Journal of Intelligent Information and Communication

Research Paper

Topology-Preserving Coronary Artery Segmentation in CTA Volumes Using an nnFormer-CRF Model

Ivan Jelen[1, *] and Dalibor Sedláček[1]

[1] Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, 61669, Czechia

*Corresponding author: jelen.iv@feec.vutbr.cz

Abstract. Coronary artery segmentation from CT angiography (CTA) is inherently difficult because the voxels of the distal vessels are small, there are changes in lumen contrast along the tree, and discontinuous or fused predictions occur due to calcified plaques. This paper develops a topology-preserving nnFormer-CRF model for full 3D coronary extraction. The encoder uses a moving volume window attention mechanism to capture long-range vascular context, and the topological branches adjust the soft centerlines and branch connections at three resolutions. Then, the graph conditional random field uses intensity, spatial, and tangential consistency potentials to fine-tune the logits, with these potentials constrained within an adaptive coronary region. To validate the model, a five-fold patient-level study was conducted, involving 420 CTA volumes from three centers. Its Dice coefficient is 0.912, the centerline Dice coefficient is 0.936, the mean symmetric surface distance is 0.31 mm, and the connection error per case is 0.71. Compared with the nnFormer backbone, the proposed method improved the recall of distal branches to 0.889 and reduced false vascular links by 34.8%. The best-case scenario of a large calcification had a Dice score of 0.79 after improvement. Inference for the resampled volume required 1.8s, and the refinement stage added an additional 0.23s. According to the results, Transformer context and structured anatomy refinement are complementary: the former can recover weak branches, while the latter avoids unreasonable shortcuts by reconstructing locally interrupted centerlines. The Framework offers good support for topology-sensitive coronary model building and quantitative CTA analysis.

Keywords: *Coronary CTA, Topology-Preserving Segmentation, nnFormer, Conditional Random Field, Vascular Connectivity*

Received on 12 November 2025, Accepted on 03 March 2026, Published on 16 March 2026

Copyright © 2026 Author, licensed to JIIC. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

Introduction

Coronary computed tomography angiography (CTA) can non-invasively show the shape and contents of the coronary arteries to determine whether there is narrowing and guide treatment plans [1]. Automatic segmentation is still an open problem because the target is a sparse, branched structure with a diameter difference of more than an order of magnitude [2]. The distal part may only have two or three voxels after conventional reconstruction, so a small localization error can eliminate the entire branch [3]. Motion, partial-volume effects and heterogeneous contrast all obscure the actual blood vessel borders [4]. Calcified plaque introduces blooming and may merge with the lumen of adjacent high-attenuation tissue [5]. Therefore, voxel overlap is not a complete index of clinical utility, as a mask with a high Dice score may still contain interruptions or anatomically implausible shortcuts [6].

Convolutional encoder-decoder networks improved volumetric segmentation by learning hierarchical appearance features, but their effective context was often lacking in distinguishing a weak coronary segment from a nearby vein or chamber boundary [7]. Attention-based architectures expand the receptive field to learn relationships among distant parts of the vessel [8]. Volumetric Transformers, including windowed designs, are

convenient compromises between global context and memory consumption [9]. Direct decoding is still based on local classification, and isolated confident errors can be carried over to the final mask [10]. Topological-aware loss has already addressed some of the issues, but the regularization during training does not ensure that all inference results meet connectivity and boundary consistency [11].

This paper takes on the problems of context representation and structured vascular inference. We introduce a topology-preserving nnFormer-CRF model that integrates a three-dimensional transformer backbone with multiscale soft-skeleton supervision and a graph-conditioned conditional random field. nnFormer is used to recover weak, extended evidence through moving window attention; the topological branches clearly show the centerline and branch points; and conditional random fields are used to eliminate false links and repair brief interruptions, considering image appearance, spatial proximity, and local tangent consistency. Evaluation of the Design by patient-level cross-validation with overlap, surface, centerline, connectivity, calibration and efficiency criteria. Therefore, the main contribution is not a separate loss term, but an end-to-end inference scheme that learns volumetric context and anatomical structure in tandem.

Related Work

Volumetric coronary segmentation

Three-dimensional U-shaped networks are still the main baseline for coronary CTA because skip connections maintain fine spatial information [12]. Residual and dense connection variants have shown improvements in optimization and feature reuse, especially when training data is limited [13]. The cascade strategy first locates the heart and then segments the vessels at a higher resolution to reduce class imbalance, but at the cost of adding extra stages [14]. Deep supervision and compound overlap-boundary objectives improve distal sensitivity but are unstable near calcification and low-contrast ostia [15]. The above ways have good local representations but fail to explicitly encode the long-range identity of a vessel tree.

Transformer-based medical segmentation replaces or adds token interaction to convolution. Window attention reduces the quadratic cost of global attention and is suitable for anisotropic volumetric patches. The hybrid encoder can link weak distal evidence with proximal coronary context, but coarse segmentation may lose the smallest branches. Thus, a practical coronary model needs to maintain high-resolution paths and allow information to travel outside the local convolutional field.

Topology-aware learning and structured refinement

Topology-sensitive segmentation has been studied through centerline losses, persistent-homology penalties, and graph-based objectives [16]. Soft-skeleton formulations provide differentiable overlap between predicted and reference centerlines and directly penalize broken thin structures [17]. Persistent-homology approaches compare connected components and loops, but their computational cost and global gradients can complicate large-volume training [18]. Graph neural representations offer explicit branch reasoning, although converting dense probability maps into stable graphs during training remains nontrivial [19]. Conditional random fields model label agreement using unary evidence and pairwise similarity, and efficient mean-field inference permits integration with neural predictions [20]. Existing pipelines typically apply these devices separately; few combine transformer context, multiscale skeleton supervision, and graph-conditioned refinement within one coronary CTA workflow.

nnFormer Representation and Topology Learning

Volumetric representation

Each CTA volume is clipped to the attenuation interval [-200, 900] HU, normalized, and resampled to an isotropic voxel spacing of 0.5 mm. A coarse heart mask defines a $192 \times 192 \times 160$ region of interest. Overlapping $4 \times 4 \times 4$ patches are projected into C-dimensional tokens, and four encoder stages subsequently process these tokens through shifted three-dimensional attention windows. For the token matrix $\mathbf{X}$ at stage l, the query, key, and value representations are obtained as

https://doi.org/10.64972/jiic.2026v4.471p15s:190-203

$$\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{W}_Q, \boldsymbol{K} = \boldsymbol{X}\boldsymbol{W}_K, \boldsymbol{V} = \boldsymbol{X}\boldsymbol{W}_V. \quad \text{Eq.(1)}$$

The attention response in each window combines content similarity with a learned 3D relative position bias. A position term can help the network differentiate between vessel evidence that looks the same but is in different parts of the body.

$$\mathcal{A}(\boldsymbol{X}) = Softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\top}}{\sqrt{d}} + \boldsymbol{B}\right)\boldsymbol{V}. \quad \text{Eq.(2)}$$

A convolutional stem halves the resolution and retains high-frequency lumen boundaries and small-scale intensity fluctuations at the same time. At each decoder stage, a spatially varying fusion gate is used to combine the Transformer representation and convolutional feature maps.

$$\boldsymbol{F}_l = \boldsymbol{g}_l \odot \boldsymbol{T}_l + (1 - \boldsymbol{g}_l) \odot \boldsymbol{C}_l, \boldsymbol{g}_l = \sigma\big(\boldsymbol{W}_g[\boldsymbol{T}_l, \boldsymbol{C}_l]\big). \quad \text{Eq.(3)}$$

The attention schedule is still anisotropy-aware after isotropic resampling. In the early stages, small windows are used to avoid mixing the coronary lumen with the chambers before the reliable vascular features appear. In the later stages, the effective receptive field is expanded by alternately using moving and stationary windows. A shifted partition moves half a window in each space so that tokens can communicate with adjacent windows without constructing a global attention matrix. Patch merging doubles the number of channels and reduces the spatial dimensions. The patch expansion in the decoder is combined with the normalized high-resolution skip features. This way can retain the contextual advantages of the Transformer encoder and keep memory consumption suitable for a clinical workstation GPU.

