3D Segmentation of Coronary Arteries in Computer Tomography Angiography

Dmitrii Evtyukhov[1][0009-0008-6346-0000], Georgy Kopanitsa[1][0000-0002-6231-8036], Oleg Metsker[2][0000-0003-3427-7932], and Sergey Kovalchuk[1][0000-0001-8828-4615]

[1] ITMO University, Saint Petersburg, Russia
dmitrii.evtyukhov@itmo.ru, kovalchuk@itmo.ru
[2] Almazov National Medical Research Centre, Saint Petersburg, Russia

Abstract. Accurate 3D coronary artery segmentation in CT angiography (CTA) is a key prerequisite for quantitative vessel analysis and downstream decision support. In this paper, we study three modern 3D segmentation backbones—a 3D U-Net baseline, a residual CNN (SegResNet), and a transformer-based model (SwinUNETR)—on the ImageCAS coronary artery dataset. We employ a unified preprocessing and patch-based training protocol implemented with the MONAI framework, including isotropic resampling, orientation normalization, intensity standardization, and ROI-based sampling.
In our experiments, SegResNet achieved the best Dice score of **0.780**, outperforming SwinUNETR (Dice 0.731) and the 3D U-Net baseline (Dice 0.616). Beyond segmentation accuracy, we further demonstrate two practical downstream modules enabled by the predicted masks: (i) centerline-based diameter estimation via 3D skeletonization and Euclidean distance transform (EDT), and a distribution-based comparison between ground truth and predictions using quantile-profile MSE; and (ii) a proof-of-concept projection-based candidate visualization module that applies a YOLOv8 detector to three orthogonal 2D projections, reported qualitatively (no standardized stenosis annotations are available for quantitative evaluation in ImageCAS). These additions support interpretable vessel geometry analysis and provide practical intermediate visual artifacts for downstream work.

Keywords: Coronary arteries · CT angiography · 3D segmentation · SegResNet · SwinUNETR · MONAI · ImageCAS · Centerline · Diameter estimation · YOLO

1 Introduction

Cardiovascular diseases remain among the leading causes of mortality worldwide. Coronary CT angiography (CTA) is widely used for non-invasive assessment of coronary anatomy, offering volumetric visualization of the coronary tree. Automatic 3D coronary artery segmentation enables efficient extraction of vessel

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

geometry, supports reproducible quantitative analysis, and serves as a foundation for subsequent tasks such as centerline extraction, plaque analysis, and stenosis assessment.

Despite rapid progress in deep learning, coronary artery segmentation in CTA remains challenging due to small vessel calibers, low contrast in distal branches, motion artifacts, and strong class imbalance (thin vessels vs. large background volume). Public benchmarks, such as ImageCAS, improve comparability across methods and accelerate the development of robust models.

This work focuses on *3D coronary artery segmentation* as the core component, and additionally reports practical *downstream experiments* that become feasible once a reliable 3D vessel mask is available: (1) extraction of a centerline proxy and estimation of local diameters for quantitative vessel profiling, and (2) stenosis-candidate localization using a lightweight 2D detector on orthogonal projections. These modules are intended as a step toward a full analysis chain, while keeping the segmentation benchmark evaluation as the main quantitative focus.

Contributions. Our main contributions are:

- We provide a unified and reproducible training/evaluation protocol for 3D coronary artery segmentation on ImageCAS, including consistent preprocessing, patch-based ROI sampling, and case-level splitting.
- We benchmark three modern 3D backbones (3D U-Net, SegResNet, SwinUNETR) under the same setup and report comparative quantitative results, highlighting SegResNet as the strongest performer in our experiments.
- We demonstrate a lightweight downstream geometry module that converts 3D vessel masks into centerline-based diameter profiles in millimeters and introduces a quantile-profile MSE for distribution-level comparison between prediction and ground truth.
- We include a qualitative proof-of-concept projection-based candidate visualization pipeline using YOLOv8 on orthogonal projections, intended as an illustrative intermediate step rather than a validated stenosis detector.
- We release code and scripts to support replication and further extensions (Section *Code Availability*).

2 Related Work

Early learning-based 3D medical segmentation was shaped by fully convolutional architectures, including 3D U-Net [4] and V-Net [5]. Later, strong baseline pipelines such as nnU-Net [6] demonstrated that careful system configuration can yield highly competitive results across datasets.

