R}, \hat{C}, \hat{J}, \hat{D}\},$$

corresponding to region, centreline, junction, and normalized radius.

A second head H_M predicts the five-class structural membership from the shared CVRD representation and the predicted structural fields. To expose the intended structural relationship explicitly during training, we also construct a geometry-derived categorical prior $\Pi(\hat{G})$:

$$\begin{aligned} \hat{G} &= \sigma(H_G(f)), \\ \ell_M &= H_M\Big(\text{Concat}[f, \hat{G}]\Big) + \beta \log\Big(\text{clamp}_\epsilon\big(\Pi(\hat{G})\big)\Big), \\ \hat{M} &= \text{softmax}(\ell_M). \end{aligned} \tag{3}$$

Here, $\hat{M} \in [0,1]^{5\times H/4\times W/4}$ is the predicted membership distribution, ϵ provides numerical stability, and β controls the contribution of the geometry-derived prior. The prior allocates predicted non-junction vessel mass among the trunk-like, branch-like, and terminal-like channels according to predicted radius while retaining separate junction and background components. Its exact construction is given in Appendix B.2.

Importantly, neither $\hat{G}$ nor $\hat{M}$ is passed to $\mathcal{R}_1$, $\mathcal{R}_2$, or H_{seg}. The auxiliary branch therefore modifies the shared representation through training supervision rather than defining a structure-conditioned inference path.

3.4 Training Objective and Reconstruction-First Protocol

We group the centreline, junction, and radius objectives into a structural geometry term,

$$\mathcal{L}_{\text{geo}} = \mathcal{L}_c + \mathcal{L}_j + \mathcal{L}_d. \tag{4}$$

The complete training objective is

$$\mathcal{L} = \mathcal{L}_{\text{seg}} + \lambda_r \mathcal{L}_r + \lambda_{\text{geo}} \mathcal{L}_{\text{geo}} + \lambda_M \mathcal{L}_M, \tag{5}$$

where $\mathcal{L}_{\text{seg}}$ supervises the final vessel prediction, $\mathcal{L}_r$ supervises region prediction, $\mathcal{L}_{\text{geo}}$ supervises centreline, junction, and radius prediction, and $\mathcal{L}_M$ supervises structural membership.

$\mathcal{L}_{\text{seg}}$ and $\mathcal{L}_r$ use BCE+Dice objectives. The centreline loss $\mathcal{L}_c$ also uses BCE+Dice, $\mathcal{L}_j$ combines focal and Dice losses for junction supervision, $\mathcal{L}_d$ uses foreground-masked Smooth-L_1 regression for normalized radius, and $\mathcal{L}_M$ is a class-balanced soft cross-entropy. We use

$$\lambda_r = 0.10, \qquad \lambda_{\text{geo}} = 0.05, \qquad \lambda_M = 0.20$$

throughout the final experiments. Further implementation details are given in Appendix B.3.

Reconstruction-first training protocol. Training proceeds in two consecutive phases. In the first phase, we optimize the reconstruction pathway using only $\mathcal{L}_{\text{seg}}$. For each dataset and model seed, the validation-selected reconstruction checkpoint then initializes a second phase in which the complete objective in Eq. (5) is optimized. The segmentation architecture remains unchanged and fully trainable during this continuation; only the auxiliary objectives are additionally activated.

This protocol first establishes a reconstruction-only solution and then measures the residual effect of auxiliary supervision from the corresponding same-seed initialization. In the principal experiments, CVRD is trained for 50 epochs and the CVRD+Aux continuation for 20 additional epochs. Because additional optimization itself may affect performance, we separately compare CVRD+Aux with a reconstruction-only continuation receiving the same 20-epoch budget in Section 4.3.

4 Experiments

Our experiments address three questions: (1) which reconstruction functions materially affect vessel recovery; (2) what benefit remains from structural auxiliary supervision once reconstruction is strong, and whether that benefit requires the intended structural semantics; and (3) whether the resulting streamlined model remains competitive under a controlled external benchmark.

4.1 Experimental Setup

We evaluate on XACV (Wu et al., 2025), DCA1 (Cervantes-Sanchez et al., 2019), and XCAD (Ma et al., 2021) using fixed train/validation/test splits of 141/35/45, 80/20/34, and 75/19/32 images, respectively. All angiograms are resized and center-cropped to 512×512. Mean per-image IoU is the primary regional metric and clDice (Shit et al., 2021) the main structure-sensitive metric.

CVRD uses a frozen OpenCLIP ViT-L/14-336 with features from layers $\{3, 6, 12, 18, 24\}$ and a decoder width of 256. The reconstruction-only model is trained for 50 epochs, after which its same-seed validation-selected checkpoint initializes a 20-epoch auxiliary continuation. Experiments use seeds 42, 222, and 3407 and report mean $\pm$ sample standard deviation. Optimization uses Adam with learning rate 10^{-4} and weight decay 10^{-4}. Additional preprocessing, augmentation, checkpoint-selection, and metric details are provided in Appendices A and C.

Table 1: CVRD component controls on XACV seed 42. Δ denotes the difference from complete CVRD in percentage points.