The fusion gate is spatially distributed, not a single value. Near the sharp cavity boundaries, more emphasis is placed on convolutional representations; in the weak distal regions, transformer responses related to identifiable proximal arterial paths are prioritized. In the first 100 cycles, slightly regularize the gate entropy to prevent the model from collapsing into either branch before both representations become effective. Deep decoder features also feed a low-cost coronary-presence head. An auxiliary head predicts whether each coarse image block contains a reference vessel and reduces the allocation of attention to large background areas. It is not required in inference and thus adds no deployment overhead.

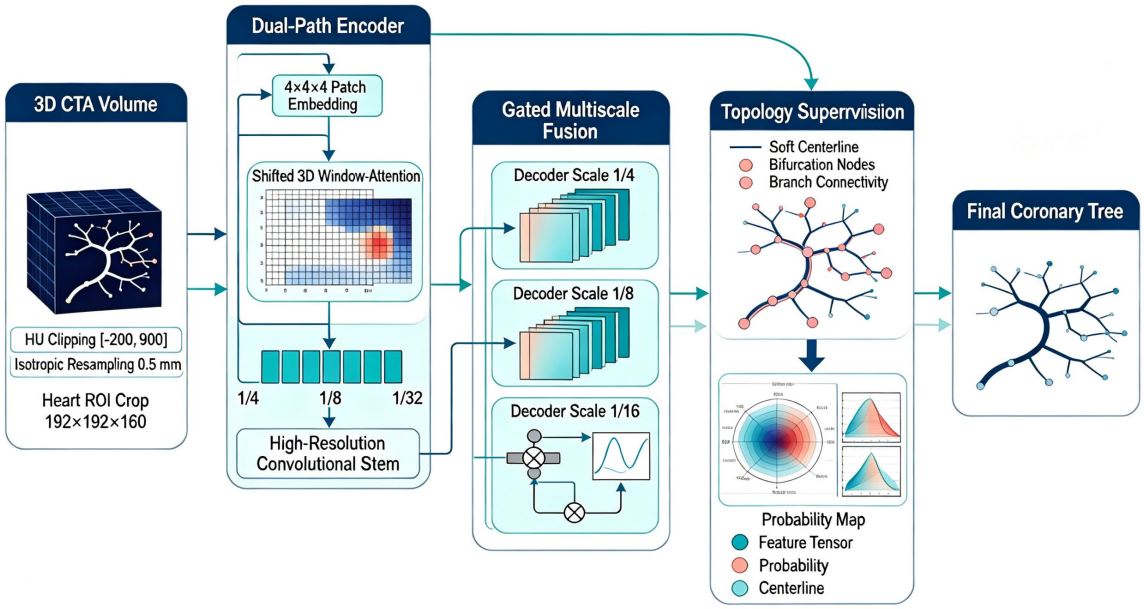


Figure 1. Topology-preserving nnFormer pipeline from CTA preprocessing to multiscale decoding and topology supervision.

Input sampling is centered on the coronary bounding box and keeps 20 mm of the surrounding anatomy. Half of the training crops are selected at centerline points with a probability inversely proportional to the local vessel radius, and the rest are uniformly sampled from the heart region. Radius-balanced sampling expands the exposure of the model to distal vessels without reducing ostial or proximal examples. Intensity augmentation maintains the attenuation order of the blood pool and nearby soft tissues. Simulated motion streaks are placed

outside and across the lumen with different probabilities. The above transformations increase acquisition-related variability without adding anatomically implausible training samples.

Figure 1 is the entire topology-preserving nnFormer workflow. The input CTA volume is attenuated-clipped, isotropically resampled and heart-region cropped before entering the dual-path encoder. Fuse volumetric transformer features and high-resolution convolutional features at multiple decoder levels. Then, the generated probability map undergoes centerline and bifurcation supervision, and is subsequently provided to the graph conditional CRF module.

Multiscale Topology Supervision

A probability head is added to the final decoder and two intermediate decoder scales. Regional overlap and focal classification terms are employed to solve the problem of severe imbalance between coronary and background voxels.

$$\mathcal{L}_{seg} = 1 - \frac{2\sum_i p_i y_i + \varepsilon}{\sum_i p_i + \sum_i y_i + \varepsilon} + \lambda_f \mathcal{L}_{focal}. \quad \text{Eq.(4)}$$

A differentiable thinning operator $\mathcal{S}_k(\cdot)$ iteratively removes peripheral probability mass while preserving ridge-like responses. Centerline agreement is measured symmetrically so that both missing branches and spurious centerline fragments contribute to the loss.

$$\mathcal{L}_{cl} = 1 - \frac{2\langle \mathcal{S}_k(\boldsymbol{p}), \mathcal{S}_k(\boldsymbol{y}) \rangle + \varepsilon}{\|\mathcal{S}_k(\boldsymbol{p})\|_1 + \|\mathcal{S}_k(\boldsymbol{y})\|_1 + \varepsilon}. \quad \text{Eq.(5)}$$

Branch points are identified from the soft skeleton through a 26 -neighborhood degree estimator d_i. Locations corresponding to high-degree reference nodes are given more weight because an error at a bifurcation can affect several downstream branches.

$$\mathcal{L}_{br} = \sum_i \omega_i |d_i(\boldsymbol{p}) - d_i(\boldsymbol{y})|, \omega_i = 1 + \eta \mathbb{I}[d_i(\boldsymbol{y}) \geq 3] \quad \text{Eq.(6)}$$

The full nnFormer objective distributes the regional, centerline and branch constraints at all levels of the supervised decoder. The low-resolution prediction does not need to reproduce the exact boundary detail of the final output.

$$\mathcal{L}_{nnF} = \sum_{s=1}^{S} \alpha_s \left(\mathcal{L}_{seg}^s + \lambda_c \mathcal{L}_{cl}^s + \lambda_b \mathcal{L}_{br}^s \right). \quad \text{Eq.(7)}$$

The goals are not the same, intentionally. Regional Segmentation is a large-scale vascular Area; Skeleton Agreement maintains continuous paths; Branch supervision prevents false bifurcations. Only optimizing the overlapping areas resulted in relatively thicker distal vessels and occasional cross-vessel attachments. In contrast, applying strong topological constraints before the model learns the appearance of the lumen can lead to early optimization instability.

Thinning depth k is scaled according to the decoder resolution. Two iterations are used for the coarsest supervised map, four for the intermediate scale, and eight for the full resolution. A fixed reduction in thickness would either over-reduce small, thick branches or fail to skeletonise large proximal vessels. Skeleton terms are evaluated in a dilated reference envelope to prevent the influence of remote false-positive regions on the connectivity gradient, and these areas are still regulated by the regional focal term. The range of branches is set by soft neighborhood aggregation and limited to six to reduce the impact of noise local clusters.