Residual CNN designs are often preferred for volumetric segmentation due to stable optimization and efficient inference. In particular, residual encoder-decoder variants have been widely adopted in brain tumor segmentation and beyond [7]. In parallel, transformer-based architectures have gained traction,

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

with Swin Transformer variants adapted to dense prediction; SwinUNETR [8] is a prominent example for 3D medical imaging.

For coronary artery segmentation specifically, multiple approaches have been explored in CTA, including traditional model-based pipelines and deep learning methods that leverage local vessel appearance and topology. Recent studies continue to report progress, but also highlight persistent difficulties in generalization and distal branch delineation.

Benchmark datasets are essential for reliable comparison. The ImageCAS dataset and benchmark were introduced to support large-scale evaluation of coronary artery segmentation methods in CTA [2]. The dataset is also distributed through Kaggle[3].

Finally, reproducible medical imaging research increasingly relies on open-source toolkits. In this work, we build upon MONAI [3], which provides domain-specific transforms, inference utilities, and validated network implementations.

For downstream vessel geometry analysis, skeletonization-based centerline proxies are commonly derived via 3D thinning algorithms (e.g., Lee *et al.* [9]). EDT is a standard tool for computing distances inside binary objects and is widely used for radius estimation; linear-time exact EDT algorithms are described by Maurer *et al.* [10]. For stenosis localization, modern object detectors such as YOLOv8 can be applied to 2D projections as a pragmatic first-stage candidate generator [11].

3 Dataset and Data Manifest

We use the ImageCAS coronary artery dataset [2], consisting of 3D CTA volumes and coronary artery annotations (binary masks). For practical replication, we follow the Kaggle distribution of the dataset.

Each case is stored as a NIfTI pair with a consistent naming scheme: `X.img.nii.gz` for the CTA volume and `X.label.nii.gz` for the corresponding binary mask, typically located in the same folder. A case is defined as the *volume+mask* pair; we do not split the data at the slice/patch level.

All segmentation experiments are driven by a CSV manifest with columns: `case_id`, `image_path`, `mask_path`, and `split` (train/test). If `split` is not provided, we create it programmatically (see Section 5).

Case discovery and pairing. Raw pairs are discovered recursively under `raw_data/**` by matching `*.img.nii.gz`. For each volume file `X.img.nii.gz`, we look for the corresponding mask `X.label.nii.gz` in the same directory. If the mask is missing, the case is skipped (tracked via a `missing_labels` counter in the split script).

Split unit (case-level). The split is performed *by case* (patient), not by slices or patches. The `case_id` is extracted from the image filename prefix `X` (the substring before `.img.nii.gz`) and normalized via `norm_id()` (e.g., `"1.0"` → `"1"`).

[3] `https://www.kaggle.com/datasets/xiaoweixumedicalai/imagecas/data`

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

Post-split layout and manifest generation. After splitting, the script copies the paired files (`MODE="copy"` by default) into:

- `data/train/images/{case_id}.nii.gz`,
- `data/train/masks/{case_id}.nii.gz`,
- `data/test/images/{case_id}.nii.gz`,
- `data/test/masks/{case_id}.nii.gz`.

It also writes `data/manifest.csv` with columns `case_id`, `image_path`, `mask_path`, and `split`.

For downstream experiments (Experiments 1–2 below), we additionally support a directory-based layout (e.g., `data/test/images` and `data/test/masks`) for direct evaluation and visualization.

4 Methods

We evaluate three 3D segmentation backbones under a unified training and evaluation protocol:

- **3D U-Net (baseline).** A standard encoder–decoder architecture with skip connections.
- **SegResNet.** A residual CNN encoder–decoder implemented in MONAI, using 3D convolutions and residual blocks.
- **SwinUNETR.** A transformer-based model with hierarchical Swin Transformer encoder blocks and a U-Net-like decoder for dense 3D prediction.

Preprocessing and ROI sampling.