Setting	IoU	Δ	clDice	Δ
Complete CVRD	73.40	–	82.66	–
Basic local pyramid	72.16	-1.24	81.35	-1.31
Non-adaptive fusion	72.88	-0.52	81.91	-0.75
Identity context	73.23	-0.17	82.46	-0.20
Direct recovery	71.55	-1.85	81.69	-0.98

Table 2: Residual effect of structural auxiliary supervision over three seeds. Values are mean $\pm$ sample standard deviation; Δ denotes CVRD+Aux minus CVRD in percentage points.

	IoU (%)			clDice (%)		
Dataset	CVRD	+Aux	Δ	CVRD	+Aux	Δ
XACV	$73.44 \pm .06$	$74.11 \pm .15$	+0.67	$82.60 \pm .13$	$83.06 \pm .07$	+0.46
DCA1	$66.24 \pm .26$	$66.84 \pm .08$	+0.61	$86.37 \pm .36$	$86.83 \pm .52$	+0.46
XCAD	$67.31 \pm .41$	$67.58 \pm .15$	+0.26	$85.35 \pm .31$	$85.43 \pm .01$	+0.08

4.2 Reconstruction Component Attribution

We first isolate the reconstruction pathway before introducing structural auxiliary supervision. Table 1 replaces one CVRD function at a time with a shape-compatible control. All four controls reduce both IoU and clDice. The largest seed-42 IoU degradations occur when staged recovery is replaced by direct recovery (-1.85 pp) and when the local pyramid is simplified (-1.24 pp). The corresponding clDice degradations are -0.98 and -1.31 pp, respectively. Non-adaptive fusion produces smaller reductions, while bypassing LRCC gives the smallest effect in the seed-42 decomposition.

Because the full component decomposition is a targeted single-seed study, we additionally repeat the LRCC control over seeds 42, 222, and 3407. Relative to the identity-context variant, complete CVRD improves IoU/clDice by $+0.17/+0.20$, $+1.09/+0.90$, and $+0.30/+0.19$ pp for the three seeds, respectively. The resulting paired mean differences are 0.52 ± 0.50 pp IoU and 0.43 ± 0.41 pp clDice.

These results support CVRD as a coordinated reconstruction design. Among the tested seed-42 controls, the clearest degradations arise from weakening local-detail extraction or replacing staged spatial recovery, while the repeated LRCC experiment provides additional evidence that the context-calibration stage is also beneficial. We therefore use the component results to identify supported reconstruction functions rather than to claim a universal ranking among them.

4.3 Residual Benefit of Structural Auxiliary Supervision

We next measure the residual effect of structural auxiliary supervision after the reconstruction pathway has already been established. For each seed, CVRD+Aux is initialized from the corresponding validation-selected reconstruction-only CVRD checkpoint.

As shown in Table 2, mean IoU increases by 0.67, 0.61, and 0.26 pp on XACV, DCA1, and XCAD, respectively, while the corresponding clDice differences are $+0.46$, $+0.46$, and $+0.08$ pp. The dataset-level direction is therefore positive for both metrics, although the effect is substantially smaller than the largest reconstruction-component degradations in Section 4.2.

Because CVRD+Aux receives an additional 20 epochs of optimization, we also compare it with a reconstruction-only continuation initialized from the same seed-specific checkpoints and trained for the same additional budget. Under this matched continuation, CVRD+Aux retains small positive mean IoU differences on all three datasets, while the structure-sensitive differences are mixed. The complete matched-budget results, including clDice and filtered ccDice_{32}, are reported in Appendix D.3. We therefore interpret the auxiliary objectives as providing a modest residual training benefit rather than a uniformly stronger structural effect.