Topology weights increase linearly for the first 80 epochs and then stay the same. This curriculum can help the decoder learn vascular appearance without adding strong structural constraints. Optimize the use of two batches of crops and perform four-step gradient accumulation. The initial learning rate is 2×10^{-4}, weight decay is 10^{-5}, and stochastic depth increases from zero to 0.15 across the encoder stages.

The model selection criteria combine Dice, centerline Dice, and normalized connection penalty. Dice alone selected checkpoints that expanded the distal masks and occasionally connected adjacent arteries. The composite criterion has a narrow probability ridge in difficult regions and penalizes excessive expansion. At inference time, the predictions of overlapping crops are combined based on a Gaussian importance map to reduce seam artifacts prior to the construction of the anatomical graph.

Graph-Conditioned CRF Refinement

Graph-Conditioned CRF Energy

Skeletonize the initial probability mask, and then convert it into an anatomical diagram. Its nodes are the ends of the coronary arteries, bifurcation points, and regularly sampled centerline positions. Each node has a position, neural probability, estimated vessel radius, local attenuation and centerline tangent. A small adaptive band around the graph limits the structured refinement to the anatomically reasonable coronary area.

Graph extraction starts with a relatively high probability threshold of 0.35 and will be performed before final segmentation. This arrangement has weak but spatially correlated distal evidence. Components shorter than 2 mm are excluded unless they are within 3 mm of a high-confidence endpoint. Graph edges are traced along skeleton paths and then resampled at 1 mm intervals. The tangent vectors are estimated using principal component analysis within a 5mm geodesic neighborhood to ensure stability when small staircase artifacts appear in the voxel skeleton. Vessel radius is obtained from the Euclidean distance transform and determines the adaptive-band width, which is restricted to the interval of $2 - 6$ mm.

Figure 2 is the structured refinement mechanism. Neural unary logits are first converted into an anatomical graph and adaptive coronary band. Appearance and tangent-consistency kernels then exchange information via recurrent mean-field iterations. The final Topology audit accepts well-supported reconnections and rejects short cycles and anatomically implausible shortcuts.

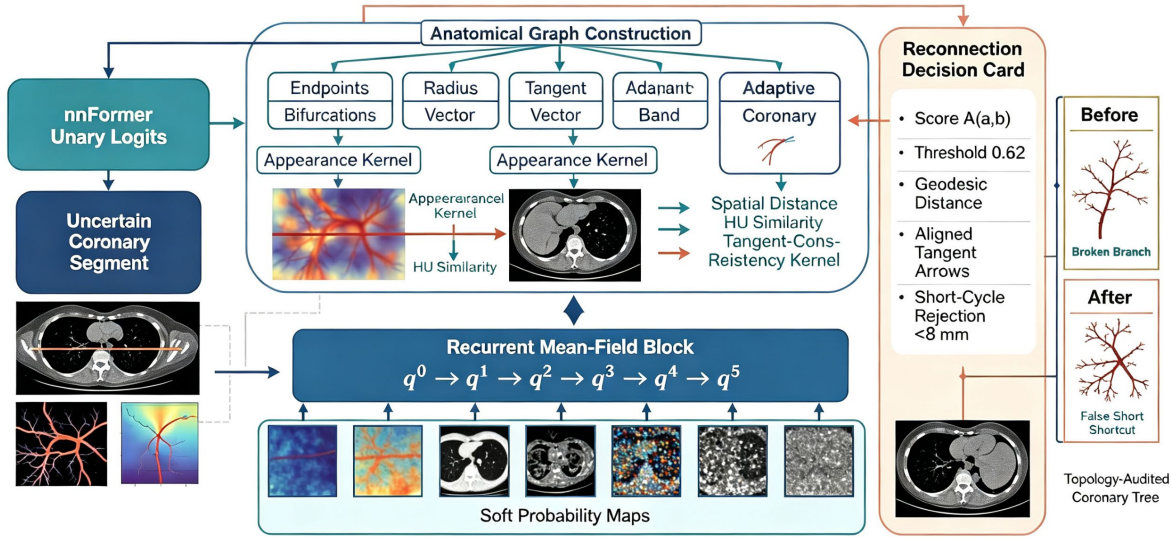


Figure 2. Graph-conditioned CRF refinement with adaptive-band message passing and topology auditing.

For the voxel-label field $\mathbf{z}$, the conditional random field minimizes an energy consisting of neural unary costs and graph-aware pairwise interactions.

$$E(\boldsymbol{z}) = \sum_{i} \psi_u(z_i) + \sum_{(i,j)\in\mathcal{N}_G} \psi_p(z_i, z_j). \qquad \text{Eq.(8)}$$

The unary cost is obtained from temperature-scaled nnFormer probabilities. The temperature parameter was optimized on the validation set to reduce systematic overconfidence while not altering the spatial order of neural predictions.

$$\psi_u(z_i) = -log\, P_\theta(z_i \mid \boldsymbol{x}_i)^{1/\tau}. \qquad \text{Eq.(9)}$$

Pairwise compatibility combines an appearance kernel and a tangent-consistency kernel. A small weight is given to the Directional component unless both voxels are close to the estimated coronary graph.

$$\psi_p(z_i, z_j) = \mu(z_i, z_j)\left[w_1 k_{app}(i,j) + w_2 k_{tan}(i,j)\right]. \qquad \text{Eq.(10)}$$

The appearance term keeps the boundary information by requiring simultaneous spatial and attenuation similarity. Normalization is performed independently in each adaptive graph band, and a large cardiac blood pool does not affect the information transmission around the narrow coronary segment.

https://doi.org/10.64972/jiic.2026v4.471p15s:190-203

$$k_{app}(i,j) = exp\left(-\frac{\left\|\boldsymbol{r}_i - \boldsymbol{r}_j\right\|_2^2}{2\sigma_r^2} - \frac{\left|I_i - I_j\right|^2}{2\sigma_I^2}\right). \quad \text{Eq.(11)}$$

The tangent term propagates the compatible labels preferentially along the estimated arterial direction. The squared inner product keeps the tangential direction of the kernel unchanged, and vectors pointing in opposite directions along the same central line remain compatible.

$$k_{tan}(i,j) = exp\left(-\frac{\left\|\boldsymbol{r}_i - \boldsymbol{r}_j\right\|_2^2}{2\sigma_g^2}\right)\left(\boldsymbol{t}_i^{\top}\boldsymbol{t}_j\right)^2. \quad \text{Eq.(12)}$$

Not all parts of the CTA crop use the same appearance kernel. The intensity bandwidth is determined by the median absolute deviation of the voxels around the high-confidence centerline. This adaptive estimate can give different weights to strongly and weakly enhanced examinations. Spatial Bandwidth follows the local radius; proximal arteries have a wider propagation and distal vessels use narrow interactions to reduce shortcuts.

The tangent kernel is disabled within 2 mm of a detected bifurcation because a single direction is not appropriate at a branching point. Compatibility in that Area is based on graph adjacency and appearance similarity. Therefore, conditional treatment will not reduce the direction of valid daughter branches or force them to align with the parent vessel.