To isolate the effect of the backbone, all models share the same preprocessing pipeline and sampling strategy:

- **Orientation normalization.** All volumes are reoriented to the canonical RAS coordinate system to ensure consistent anatomical axes across subjects.
- **Resampling.** CTA images are resampled to isotropic spacing $(1.0, 1.0, 1.0)$ mm using trilinear interpolation. Binary masks are resampled to the image grid with nearest-neighbor interpolation to preserve label topology and avoid mixing.
- **Intensity normalization.** We apply nonzero, channel-wise z-score normalization. With **nonzero** statistics, the mean and standard deviation are computed only over voxels with nonzero intensity, which prevents padded / background regions (introduced by fixed-size cropping and padding) from biasing normalization. With **channel-wise** normalization, statistics are computed independently per channel; since CTA is single-channel, this reduces to one (μ, σ) pair per volume.
- **Patch-based training.** Because coronary arteries occupy a small fraction of the volume, we train on fixed-size ROIs of $96 \times 96 \times 96$ voxels and use label-aware sampling (`RandCropByPosNegLabel`) to mitigate class imbalance. This increases the probability that a crop contains vessel voxels while retaining background-only crops for specificity.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

Augmentations and inference.

During training, we apply lightweight 3D augmentations (random rotations, flips, zoom/affine transforms, and intensity/noise perturbations) to improve robustness to anatomical variability and acquisition differences. At inference time, predictions are produced on full volumes; when memory constraints require it, we use sliding-window inference with overlap and blending to obtain consistent evaluation across backbones.

Downstream Geometry: Centerline and Diameter Estimation

Given a binary vessel mask $M \in \{0,1\}^{X\times Y\times Z}$, we derive a sparse centerline proxy C using 3D skeletonization (medial axis / thinning) [9]. We then compute a voxel-wise Euclidean distance transform inside the vessel:

$$D(\mathbf{p}) = \mathrm{EDT}(M)(\mathbf{p}), \quad \mathbf{p} \in \Omega, \tag{1}$$

where EDT returns the distance from voxel $\mathbf{p}$ to the closest background voxel. When voxel spacing is anisotropic, EDT is computed in physical units by supplying voxel spacing (sampling) [10].

For each centerline voxel $\mathbf{p} \in C$, we estimate radius and diameter in millimeters:

$$r(\mathbf{p}) = D(\mathbf{p}), \quad d(\mathbf{p}) = 2D(\mathbf{p}). \tag{2}$$

This yields a diameter sample set $\{d(\mathbf{p})\}_{\mathbf{p}\in C}$, which can be used for quantitative profiling and stenosis-candidate analysis.

Distribution-Based Comparison of Diameter Profiles

Centerline points from different masks (ground truth vs. prediction) do not correspond one-to-one. To compare diameter distributions robustly, we compute a fixed-length *quantile profile*. Given diameter samples $d_1, \ldots, d_n$, we define a quantile vector $\mathbf{q} \in \mathbb{R}^K$:

$$\mathbf{q}[k] = Q\left(\frac{k}{K-1}\right), \quad k = 0, \ldots, K-1, \tag{3}$$

where $Q(\tau)$ is the empirical quantile function at probability τ. We then compare the ground truth and predicted profiles via mean squared error:

$$\mathrm{MSE}(\mathbf{q}^{gt}, \mathbf{q}^{pr}) = \frac{1}{K}\sum_{k=0}^{K-1}\left(\mathbf{q}^{gt}[k] - \mathbf{q}^{pr}[k]\right)^2. \tag{4}$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

Projection-Based Candidate Visualization with YOLOv8 (Qualitative)

As a proof-of-concept downstream visualization module, we generate three orthogonal 2D projections of a 3D vessel mask (and, optionally, the corresponding CTA volume), corresponding to axial/coronal/sagittal views. We then apply a YOLOv8 detector [11] independently to each projection to obtain bounding boxes and confidence scores that highlight *candidate regions* for further inspection.

Importantly, we do not claim clinically validated stenosis detection in this work. In the absence of standardized stenosis annotations within ImageCAS, this module is presented as a qualitative candidate visualization step that produces interpretable intermediate artifacts and can be used as a starting point for a rigorously annotated detection protocol.

5 Training and Evaluation Protocol

The complete training and evaluation configuration is summarized in Table 1. Below we provide brief clarifications for individual components.

Loss and Optimization

We train with a Dice-based objective (softmax, two classes: background and vessel). For SegResNet and SwinUNETR we additionally use a weighted cross-entropy term:

$$\mathcal{L} = \lambda_{\text{dice}}\mathcal{L}_{\text{Dice}} + \lambda_{\text{ce}}\mathcal{L}_{\text{CE}}. \tag{5}$$

We use Adam/AdamW optimizers with gradient clipping and cosine annealing with warmup for selected models.