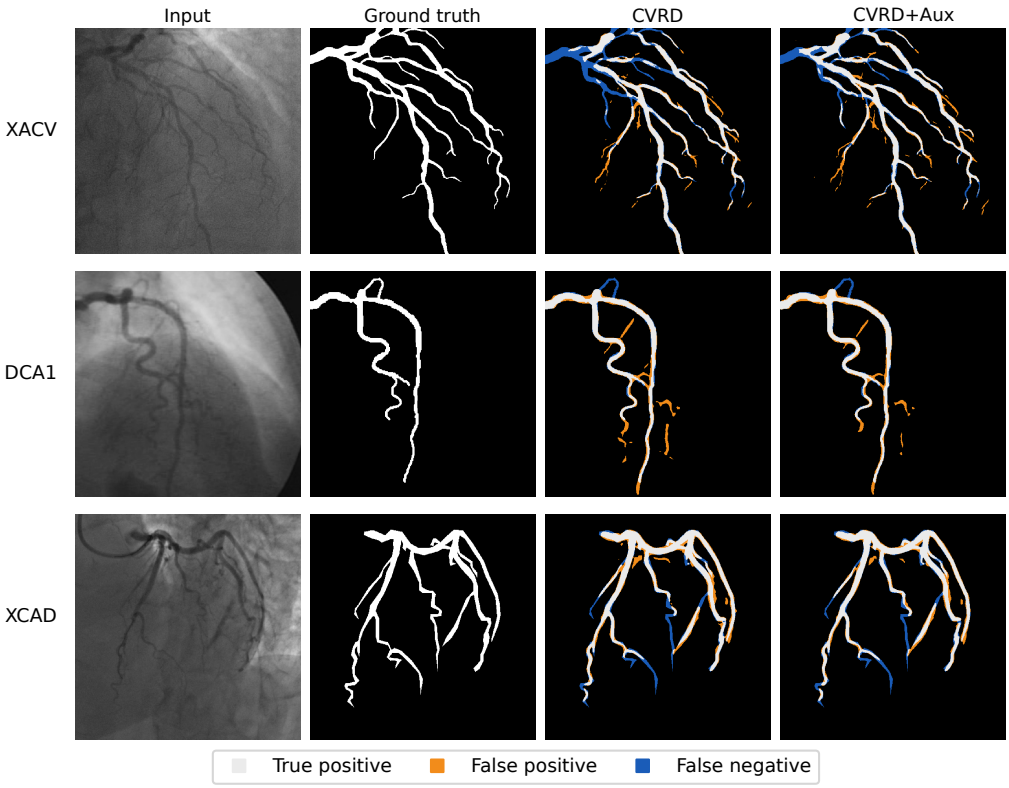


Figure 4: Qualitative CVRD and CVRD+Aux predictions on retained-test examples. White, orange, and blue denote true positives, false positives, and false negatives, respectively.

Table 3: XACV seed-42 auxiliary-objective screen with a matched 30-epoch continuation budget. Δ is relative to reconstruction-only continuation.

Setting	IoU	Δ	clDice	Δ
CVRD continuation	73.65	–	82.88	–
Geometry losses only	73.76	+0.10	82.84	-0.04
Geometry + membership	73.96	+0.31	82.96	+0.09

Figure 4 shows representative paired predictions. The examples illustrate local error patterns rather than their prevalence; the quantitative conclusions rely on the multi-seed comparisons.

Complete seed-level Dice, cbDice, and secondary connectivity results are reported in Appendix D.

4.4 Auxiliary-Objective Screen

We further examine the auxiliary objectives using an XACV seed-42 screen in which all configurations start from the same validation-selected CVRD checkpoint and receive a matched 30-epoch continuation budget. As shown in Table 3, geometry losses alone produce only marginal differences from reconstruction-only continuation ($+0.10$ pp IoU, -0.04 pp clDice), whereas adding membership supervision increases the differences to $+0.31$ pp IoU and $+0.09$ pp clDice.

The combined auxiliary configuration therefore shows the clearest improvement in this screen. Because this is a single-seed 30-epoch diagnostic, distinct from the final 20-epoch three-seed protocol, we do not treat it as an additive decomposition of the final model. Additional protocol details are provided in Appendix D.4.

Table 4: Controlled seed-42 external benchmark. Dataset columns report mean per-image IoU (%). Macro IoU, Dice, and clDice are unweighted means across XACV, DCA1, and XCAD.

Method	XACV	DCA1	XCAD	mIoU	mDice	mclDice
CVRD+Aux (ours)	**73.99**	**66.84**	**67.61**	**69.48**	**81.83**	**85.16**
nnU-Net v2	71.45	65.53	62.87	66.61	79.73	82.53
MDSA-UNet	68.22	59.20	57.19	61.53	75.84	77.62
U-KAN	67.07	60.48	56.81	61.45	75.79	79.38
MedCLIPSeg	65.64	60.54	57.93	61.37	75.86	78.65
MK-UNet	68.21	59.63	55.98	61.27	75.59	78.70
FR-UNet	69.51	61.43	51.53	60.82	75.01	76.86
HiFormer	64.90	58.83	56.08	59.94	74.50	76.71
Rolling-Unet	64.83	55.90	55.32	58.68	73.65	78.70
MISSFormer-512	63.10	57.82	52.61	57.84	72.89	72.64

4.5 Controls for Geometry-Specific Explanations

The positive auxiliary effect does not by itself show that the intended structural semantics are responsible for the gain. We therefore compare the geometry-derived auxiliary targets with a capacity- and budget-matched control in which the same auxiliary heads are trained on repeated vessel-occupancy targets.

On the XACV seed-42 validation screen, geometry-derived supervision differs from the repeated-occupancy control by only -0.027 pp in IoU and $+0.011$ pp in clDice. Both configurations select epoch 6. This comparison provides no clear evidence that the intended geometry-specific target semantics outperform matched generic auxiliary supervision in the tested setting. Because this experiment is a validation-only, single-seed diagnostic, we use it to bound the interpretation of the auxiliary gain rather than to make a broader claim about structure-aware learning. Additional staged controls and complete diagnostic results are reported in Appendix D.

4.6 Controlled External Benchmark

Finally, we evaluate whether the streamlined reconstruction-first model remains competitive beyond the internal attribution experiments. We compare CVRD+Aux with established and recent segmentation baselines using common file-level splits, validation-IoU checkpoint selection, and retained-test evaluation. All entries in Table 4 use a single seed-42 run, a fixed probability threshold of 0.5, and no test-time augmentation or ensembling. Architecture-specific initialization and optimization settings are retained where appropriate; complete implementation details are provided in Appendix E.