Differentiable Mean-Field Inference

Mean-field inference is unrolled for five iterations. Message passing is implemented using separable Gaussian filtering within the adaptive coronary band, followed by a compatibility transformation and probability normalization.

$$m_i^{(t)}(l) = \sum_{j \in \mathcal{N}_G(i)} k(i,j) q_j^{(t-1)}(l). \quad \text{Eq.(13)}$$

The posterior update retains the nnFormer unary evidence while incorporating the compatible labels of neighboring voxels. The normalization factor Z_i ensures that the updated label probabilities sum to one.

$$q_i^{(t)}(l) = \frac{1}{Z_i} exp\left[-\psi_u(l) - \sum_{l'} \mu(l,l') m_i^{(t)}(l')\right]. \quad \text{Eq.(14)}$$

Short gaps are considered for reconnection only when their opposing endpoints have compatible tangent directions and the path between them has a sufficiently low geodesic cost. Let a and b denote two candidate endpoints. Their reconnection score is defined by the mean path probability $\bar{p}$, geodesic distance $d_g(a,b)$, and endpoint tangent agreement.

$$R(a,b) = \bar{p} exp\left(-\frac{d_g(a,b)}{\rho}\right) max(0, \boldsymbol{t}_a^{\top}\boldsymbol{t}_b). \quad \text{Eq.(15)}$$

A candidate connection is accepted when $R(a,b) > 0.62$ and the added path does not produce a cycle shorter than 8 mm. Each endpoint can participate in at most one accepted reconnection, preventing a single uncertain location from generating several artificial daughter branches.

The final learning criterion supplements the multiscale nnFormer loss with the negative log-likelihood of the CRF posterior.

$$\mathcal{L}_{total} = \mathcal{L}_{nnF} + \lambda_r \left[-\sum_i log\, q_i(y_i)\right]. \quad \text{Eq.(16)}$$

The unrolled iterations share the same kernel parameters, have negligible parameter growth, and thus converge to a stable fixed point. The gradient is processed through filtering, compatibility, and normalization layers during fine-tuning. Graph construction is treated as a stopped-gradient operation because differentiation via discrete skeletonization resulted in unstable training without improving validation topology.

A residual coefficient of 0.6 is used to combine each updated posterior with the previous estimate, and oscillations near calcified boundaries are thus damped. Voxels outside the adaptive band keep their neural

probabilities and are subjected to the standard threshold; only voxels inside the band are updated via topology-aware posterior refinement.

Topological auditing is conducted after binarization. Each connected component is mapped back to the anatomical graph, and at the endpoints, whether they are close to the expected ostial and distal regions determines whether they are classified as anatomical or artificial. A short bridge for a suspect is tested by briefly removing its middle part. The bridge is not selected because, after removal, the two high-confidence paths point in different directions.

Topology analysis is performed, and then a sub-voxel surface is generated via Marching Cubes. The two curvature-constrained smoothing iterations eliminate isolated stair-step artifacts without moving the estimated centerline position. Therefore, it is necessary to limit the scope of smoothing to avoid reducing the size of small branches or shifting branch points.

Training uses AdamW for 600 epochs with cosine learning-rate decay, mixed-precision arithmetic, random spatial rotation, gamma perturbation, and simulated motion streaks. The transformer and topology-supervision branches are optimized for the first 500 epochs. Graph-conditioned inference is then jointly fine-tuned for another 100 epochs using a reduced learning rate.

Experimental Evaluation and Analysis

Data, protocol, and evaluation

A total of 420 coronary CTA volumes retrospectively assembled from three centers were used for evaluation: 260 for training, 80 for validation, and 80 for holdout testing in each fold, using patient-level five-fold cross-validation. Scanner vendors, tube voltages, and reconstruction kernels were deliberately mixed. Two cardiovascular imaging readers generated the lumen masks and centerlines, with discrepancies resolved by a senior reader. The three cohorts of mild, moderate and severe calcification included 146, 168 and 106 cases respectively. Institutional data handling followed the de-identification and access-control procedures generally required for multicenter imaging research [21].

The baselines were 3D U-Net, nnU-Net, V-Net, UNETR, Swin UNETR and the unmodified nnFormer. All models used identical crops and augmentation budgets. Dice, Jaccard, centerline Dice (clDice), centerline recall, average symmetric surface distance (ASSD), 95th percentile Hausdorff distance, disconnected components, false links, expected calibration error, parameters, memory, and latency are recorded. Using 2,000 patient-level bootstrap replicates, confidence intervals were estimated to show accepted paired evaluation practice [22]. Two-sided Wilcoxon tests with Holm correction were employed to compare models [23].

Figure 3(a) reports Dice distributions of 0.856, 0.871, 0.884, 0.893, 0.901, 0.904, and 0.912 for the seven methods, while Figure 3(b) pairs clDice with ASSD: the proposed model reaches 0.936 clDice and 0.31 mm ASSD, compared with 0.908 and 0.39 mm for nnFormer. The distribution width also narrows, indicating fewer catastrophic cases rather than only a higher mean.

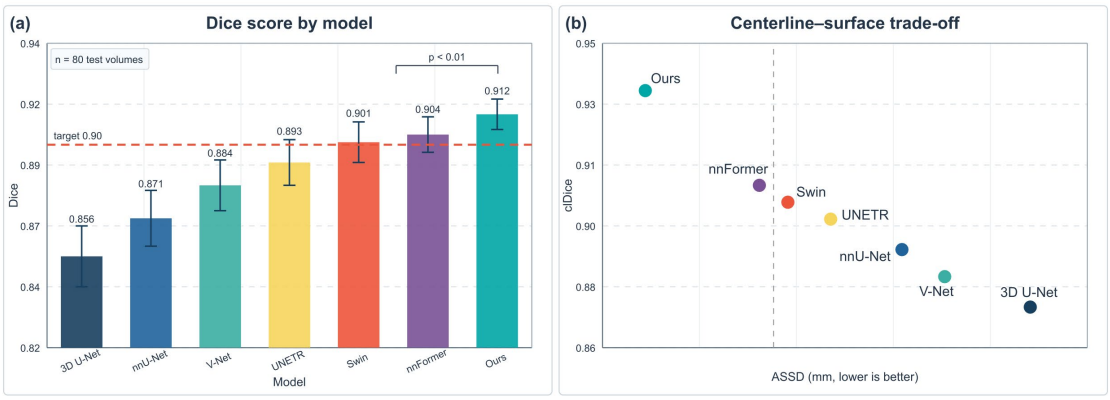


Figure 3. Overall segmentation performance: (a) Dice distributions across seven methods; (b) clDice–ASSD comparison with confidence intervals.

https://doi.org/10.64972/jiic.2026v4.471p15s:190-203

Patient-level splitting avoided crossing folds in repeated examinations. Preprocessing parameters were learned or selected only on the training and validation sets, including the heart crop margin, probability temperature and reconnection threshold. Quantification of reader agreement before adjudication: the two initial masks achieved 0.928 Dice and 0.946 clDice, and most disagreements were in areas less than 1.2 mm. The reference centerline is generated from the defined mask and is corrected at the opening, bifurcation, and distal ends. Exclusion set does not include background regions outside the heart area; false positives remain in the heart area.

Pairwise statistics were conducted by case. In addition to the group mean, the analysis also retained the distribution of each case, the acquisition center, the calcification layer, the contrast-to-noise ratio, and the estimated motion level. A result was deemed practically significant only if the bootstrap interval of the paired difference excluded zero and the topology change exceeded 0.1 errors per case. Both rules avoid over-interpreting small numerical increases due to a large sample. Efficiency measurements used a fixed batch of ten warm-up volumes and included preprocessing, neural inference, CRF refinement and export; data loading from disk was excluded.