Splits

If the manifest includes `split`, we strictly respect it. Otherwise, we generate a reproducible train/test split programmatically at the *case level* (by `case_id`) to avoid information leakage between patches/slices from the same subject.

Concretely, we collect all eligible `case_id`s from discovered `volume+mask` pairs (Section 5, Materials). We then apply `sklearn.model_selection.train_test_split` with `shuffle=True`, `test_size=0.20`, and a fixed seed `SEED=42` (`random_state=42`). This results in an 80%/20% split (train/test) that is fully reproducible.

Train/val/test note. The provided split script creates only `train` and `test`. If a separate validation set is required, it should be obtained by an additional split of the training pool (using the same case-level procedure); however, this step is not part of the referenced script and is not assumed by default.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

Table 1. Training setup used in all segmentation experiments.

Component	**Setting**
Dataset	ImageCAS CTA volumes with binary coronary artery masks (Kaggle distribution).
Split unit	Case-level (`case_id`); no slice/patch-level splitting.
Train/test split	`train_test_split` with `shuffle=True`, `test_size=0.20`, `random_state=42` ⇒ 80% train / 20% test.
Validation	Not created by the referenced split script; can be obtained by an additional case-level split of the training pool if required.
Preprocessing	Orientation to RAS; resampling to $(1.0, 1.0, 1.0)$ mm (labels: nearest-neighbor); nonzero, channel-wise intensity normalization.
Training strategy	Patch-based 3D training with ROI sampling to mitigate class imbalance.
ROI size	$96 \times 96 \times 96$ voxels.
ROI sampling	Positive/negative label sampling (foreground-aware).
Augmentations	Lightweight 3D augments: random rotations/flips, zoom/affine, intensity/noise perturbations.
Framework	MONAI-based pipeline for preprocessing, training, and inference.
Backbones	3D U-Net (baseline), SegResNet, SwinUNETR.
Loss	Dice loss (two classes: background/vessel); for SegResNet and SwinUNETR: Dice + weighted cross-entropy.
Optimizer & schedule	Adam/AdamW; gradient clipping; cosine annealing with warmup for selected models (as configured in scripts).
Inference	Sliding-window inference with Gaussian blending.
Inference params	ROI 96^3, overlap 0.25, batch size 1.
Metric	Dice similarity coefficient (DSC), computed for the vessel class (background excluded).

Intensity normalization (nonzero, channel-wise).

We normalize CTA intensities using MONAI's `NormalizeIntensity` transform with `nonzero=True` and `channel_wise=True`. Let $x \in \mathbb{R}^{C\times H\times W\times D}$ denote the input volume. With `channel_wise=True`, normalization is performed independently for each channel c (in our case $C = 1$), using per-volume statistics computed over a selected set of voxels Ω_c. The option `nonzero=True` defines Ω_c as the set of voxels with nonzero intensity, i.e., $\Omega_c = \{v \mid x_c(v) \neq 0\}$, and computes

$$\mu_c = \frac{1}{|\Omega_c|} \sum_{v\in\Omega_c} x_c(v), \qquad \sigma_c = \sqrt{\frac{1}{|\Omega_c|} \sum_{v\in\Omega_c} \left(x_c(v) - \mu_c\right)^2 + \epsilon}, \tag{6}$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

followed by the z-score transformation

$$\hat{x}_c(v) = \frac{x_c(v) - \mu_c}{\sigma_c}, \quad \forall v \in \Omega_c. \tag{7}$$

Voxels outside the selected set (i.e., $x_c(v) = 0$) are left unchanged, which prevents padded/background regions from dominating the mean and variance. This is particularly important in our patch-based pipeline where `SpatialPadd` introduces large zero-valued areas to match a fixed ROI size; without `nonzero=True`, these zeros would bias μ_c toward 0 and shrink σ_c, leading to weaker normalization inside the anatomy. Finally, since we operate on single-channel CTA volumes, `channel_wise=True` reduces to computing a single pair (μ, σ) per volume, ensuring consistent scaling across all patches sampled from that volume.