CVRD+Aux obtains IoUs of 73.99%, 66.84%, and 67.61% on XACV, DCA1, and XCAD, respectively. Its corresponding three-dataset macro IoU, Dice, and clDice are 69.48%, 81.83%, and 85.16%. Under the same evaluation protocol, nnU-Net v2 obtains 66.61%, 79.73%, and 82.53%, respectively.

This benchmark provides complementary evidence that the final streamlined model remains competitive across both regional and structure-sensitive metrics. However, because the external comparison is a single seed-42 benchmark and does not match training compute across architectures, it is not used to attribute the source of the improvement. The reconstruction and auxiliary conclusions of this paper instead rely on the controlled internal analyses in Sections 4.2–4.5.

5 Conclusion

We studied coronary vessel segmentation from a reconstruction-first perspective and introduced CVRD to preserve fine local evidence, integrate multi-level context, and progressively restore full-resolution predictions. Controlled ablations show the largest observed degradations when local-detail extraction or staged recovery is weakened, whereas structural auxiliary supervision provides smaller repeatable gains across XACV, DCA1, and XCAD. Matched continuation and non-structural auxiliary-target controls do not support a geometry-specific explanation for these gains. Overall, reconstruction quality emerges as a key contributor to reliable vessel recovery, while the evaluated structural objectives provide a modest residual training benefit.

AI USE STATEMENT

ChatGPT (GPT-5.6) was used for language polishing and stylistic refinement to improve the clarity and readability of the manuscript. All AI-assisted text was reviewed and revised as necessary by the authors, who take full responsibility for the accuracy and final content of the work.

REPRODUCIBILITY STATEMENT

The main paper specifies the dataset partitions, input resolution, model configuration, training schedule, optimizer, random seeds, checkpoint selection, and evaluation protocol. The Appendix provides additional preprocessing and augmentation details, exact structural-target and prior construction, auxiliary-loss implementation, per-seed results, secondary metrics, and supplementary controls. Principal experiments use fixed file-level splits and seeds 42, 222, and 3407, with checkpoints selected by validation mean per-image IoU before a single retained-test evaluation. External baselines use the same data partitions and evaluation conventions, although training compute is not matched across architectures. Known limitations concerning archived geometry-cache provenance are also reported in the Appendix.

REFERENCES

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. doi: 10.3390/app9245507. URL https://www.mdpi.com/2076-3417/9/24/5507.

Yunlong Gao, Yuanyuan Wang, Danni Ai, Feixu Shang, Hong Song, Jingfan Fan, Deqiang Xiao, Yining Wang, and Jian Yang. Iterative joint learning integrating temporal and geometric information for vessel segmentation in x-ray coronary angiography. *Medical Physics*, 53(2):e70317, 2026. doi: 10.1002/mp.70317.

Zijun Gao, Lu Wang, Reza Soroushmehr, Alexander Wood, Jonathan Gryak, Brahmajee Nallamothu, and Kayvan Najarian. Vessel segmentation for X-ray coronary angiography using ensemble methods with deep learning and filter-based features. *BMC Medical Imaging*, 22(1):10, 2022. doi: 10.1186/s12880-022-00734-4.

Haorui He, Abhirup Banerjee, Robin P. Choudhury, and Vicente Grau. Deep learning based coronary vessels segmentation in X-ray angiography using temporal information. *Medical Image Analysis*, 102:103496, 2025a. doi: 10.1016/j.media.2025.103496.

Yuanqing He, Zhenhuan Lyu, Yayue Mai, Si Li, and Chen-kai Hu. VM-CAGSeg: A vessel structure-aware state space model for coronary artery segmentation in angiography images. *Frontiers in Medicine*, 12:1661680, 2025b. doi: 10.3389/fmed.2025.1661680.

Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, 2021. doi: 10.1038/s41592-020-01008-z.

Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, and Ismail Ben Ayed. Boundary loss for highly unbalanced segmentation. In *International Conference on Medical Imaging with Deep Learning*, Proceedings of Machine Learning Research, pp. 285–296, 2019.

Wentao Liu, Huihua Yang, Tong Tian, Zhiwei Cao, Xipeng Pan, Weijin Xu, Yang Jin, and Feng Gao. Full-resolution network and dual-threshold iteration for retinal vessel and coronary angiograph segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(9):4623–4634, 2022. doi: 10.1109/JBHI.2022.3188710.

Yuxin Ma, Yang Hua, Hanming Deng, Tao Song, Hao Wang, Zhengui Xue, Heng Cao, Ruhui Ma, and Haibing Guan. Self-supervised vessel segmentation via adversarial learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7536–7545,

2021. doi: 10.1109/ICCV48922.2021.00744. URL https://openaccess.thecvf.com/content/ICCV2021/html/Ma_Self-Supervised_Vessel_Segmentation_via_Adversarial_Learning_ICCV_2021_paper.html.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, volume 9351 of *Lecture Notes in Computer Science*, pp. 234–241. Springer, 2015. doi: 10.1007/978-3-319-24574-4_28. URL https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28.