Reproducibility tests were conducted before the final fold was unfolded. Random seeds control crop selection, augmentation, parameter initialization, and bootstrap resampling. The same post-processing allowance was given to all baselines; only their validation sets could determine a component-size threshold, and none of the baselines received the proposed graph or CRF. Map back to the original resolution and calculate the metric at the actual coordinates. Sample the centerline distance at 0.5 mm intervals. If the predicted point falls within the local reference radius plus one voxel, it is considered recovered. This rule considering the radius will not penalize the correct centerline due to boundary uncertainty in the broad proximal part.

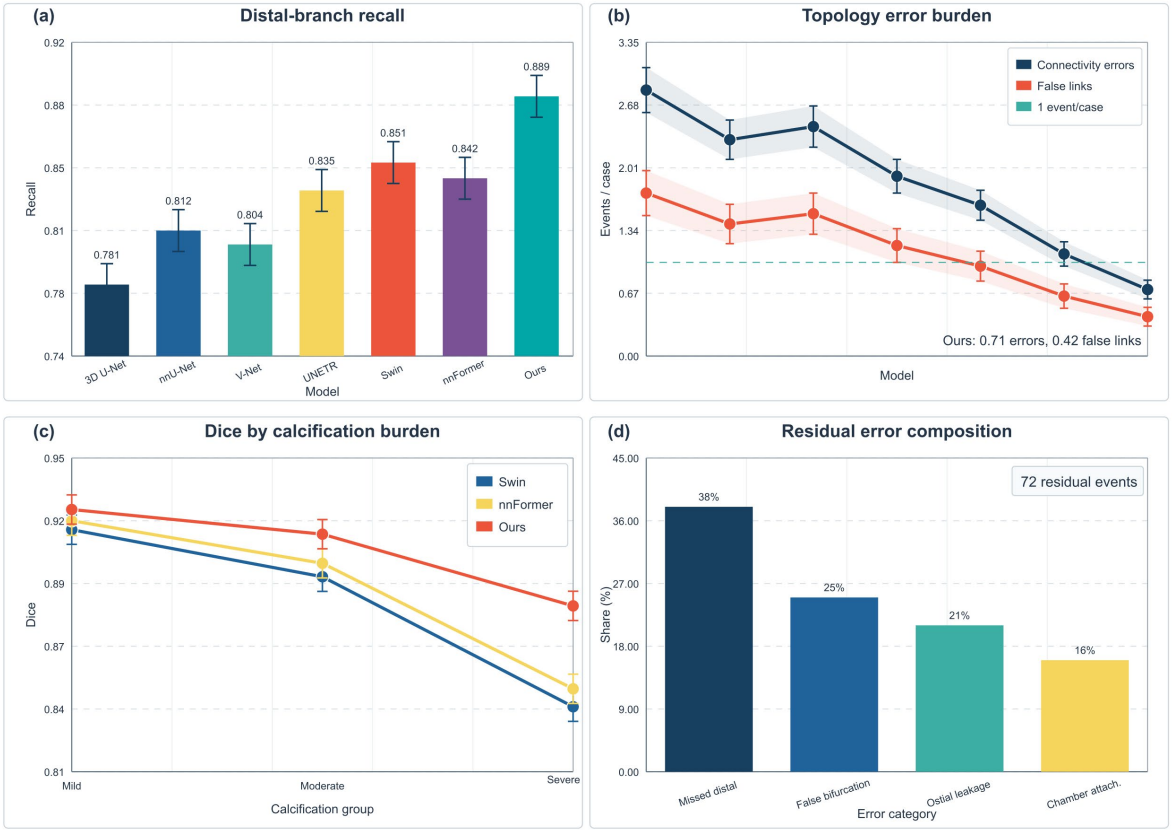


Figure 4. Topology and stratified accuracy: (a) distal-branch recall; (b) connectivity errors; (c) Dice by calcification burden; (d) residual error composition.

Comparative and stratified results

Figure 4 contains four complementary views. Figure 4(a) shows distal-branch recall rising from 0.781 for 3D U-Net to 0.889 for nnFormer-CRF; Figure 4(b) shows connectivity errors falling from 2.84 to 0.71 per case. Figure

4(c) plots Dice against calcification burden: in the severe group the proposed model attains 0.884, whereas nnFormer reaches 0.847 and Swin UNETR 0.839. Figure 4(d) presents an error decomposition in which missed distal segments, false bifurcations, ostial leakage, and chamber attachment account for 38%, 25%, 21%, and 16% of residual errors. The severe-calcification advantage is consistent with the need for explicit structural context when local intensities become ambiguous [24].

Figure 5(a) maps performance across vessel diameter and contrast-to-noise ratio; Dice remains above 0.88 when diameter exceeds 1.5 mm, but declines in the joint low-diameter, low-contrast corner. Figure 5(b) gives reliability curves: temperature-scaled CRF output reduces expected calibration error from 0.041 to 0.021 and keeps the curve close to the identity line. Calibration matters because confidence can guide manual review and downstream measurement [25]. The heat map reveals that representation, not refinement, remains the limiting factor for submillimeter branches.

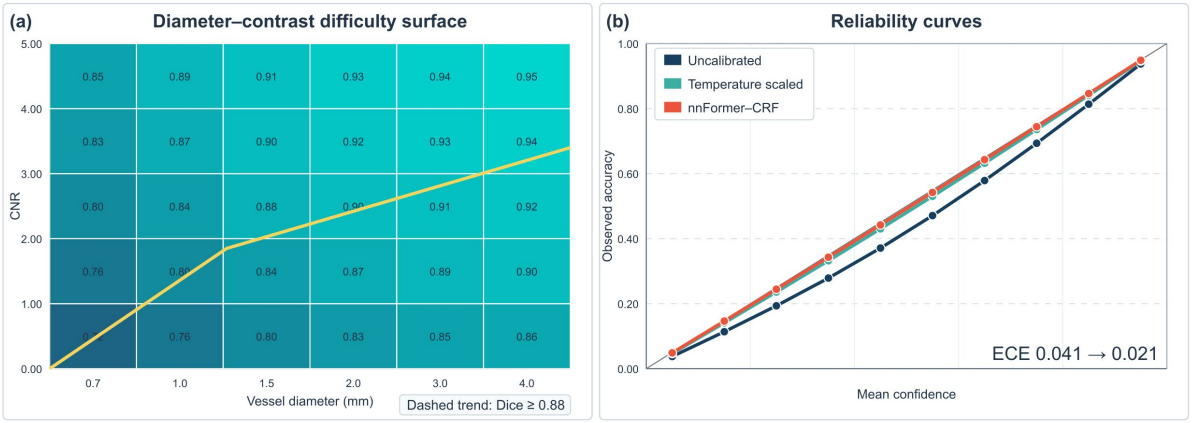


Figure 5. Difficulty and calibration analysis: (a) Dice over vessel diameter and contrast-to-noise ratio; (b) reliability curves and expected calibration error.