Inference

Because CTA volumes are high-resolution 3D scans, direct full-volume inference can exceed GPU memory for the larger backbones. We therefore evaluate models using sliding-window inference with Gaussian blending, which is standard for patch-based 3D segmentation. Specifically, we partition the volume into overlapping ROIs of size 96^3 voxels and run the network on each ROI independently. Overlapping predictions are merged by a spatially varying Gaussian weight map that assigns higher weight to the center of each ROI and lower weight to its borders, thereby reducing boundary artifacts caused by limited receptive field and padding at patch edges. We use an overlap of 0.25 and sliding-window batch size 1, which provides a stable trade-off between memory usage and throughput. For backbones that fit into memory, the same procedure yields identical output to a full-volume forward pass up to numerical precision, while for memory-constrained settings it enables consistent evaluation across all architectures.

Metrics

We report the Dice similarity coefficient (DSC) as the primary segmentation metric:

$$\mathrm{Dice}(A, B) = \frac{2|A \cap B|}{|A| + |B|}, \tag{8}$$

where A is the predicted vessel mask and B is the ground-truth mask. Dice is computed on the foreground class only (background excluded). We choose Dice because (i) it directly measures overlap and is widely used for medical segmentation, enabling comparison to prior work, and (ii) it partially mitigates class imbalance by emphasizing agreement on the relatively small foreground region compared to the background. In coronary CTA, vessels occupy a small fraction of the volume; therefore, metrics dominated by true negatives (e.g., voxel accuracy) are less informative, while Dice provides a more meaningful summary of segmentation quality.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

At the same time, we note that Dice can be insensitive to clinically relevant errors in thin distal branches and along vessel boundaries: small spatial shifts may yield a limited change in overlap despite substantial geometric deviation. For this reason, we complement Dice with geometry-aware analyses. In addition to the downstream diameter-profile study (Experiment 1), we report the 95th-percentile Hausdorff distance (HD95), the absolute error in total skeleton-based centerline length, and the absolute error in the number of 26-connected components as a simple proxy for branch continuity.

6 Experiments and Results

Unless stated otherwise, the quantitative segmentation results below correspond to a paired reevaluation of the best saved checkpoint of each model on the full 200-case test split. Table 2 summarizes the primary overlap-based results.

Table 2. Primary segmentation results on ImageCAS (mean ± standard deviation across 200 test cases).

Model	Dice	IoU	Parameters (M)
3D U-Net (baseline)	0.616 ± 0.070	0.449 ± 0.072	22.625
SegResNet (MONAI)	**0.780 ± 0.050**	**0.642 ± 0.066**	18.796
SwinUNETR (MONAI)	0.731 ± 0.055	0.579 ± 0.066	15.703

To assess whether the observed Dice differences are consistent at the case level, we additionally performed paired Wilcoxon signed-rank tests. The differences were statistically significant for all model pairs: SegResNet vs. SwinUNETR ($p = 1.60 \times 10^{-34}$), SegResNet vs. 3D U-Net ($p = 1.44 \times 10^{-34}$), and SwinUNETR vs. 3D U-Net ($p = 1.44 \times 10^{-34}$).

Table 3. Geometry-aware validation metrics on ImageCAS (mean ± standard deviation across 200 test cases; lower is better). Here, $|\Delta L_{cl}|$ is the absolute error of total skeleton-based centerline length and $|\Delta \#CC|$ is the absolute difference in the number of 26-connected components.

Model	HD95 (mm) ↓	$\lvert\Delta L_{cl}\rvert$ (mm) ↓	$\lvert\Delta \#CC\rvert$ ↓
3D U-Net (baseline)	30.69 ± 6.89	159.23 ± 41.58	44.13 ± 9.80
SegResNet (MONAI)	**10.10 ± 2.88**	**69.38 ± 17.30**	**8.23 ± 2.16**
SwinUNETR (MONAI)	15.82 ± 4.56	91.47 ± 18.83	35.24 ± 10.25

To illustrate typical successes and failure modes, we include multi-planar overlays (sagittal/coronal/axial) of predicted masks against ground truth (Fig. 1).