Pierre Rougé, Olivier Merveille, and Nicolas Passat. ccDice: A topology-aware dice score based on connected components. In *Topology- and Graph-Informed Imaging Informatics*, volume 15239 of *Lecture Notes in Computer Science*, pp. 11–21. Springer, 2024. doi: 10.1007/978-3-031-73967-5_2.

Pengcheng Shi, Jun Hu, Yizhou Yang, Zikang Gao, Wei Liu, and Tao Ma. Centerline boundary dice loss for vascular segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume 15008 of *Lecture Notes in Computer Science*, pp. 46–56. Springer, 2024. doi: 10.1007/978-3-031-72111-3_5.

Tianyi Shi, Xiaohuan Ding, Wei Zhou, Feng Pan, Zengqiang Yan, Xiang Bai, and Xin Yang. Affinity feature strengthening for accurate, complete and robust vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, 27(8):4006–4017, 2023. doi: 10.1109/JBHI.2023.3274789.

Suprosanna Shit, Johannes C. Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien P. W. Pluim, Ulrich Bauer, and Bjoern H. Menze. clDice: A novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16560–16569, 2021. doi: 10.1109/CVPR46437.2021.01629. URL https://openaccess.thecvf.com/content/CVPR2021/html/Shit_clDice_-_A_Novel_Topology-Preserving_Loss_Function_for_Tubular_Structure_CVPR_2021_paper.html.

Chun-Hung Wu, Shih-Hong Chen, Chih-Yao Hu, Hsin-Yu Wu, Kai-Hsin Chen, Yu-You Chen, Chih-Hai Su, Chih-Kuo Lee, and Yu-Lun Liu. DeNVeR: Deformable neural vessel representations for unsupervised video vessel segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15682–15692, 2025. doi: 10.1109/CVPR52734.2025.01462. URL https://openaccess.thecvf.com/content/CVPR2025/html/Wu_DeNVeR_Deformable_Neural_Vessel_Representations_for_Unsupervised_Video_Vessel_Segmentation_CVPR_2025_paper.html.

Xiuhan Zhang, Peng Lu, Zongsheng Zheng, and Wenhui Li. HR-UMamba++: A high-resolution multi-directional mamba framework for coronary artery segmentation in X-ray coronary angiography. *Fractal and Fractional*, 10(1):43, 2026. doi: 10.3390/fractalfract10010043.

APPENDIX

A Additional Reproducibility and Experimental Protocol Details

The main paper reports the dataset sizes, input resolution, encoder configuration, optimizer, training duration, model seeds, and test-time conventions. This section records the remaining implementation details needed to reproduce the reported experiments.

A.1 Split and Preprocessing Details

The fixed file-level partitions are generated with split seed 20260723 for XACV and DCA1 and split seed 20260823 for the current XCAD repartition. Model seeds 42, 222, and 3407 do not alter these partitions.

The same sampled spatial transformation is applied to the angiogram, binary mask, and cached geometry fields. XACV uses no random flips. DCA1 and XCAD use independent horizontal and vertical flips, each with probability 0.5 during training.

The retained binary-mask preprocessing uses bilinear resizing followed by tensor conversion. Cache-derived discrete structural fields are resized using nearest-neighbour interpolation, whereas the continuous radius field is resized bilinearly. Evaluation uses the strict probability rule > 0.5.

A.2 Continuation and Checkpoint Details

For the continuation experiments described in the main paper, training uses micro-batches of 4 with gradient accumulation over two steps, giving the reported effective batch size of 8.

A continuation checkpoint is eligible for selection from epoch 1 onward. The checkpoint with the highest validation mean per-image IoU is retained, and only that selected checkpoint is evaluated once on the held-out test partition.

The current XCAD repartition is distinct from older XCAD experiments. Results obtained under earlier XCAD partitions are not mixed with the experiments reported in this paper.

B Structural Target and Auxiliary Implementation Details

The main paper defines the structural targets, auxiliary prediction heads, training objective, and their relationship to the CVRD recovery pathway. Here we provide the exact target-generation rules and the construction of the geometry-derived membership prior.

B.1 Exact Geometry-Target Construction

The centreline is obtained using two-dimensional skeletonization of the binary vessel mask. The centreline supervision target includes a small local expansion to reduce sensitivity to single-pixel displacement.

The normalized-radius field is computed from the foreground Euclidean distance transform and normalized by its within-image maximum.

Skeleton connectivity is evaluated in a local 3×3 neighbourhood. A skeleton pixel with at least three neighbouring skeleton pixels is treated as a junction candidate. The resulting junction region is dilated using an elliptical kernel with radius two pixels and then restricted to the vessel foreground. A skeleton pixel with one neighbouring skeleton pixel is treated as an endpoint.

For the five-channel structural membership target, non-junction foreground pixels are assigned using normalized radius and endpoint proximity. The cache-generator defaults are:

$$\tau_{\mathrm{terminal}} = 0.15, \qquad \tau_{\mathrm{trunk}} = 0.50,$$

with an endpoint-distance threshold of two pixels and a junction-dilation radius of two pixels.