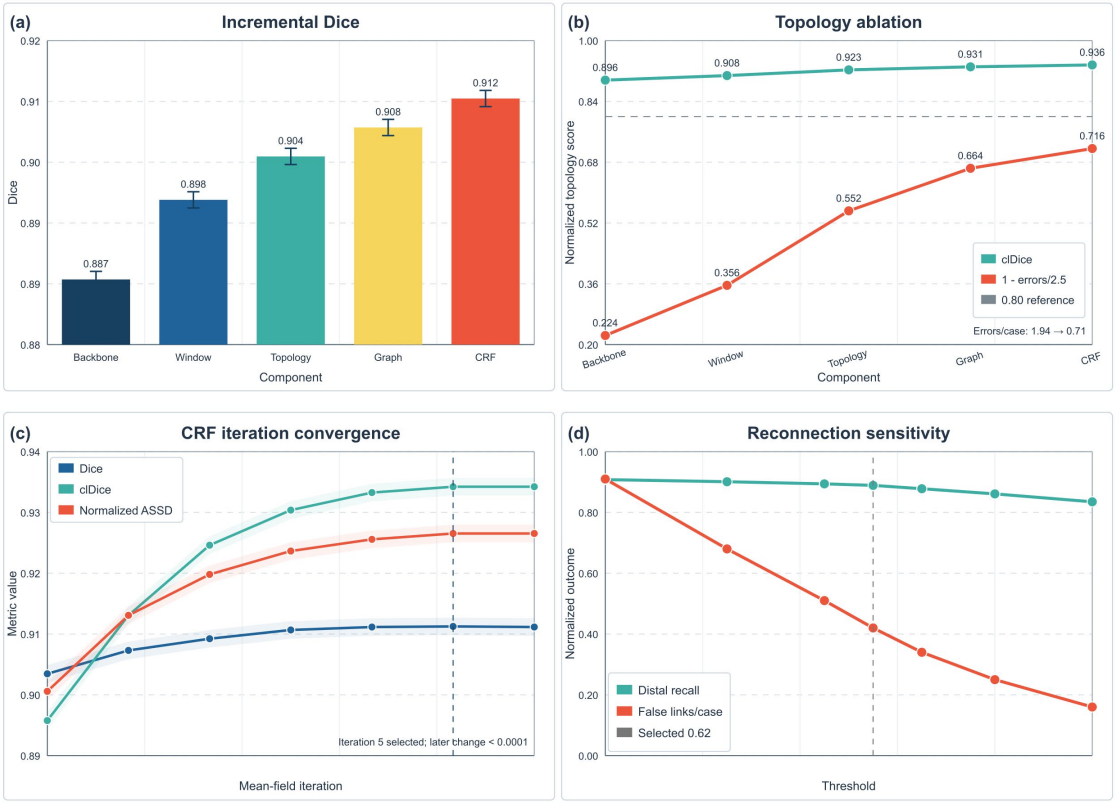


Figure 6. Ablation and sensitivity analysis: (a) incremental Dice; (b) topology metrics; (c) CRF iteration convergence; (d) reconnection-threshold trade-off.

https://doi.org/10.64972/jiic.2026v4.471p15s:190-203

Figure 6 isolates the contributions of each component. Figure 6(a) shows that window attention, topology supervision, graph conditioning, and CRF refinement successively increase Dice from 0.887 to 0.912. Figure 6(b) demonstrates a larger change in topology: clDice increases from 0.896 to 0.936 and connectivity errors decrease from 1.94 to 0.71. Figure 6(c) traces the CRF iteration curve; most improvement occurs within three iterations and saturates at five. Figure 6(d) plots reconnection threshold sensitivity, with 0.62 providing the best balance between 0.889 distal recall and 0.42 false links. These ablations support the use of complementary regional and structural criteria rather than a single overlap target [26].

Paired testing showed that the 0.008 Dice advantage over nnFormer was consistent across folds, with a 95% confidence interval of 0.005 to 0.011 and a corrected p-value less than 0.01. The corresponding clDice increase was larger at 0.028, and the median number of disconnected components decreased from 1.0 to 0.0. Improvement over Swin UNETR: 0.011 Dice and 0.033 clDice. Regional indicators therefore underreported the benefits shown by the vessel tree. The three Transformer models had relatively small differences in proximal recall, but the difference was more pronounced for distal recall.

False-positive volume did not increase after reconnection. The proposed model achieved a production rate of 0.46 ml per case and was lower than nnFormer (0.51 ml) and Swin UNETR (0.63 ml). A reduction in false links was observed near the right atrium and cardiac veins, and context alone was not sufficient to explain them. Ostial leakage was still relatively rare at 0.14 events per case. Surface errors were concentrated at the heavily calcified proximal segment, but the median 95th-percentile Hausdorff distance of the nnFormer was 3.48 mm and that of the proposed model was 2.91 mm.

The diameter-by-contrast surface offers a more precise reading than a single severity label. At diameter below 1.0 mm and contrast-to-noise ratio of one, Dice was 0.72, but it rose to 0.85 when contrast-to-noise ratio reached five. For vessels of 2.0 mm or larger, all cells exceeded 0.87, and topology refinement provided most of its gain by closing short gaps rather than shifting boundaries. Reliability analysis showed no systematic overconfidence above 0.8 predicted probability. The greatest residual miscalibration occurred between 0.4 and 0.6, a range dominated by partial-volume distal voxels.

Ablation also separated the functions of graph construction and mean-field refinement. Adding topology supervision without the graph increased clDice by 0.015 but left 1.12 connectivity errors per case. Graph conditioning reduced the count to 0.84, and then CRF inference reduced it further to 0.71. Tangent consistency was replaced with intensity-only smoothing, and it achieved 0.906 Dice and 0.918 clDice; thus, direction did not significantly enhance continuity. Expanding the CRF band beyond 6 mm increased false links near the chambers, and a 2 mm fixed band failed to cover the wide proximal boundary correction.

The folded state behavior is now stable. The range of the proposed-model Dice was 0.908-0.916, and the range of clDice was 0.931-0.940. The largest paired gain was in the fold with the highest proportion of severe calcification, but even after removing this fold, it was still 0.006 better than nnFormer. Endpoint analysis showed that 93.1% of the reference distal endpoints were within 2 mm of the predicted path, and only 88.4% for nnFormer. The number of extra endpoints dropped by 22.6%, and thus, increased recall was not due to uncontrolled fragmentation or branching.

Robustness, efficiency, and failure analysis

Figure 7(a) compares robustness under increasing simulated motion: at the strongest level, Dice is 0.861 for the proposed model, 0.826 for nnFormer, and 0.801 for nnU-Net. Figure 7(b) shows a Pareto plot of accuracy and latency; nnFormer-CRF requires 1.8 s per case with 56.2 million parameters, remaining faster than the 2.6 s Swin UNETR configuration. Figure 7(c) displays per-center performance, with Dice values of 0.916, 0.909, and 0.907 and overlapping 95% confidence intervals. Multicenter variation of reconstruction and enhancement is a known source of domain shift [27], and harmonized preprocessing can reduce but not remove it [28].

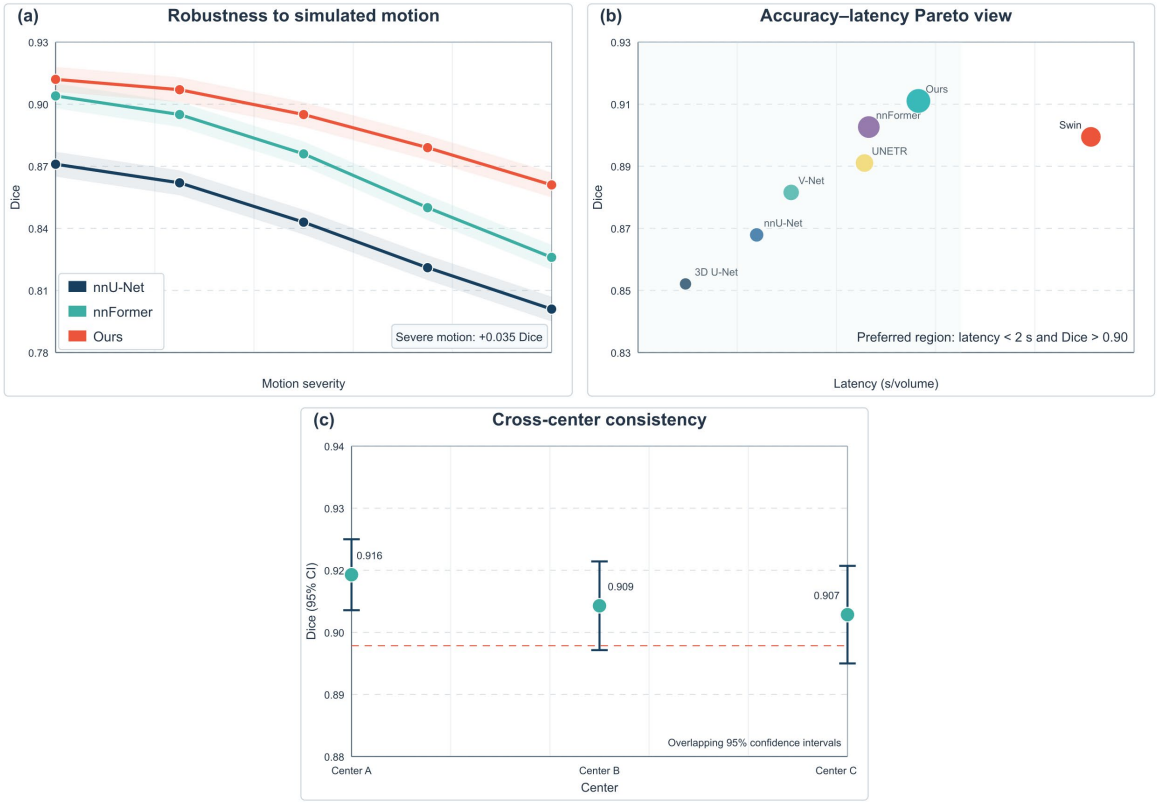


Figure 7. Robustness and deployment: (a) motion robustness; (b) accuracy–latency Pareto front; (c) center-wise Dice with 95% confidence intervals.

The method still failed in nine test cases. Four involved severe stair-step motion, three contained metal-induced streaks, and two had anomalous coronary origins outside the learned crop prior. Review showed that CRF refinement cannot reconstruct a long branch when the transformer supplies no continuous probability ridge. Conversely, a strong but incorrect venous ridge may be preserved when its tangent aligns with the coronary graph. External validation and anomaly-enriched sampling are therefore important before clinical deployment [29]. Runtime profiling indicated 1.34 s for nnFormer prediction, 0.23 s for graph construction and mean-field refinement, and 0.23 s for resampling and export. Memory peaked at 9.7 GB for the 192 x 192 x 160 crop. These costs fit offline CTA analysis, while further pruning or tiling would be required for commodity hardware [30].

Motion experiments used synthetic displacement fields calibrated with observed stair-step artifacts. Degradation was roughly linear for all models under moderate motion, and then the convolutional baseline dropped more sharply. The proposed model improved the Dice coefficient by 0.035 over nnFormer at the severe setting, but an absolute value of 0.861 shows that topology cannot fully compensate for missing image evidence. The central analysis shows that there is no significant interaction between the model and the center. The widest confidence interval was at Centre C because it had fewer severe cases and not because its mean was significantly lower.

nnFormer backbone accounted for most of the parameter and memory cost. CRF added only 0.18 million trainable parameters and increased the peak memory by 0.6 GB during joint fine-tuning. Sparse graph storage at inference was still less than 40MB per case. Five iterations were selected because a sixth changed the mean Dice by less than 0.0001. A CPU-only setup had graph extraction and filtering in 2.9 seconds, and thus the structured stage could still be applied in practice after remotely generating or caching the neural probabilities.

Qualitative analysis did not alter the numerical labels and explored failure modes. In successful cases of severe calcification, the neurography includes two low-probability ridges near the expanded plaques, while tangent-aware refinement retains ridges aligned with the upstream and downstream centerlines. In the motion failure, neither ridge was present for a length exceeding 12 mm, and thus no defensible candidate path was found. In anomalous-origin cases, the crop truncated the initial ostial segment and caused an artificial endpoint. Therefore,

the goal of optimization is to enhance localization and evidence retrieval, rather than increasing pairwise smoothing intensity.

Threshold transfer was examined by selecting all CRF settings at Center A and applying them unchanged to Centers B and C. Dice changed by less than 0.002 relative to the center-specific validation, and connectivity errors changed by no more than 0.06 per case. Therefore, the reduced intensity-bandwidth product is associated with a lower acquisition variation. It does not have universal applicability because these three centers used different contrast agent protocols and adult populations. Pediatric examinations, very low-dose scans, bypass grafts and stented segments were not sufficiently represented to support separate claims.

Conclusion

This paper introduces a topology-preserving nnFormer-CRF framework for coronary artery segmentation in CTA volumes. Shifted volumetric attention provided long-range contextual information, multiscale skeleton and branch supervision made vascular continuity visible during learning, and graph-conditioned conditional random field inference reconciled local appearance with anatomical direction. The resulting Design considers overlap, boundary placement and tree integrity as related but different objectives, providing a unified path for volume preprocessing to topology-audited coronary masks.

Some Deficiencies Remain. The anatomical graph is initialized by the neural probability field, and therefore cannot recover a long vessel segment when the initial evidence is entirely missing. Its crop prior may also be unreliable for unusual coronary origins or major anatomical variations. The study design is retrospective, uses fixed acquisition families, and manual reference masks have uncertainty in very small distal branches. Additionally, the iterative refinement stage has increased the complexity and slightly increased the latency of implementation.

In the future, it will be necessary to link a graph with uncertainty-aware active tracing and suggest missing sections outside the original mask. Extend prospective validation to extended scanner protocols, congenital variants and post-intervention cases. Self-supervised Volumetric Pre-training, Domain Adaptation and Lightweight Sparse Attention Boost Transfer and Efficiency. Another approach is to jointly estimate the attributes of lumen, plaque, centerline and stenosis under shared topological constraints, thus enabling clinically meaningful quantitative analysis of segmentation results rather than producing independent end products of image processing.