Archived membership fields are mapped to the channel order

$$\{\text{trunk-like}, \text{branch-like}, \text{terminal-like}, \text{junction}, \text{background}\}.$$

The archived field named `mask_low_contrast` is used as the branch-like field. This name is inherited from the engineering cache format and should not be interpreted as an independently measured low-contrast anatomical category.

B.2 Geometry-Derived Membership Prior

Let

$$\hat{G} = \{\hat{R}, \hat{C}, \hat{J}, \hat{D}\}$$

denote the four predicted structural fields introduced in the main paper.

We first construct a soft vessel-support quantity

$$V = \text{clip}\left(\hat{R} + 0.25\hat{C}(1-\hat{R}), 10^{-4}, 1-10^{-4}\right). \tag{6}$$

The predicted normalized radius is then converted into three soft assignments:

$$a_j = \text{softmax}_j\left[-\left(\frac{\hat{D}-\mu_j}{0.22}\right)^2\right], \qquad (\mu_1, \mu_2, \mu_3) = (0.82, 0.50, 0.18), \tag{7}$$

where j indexes the trunk-like, branch-like, and terminal-like channels.

The five-channel geometry-derived prior is

$$\Pi(\hat{G}) = \text{Norm}_5\left[V(1-\hat{J})a_1,\ V(1-\hat{J})a_2,\ V(1-\hat{J})a_3,\ V\hat{J},\ 1-V\right]. \tag{8}$$

The resulting categorical prior enters the membership logits through the scaled log-prior term in Eq. (3). The prior scale is trainable and clipped to $[0, 4]$ during the forward pass.

B.3 Additional Loss-Implementation Details

The main paper specifies the loss family associated with each auxiliary prediction. Two implementation details are relevant here.

First, the Smooth-L_1 radius regression loss is evaluated only over foreground vessel pixels. Second, the five-channel membership objective is implemented as a class-balanced soft cross-entropy: for each class present in the resized target, the loss is first averaged within that class and then averaged across the classes present in the sample.

C Evaluation Protocol and Metrics

C.1 Primary Metrics

For a binarized prediction, let TP, FP, and FN denote the true-positive, false-positive, and false-negative pixel counts. IoU and Dice are computed as

$$\text{IoU} = \frac{TP}{TP + FP + FN + 10^{-7}},$$

and

$$\text{Dice} = \frac{2TP}{2TP + FP + FN + 10^{-7}}.$$

Metrics are first computed independently for each image and then averaged arithmetically over the corresponding test set. We do not pool confusion counts across the entire test set.

For clDice (Shit et al., 2021), skeleton precision measures the fraction of predicted-skeleton pixels contained in the target mask, whereas skeleton recall measures the fraction of target-skeleton pixels contained in the prediction. Their harmonic mean gives the per-image clDice. The implementation uses two-dimensional skeletonization and 10^{-7} denominator smoothing.

Table 5: Individual retained-test results for the validation-selected CVRD and CVRD+Aux checkpoints. $ccDice_{32}$ denotes the filtered connectivity diagnostic defined in Section C.2.

Dataset	Seed	Model	IoU	Dice	clDice	cbDice	$ccDice_{32}$
XACV	42	CVRD	73.40	84.47	82.66	77.44	18.86
		CVRD+Aux	73.99	84.88	82.97	77.60	20.53
	222	CVRD	73.51	84.55	82.68	77.49	24.53
		CVRD+Aux	74.06	84.93	83.09	77.70	21.46
	3407	CVRD	73.41	84.46	82.45	77.24	25.83
		CVRD+Aux	74.28	85.07	83.11	77.88	25.16
DCA1	42	CVRD	66.22	79.60	85.96	72.37	39.89
		CVRD+Aux	66.84	80.05	87.06	73.39	46.60
	222	CVRD	65.99	79.42	86.60	71.83	42.32
		CVRD+Aux	66.92	80.10	87.19	73.23	44.62
	3407	CVRD	66.50	79.78	86.55	73.44	51.68
		CVRD+Aux	66.77	79.99	86.24	72.77	49.58
XCAD	42	CVRD	66.99	80.10	85.00	75.61	44.79
		CVRD+Aux	67.61	80.55	85.45	76.27	46.31
	222	CVRD	67.18	80.23	85.46	75.82	38.88
		CVRD+Aux	67.41	80.40	85.43	76.10	39.44
	3407	CVRD	67.77	80.68	85.59	76.86	52.98
		CVRD+Aux	67.71	80.63	85.43	76.36	51.95

C.2 Secondary Diagnostics

The retrospective cbDice diagnostic (Shi et al., 2024) uses a hard-mask, two-dimensional, division-based cl-MIN-Dice implementation with distance/radius weighting.

For the retrospective filtered $ccDice_{32}$ diagnostic (Rougé et al., 2024), connected foreground components smaller than 32 pixels are first removed independently from the prediction and target using eight-connectivity. Components are then greedily paired when their overlap reaches at least half of the source-component area. Directional match counts are normalized by the corresponding total numbers of components.

C.3 Evaluation Conventions

All metrics are evaluated on predictions binarized using the strict rule > 0.5. IoU, Dice, clDice, cbDice, and filtered $ccDice_{32}$ are computed per image and then averaged. The secondary diagnostics are not used for checkpoint selection and are not training objectives.

D Additional Internal Results

D.1 Full Seed-Level Paired Results

Table 5 provides the individual retained-test results underlying the three-seed CVRD and CVRD+Aux summaries in the main paper. Each pair uses the same data partition and model seed. All values are percentages.

D.2 Secondary-Metric Summary

IoU and clDice summaries are reported in the main paper. Table 6 therefore reports only the additional Dice, cbDice, and filtered $ccDice_{32}$ statistics over the same three model seeds.

The secondary metrics do not show a uniformly positive connectivity pattern. In particular, the mean filtered $ccDice_{32}$ on XACV decreases from 23.07 to 22.38 despite the positive IoU and clDice differences reported in the main paper.

D.3 Matched-Budget Continuation Control

To separate auxiliary supervision from the effect of additional optimization, we continue the same seed-specific CVRD checkpoints for 20 epochs without activating the auxiliary objectives. CVRD+Aux

Table 6: Secondary retained-test metrics over seeds 42, 222, and 3407. Values are mean ± sample standard deviation.

Dataset	Metric	CVRD	CVRD+Aux
XACV	Dice	84.49 ± 0.05	84.96 ± 0.10
	cbDice	77.39 ± 0.13	77.73 ± 0.14
	ccDice$_{32}$	23.07 ± 3.70	22.38 ± 2.45
DCA1	Dice	79.60 ± 0.18	80.05 ± 0.05
	cbDice	72.55 ± 0.82	73.13 ± 0.32
	ccDice$_{32}$	44.63 ± 6.23	46.93 ± 2.50
XCAD	Dice	80.34 ± 0.31	80.53 ± 0.12
	cbDice	76.09 ± 0.67	76.24 ± 0.13
	ccDice$_{32}$	45.55 ± 7.08	45.90 ± 6.26

Table 7: Matched 20-epoch continuation control. Values are mean ± sample standard deviation over three seeds. Δ denotes CVRD+Aux minus the reconstruction-only continuation, in percentage points.

Dataset	Metric	CVRD continuation	CVRD+Aux	Paired Δ
XACV	IoU	73.82 ± 0.15	74.11 ± 0.15	$+0.30 \pm 0.06$
	clDice	82.85 ± 0.02	83.06 ± 0.07	$+0.21 \pm 0.07$
	ccDice$_{32}$	21.88 ± 1.45	22.38 ± 2.45	$+0.50 \pm 1.02$
DCA1	IoU	66.68 ± 0.22	66.84 ± 0.08	$+0.16 \pm 0.27$
	clDice	86.95 ± 0.26	86.83 ± 0.52	-0.12 ± 0.31
	ccDice$_{32}$	47.61 ± 1.20	46.93 ± 2.50	-0.67 ± 1.39
XCAD	IoU	67.49 ± 0.15	67.58 ± 0.15	$+0.09 \pm 0.04$
	clDice	85.33 ± 0.09	85.43 ± 0.01	$+0.11 \pm 0.10$
	ccDice$_{32}$	44.77 ± 6.50	45.90 ± 6.26	$+1.13 \pm 0.89$

uses the same initialization and continuation budget. Table 7 reports the resulting IoU, clDice, and filtered ccDice$_{32}$ statistics over seeds 42, 222, and 3407.

Mean IoU remains positive for CVRD+Aux on all three datasets under the matched continuation. However, the structure-sensitive diagnostics are not uniformly positive: DCA1 decreases by 0.12 pp in clDice and 0.67 pp in filtered ccDice$_{32}$. The control therefore limits a simple additional-training explanation for the IoU improvement, while providing no evidence for a uniform connectivity advantage from the auxiliary objectives.

D.4 Protocol Details for the Auxiliary-Objective Screen

The auxiliary-objective screen in Section 4.4 is a targeted XACV seed-42 diagnostic. All three configurations are initialized from the same validation-selected 50-epoch CVRD checkpoint and receive 30 additional training epochs. The reconstruction-only continuation therefore serves as the same-budget reference for the two auxiliary configurations.

This screen is intentionally separate from the final CVRD+Aux protocol. In particular, the staged screen uses a membership weight of $\lambda_M = 0.10$, whereas the final 20-epoch CVRD+Aux experiments use $\lambda_M = 0.20$. The screen should therefore be interpreted as evidence about the relative behavior of the tested auxiliary configurations under that diagnostic protocol, rather than as an additive decomposition of the final model's three-seed performance.

D.5 Specificity-Control Protocol Note

The geometry-derived versus repeated-occupancy comparison reported in Section 4.5 is a validation-only diagnostic. Neither configuration was subsequently evaluated on the retained test partition for that screen. Accordingly, it is used only to test the geometry-specific interpretation of the auxiliary signal and is not included in the retained-test performance comparisons.