Author Contributions

Ivan Jelen contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Dalibor Sedláček contributes to validation, analysis, investigation, data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Pan, L.-S., Li, C.-W., Su, S.-F., Tay, S.-Y., Tran, Q.-V., & Chan, W. P. (2021). Coronary artery segmentation under class imbalance using a U-Net based architecture on computed tomography angiography images. Scientific Reports, 11(1), 14493. https://doi.org/10.1038/s41598-021-93889-z

[2] Baskaran, L., Al'Aref, S. J., Maliakal, G., Lee, B. C., Xu, Z., Choi, J. W., et al. (2020). Automatic segmentation of multiple cardiovascular structures from cardiac computed tomography angiography images using deep learning. PLOS ONE, 15(5), e0232573. https://doi.org/10.1371/journal.pone.0232573

[3] Shin, C. I., Park, S. J., Kim, J. H., Yoon, Y. E., Park, E. A., Koo, B. K., & Lee, W. (2020). Coronary artery lumen segmentation using location–adaptive threshold in coronary computed tomographic angiography: A proof-of-concept. Korean journal of radiology, 22(5), 688. https://doi.org/10.3348/kjr.2020.0296

[4] Li, Y., Wu, Y., He, J., Jiang, W., Wang, J., Peng, Y., et al. (2022). Automatic coronary artery segmentation and diagnosis of stenosis by deep learning based on computed tomographic coronary angiography. European Radiology, 32(9), 6037–6045. https://doi.org/10.1007/s00330-022-08761-z

[5] Jávorszky, N., Homonnay, B., Gerstenblith, G., Bluemke, D., Kiss, P., Török, M., et al. (2022). Deep learning-based atherosclerotic coronary plaque segmentation on coronary CT angiography. European Radiology, 32(10), 7217–7226. https://doi.org/10.1007/s00330-022-08801-8

[6] Pathak, A., Anwer, T., Verma, A., Alhujaily, M., Alahmari, M., Alshahrani, S., et al. (2026). Integrating precision medicine and artificial intelligence to prevent cardiotoxicity in cardiovascular drug therapy. Precision Medication, 3(1), Article 100078. https://doi.org/10.1016/j.prmedi.2026.100078

[7] Lin, A., Kolossváry, M., Motwani, M., Išgum, I., Maurovich-Horvat, P., Slomka, P. J., & Dey, D. (2022). Artificial intelligence in cardiovascular imaging for risk stratification in coronary artery disease. Radiology: Cardiothoracic Imaging, 4(1), e210108. https://doi.org/10.1148/ryct.210108

[8] Karimi, D., Dou, H., & Gholipour, A. (2022). Medical image segmentation using transformer networks. IEEE Access, 10, 29322–29332. https://doi.org/10.1109/ACCESS.2022.3156894

[9] Shit, S., Paetzold, J. C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J. P. W., Bauer, U., & Menze, B. H. (2021). clDice—A novel topology-preserving loss function for tubular structure segmentation. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 16555–16564. https://doi.org/10.1109/CVPR46437.2021.01629

[10] Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H. R., & Xu, D. (2022). UNETR: Transformers for 3D medical image segmentation. 2022 IEEE/CVF Winter Conference on Applications of Computer Vision, 574–584. https://doi.org/10.1109/WACV51458.2022.00069

[11] Gao, Y., Zhou, M., & Metaxas, D. N. (2021). UTNet: A hybrid transformer architecture for medical image segmentation. Medical Image Computing and Computer Assisted Intervention—MICCAI 2021, 12903, 61–71. https://doi.org/10.1007/978-3-030-87199-4_6

[12] Shi, Y., Zhang, J., Ling, T., Lu, J., Zheng, Y., Yu, Q., Qi, L., & Gao, Y. (2021). Inconsistency-aware uncertainty estimation for semi-supervised medical image segmentation. IEEE Transactions on Medical Imaging, 41(3), 608–620. https://doi.org/10.1109/TMI.2021.3117888

[13] Luo, X., Wang, G., Liao, W., Chen, J., Chen, Y., Song, T., et al. (2022). Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. Medical Image Analysis, 80, 102517. https://doi.org/10.1016/j.media.2022.102517

[14] Xiao, Z., Su, Y., Deng, Z., & Zhang, W. (2022). Efficient combination of CNN and transformer for dual-teacher uncertainty-guided semi-supervised medical image segmentation. Computer Methods and Programs in Biomedicine, 226, 107099. https://doi.org/10.1016/j.cmpb.2022.107099

[15] Malhotra, P., Gupta, S., Koundal, D., Zaguia, A., & Enbeyle, W. (2022). Deep neural networks for medical image segmentation. Journal of Healthcare Engineering, 2022, 9580991. https://doi.org/10.1155/2022/9580991

[16] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 18(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z

[17] Bertels, J., Eelbode, T., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., & Blaschko, M. B. (2021). Theoretical analysis and experimental validation of volume bias of soft Dice optimized segmentation maps in the context of inherent uncertainty. Medical Image Analysis, 67, 101833. https://doi.org/10.1016/j.media.2020.101833

[18] Mehrtash, A., Wells, W. M., Tempany, C. M., Abolmaesumi, P., & Kapur, T. (2020). Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. IEEE Transactions on Medical Imaging, 39(12), 3868–3878. https://doi.org/10.1109/TMI.2020.3006437

[19] Jungo, A., Meier, R., Ermis, E., Blatti-Moreno, M., Herrmann, E., Wiest, R., & Reyes, M. (2020). On the effect of inter-observer variability for a reliable estimation of uncertainty of medical image segmentation. Medical Image Analysis, 64, 101718. https://doi.org/10.1016/j.media.2020.101718

[20] Zhang, D., Tai, X.-C., & Lui, L. M. (2021). Topology- and convexity-preserving image segmentation based on image registration. Applied Mathematical Modelling, 100, 218–239. https://doi.org/10.1016/j.apm.2021.08.017

[21] Zeleznik, R., Foldyna, B., Eslami, P., Weiss, J., Alexander, I., Taron, J., et al. (2021). Deep convolutional neural networks to predict cardiovascular risk from computed tomography. Nature Communications, 12(1), 715. https://doi.org/10.1038/s41467-021-20966-2

[22] Lin, A., Manral, N., McElhinney, P., Killekar, A., Matsumoto, H., Kwiecinski, J., et al. (2022). Deep learning-enabled coronary CT angiography for plaque and stenosis quantification and cardiac risk prediction: An international multicentre study. The Lancet Digital Health, 4(4), e256–e265. https://doi.org/10.1016/S2589-7500(22)00022-X

[23] Karimi, D., & Salcudean, S. E. (2020). Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks. IEEE Transactions on Medical Imaging, 39(2), 499–513. https://doi.org/10.1109/TMI.2019.2930068

[24] Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., & Martel, A. L. (2021). Loss odyssey in medical image segmentation. Medical Image Analysis, 71, 102035. https://doi.org/10.1016/j.media.2021.102035

[25] Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., ... & Wu, J. (2020, May). Unet 3+: A full-scale connected unet for medical image segmentation. In ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 1055-1059). Ieee. https://doi.org/10.1109/ICASSP40776.2020.9053405

[26] Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., & Li, J. (2021). TransBTS: Multimodal brain tumor segmentation using transformer. Medical Image Computing and Computer Assisted Intervention—MICCAI 2021, 12901, 109–119. https://doi.org/10.1007/978-3-030-87193-2_11

[27] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. 2021 IEEE/CVF International Conference on Computer Vision, 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986

[28] Karimi, D., Dou, H., Warfield, S. K., & Gholipour, A. (2020). Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. Medical Image Analysis, 65, 101759. https://doi.org/10.1016/j.media.2020.101759

[29] Cai, L., Gao, J., & Zhao, D. (2020). A review of the application of deep learning in medical image classification and segmentation. Annals of Translational Medicine, 8(11), 713. https://doi.org/10.21037/atm.2020.02.44

[30] Ahmad, H. M., Khan, M. J., Yousaf, A., Ghuffar, S., & Khurshid, K. (2020). Deep learning: A breakthrough in medical imaging. Current Medical Imaging, 16(8), 946–956. https://doi.org/10.2174/1573405615666191219100824