E External Benchmark Details

E.1 Benchmark Implementation Notes

The external baselines use the same dataset-specific file lists and binary evaluation convention as the benchmark reported in the main paper. Each method is selected by validation mean per-image IoU and evaluated once on the corresponding retained test partition.

Architecture-specific initialization, optimizer settings, and convergence adjustments are retained where needed. Consequently, the common data partitions and evaluation protocol should not be interpreted as matched training compute across architectures.

The MDSA-UNet result is obtained from a local architecture reproduction based on the available specification rather than from a recovered official training configuration. The external comparison is therefore intended as a controlled reference benchmark rather than a compute-matched statistical comparison.

E.2 Qualitative Benchmark Examples

Figure 5 provides qualitative comparisons between CVRD+Aux and selected external methods under the same seed-42 evaluation setting as the benchmark in the main paper. Two examples per dataset are selected according to ground-truth properties rather than relative model performance.

F Cached-Geometry Provenance

The retained DCA1 and XCAD manifests record the intended cache-generation settings, but they also indicate that pre-existing cache files were skipped during those invocations. The original XACV cache-generation manifest was not retained.

Consequently, exact parameter equivalence of every archived cache file cannot be independently established from the retained manifests alone. The rules and default parameters reported in Section B.1 describe the intended cache-generation procedure used by the current implementation.

This provenance limitation should be distinguished from the deterministic use of the already archived auxiliary targets during the reported training runs.

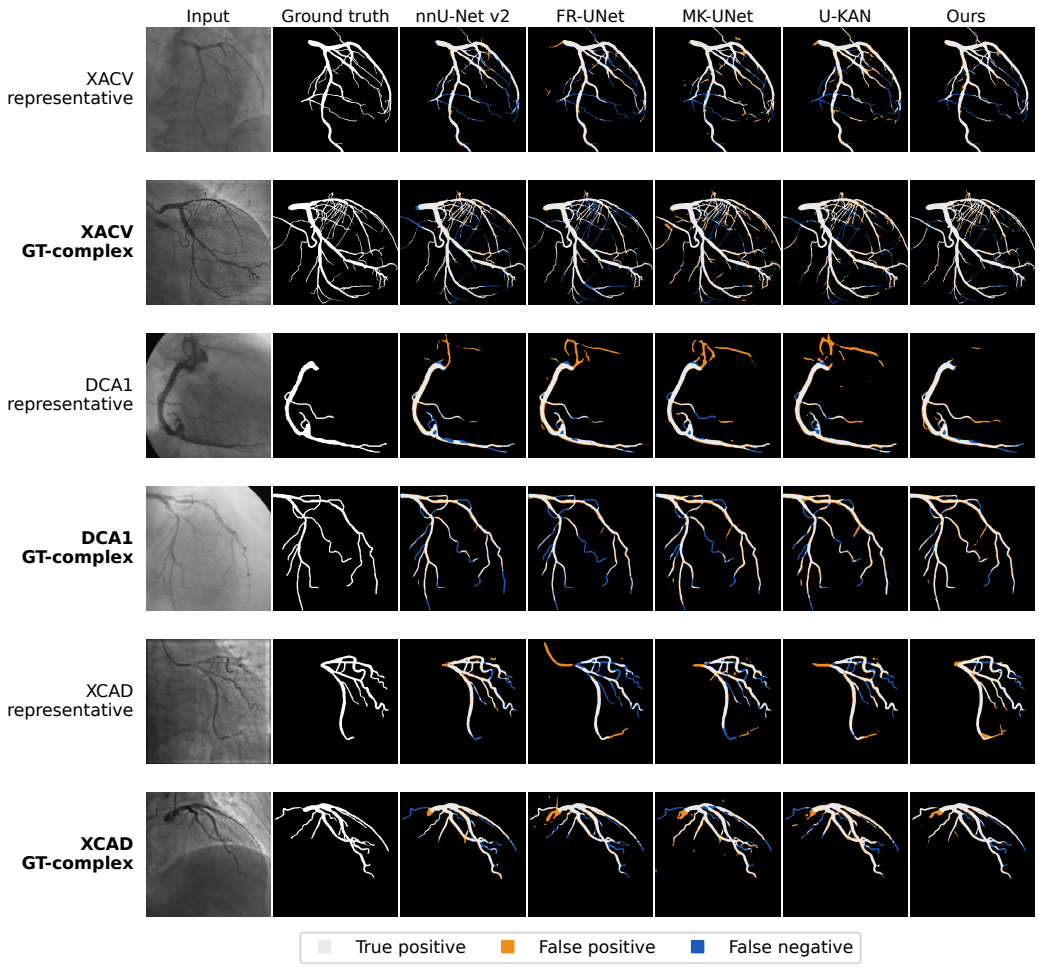


Figure 5: Qualitative comparison of CVRD+Aux with selected reproduced seed-42 external methods on XACV, DCA1, and XCAD. White denotes true-positive vessel pixels, orange denotes false positives, and blue denotes false negatives under the shared probability threshold > 0.5. These examples illustrate error patterns rather than their dataset-level prevalence.