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

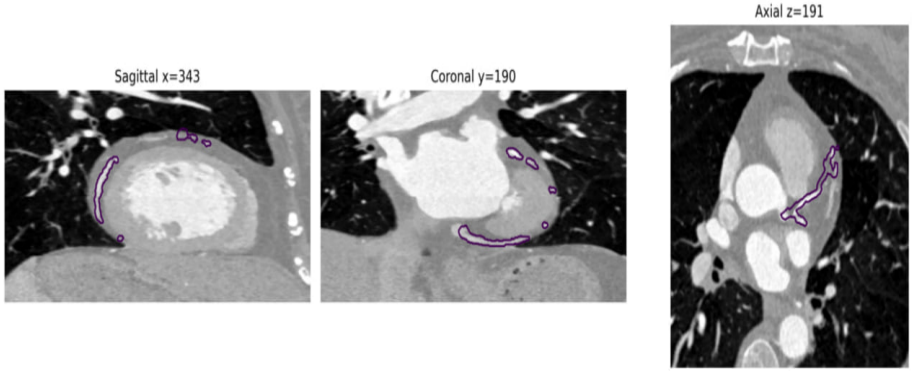

Fig. 1. Qualitative comparison of 3D coronary artery segmentation (example multi-planar overlays).

6.1 Experiment 1: Centerline-Based Diameter Profiles and Quantile-Profile MSE

In Experiment 1, we evaluate how segmentation errors impact downstream vessel geometry. For each test case, we compute centerlines and diameter samples for both ground truth and predicted masks (Section 4). We then compare diameter distributions via quantile-profile MSE (with K quantiles, fixed across cases).

Quantile resolution. We set the number of quantiles to $K = 1000$ (configurable via `-n-quantiles`, default: 1000), using uniformly spaced quantile levels in $[0, 1]$.

Invalid and failed cases. We distinguish (a) *failed* cases due to I/O or preprocessing errors and (b) *degenerate* cases with empty masks/centerlines. Failed cases are recorded with an error flag and excluded from summary statistics; we report the total number of failed cases. For degenerate cases, skeletonization returns an empty centerline (0 points), and the corresponding quantile profile is defined as an all-zero vector. We explicitly log the number of centerline points for both ground truth and prediction (`gt_centerline_points`, `pred_centerline_points`), enabling transparent reporting of how many cases are affected.

Pipeline scheme. Figure 2 provides a schematic overview of Experiment 1, from segmentation masks to centerline extraction, diameter estimation and quantile-profile comparison.

3D visualization with centerline and sampling markers. To produce visuals, we render the 3D vessel surface together with the extracted centerline (red) and a chain of sampling markers (green spheres) placed along the centerline (Fig. 3). This visualization is used both for debugging and for communicating the measurement procedure in a human-interpretable form.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

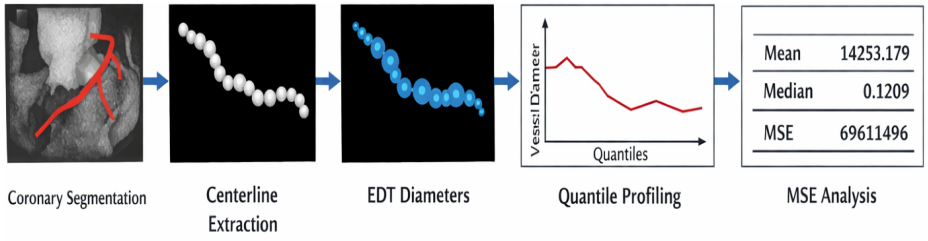

Fig. 2. Experiment 1 scheme: segmentation mask → centerline → EDT diameters → quantile-profile MSE.

Quantitative result. Across the evaluated test set, we obtained the following summary statistics for quantile-profile MSE:

$$\mathrm{MSE}_{mean} = 0.1425, \quad \mathrm{MSE}_{median} = 0.1301. \tag{9}$$

Remark on outliers. The gap between mean and median indicates the presence of outliers (e.g., rare cases with degenerate centerlines, empty diameter sets, or severe prediction artifacts).

6.2 Experiment 2: Orthogonal Projections and YOLOv8 Stenosis Candidate Detection

In Experiment 2, we present a qualitative proof-of-concept module that produces candidate visual cues on 2D projections derived from 3D vessel masks (Section 4). Specifically, we compute three orthogonal projections (axial/coronal/sagittal) and apply YOLOv8 to each view to obtain bounding boxes and confidence scores. The goal is to demonstrate a practical intermediate artifact for downstream analysis and visualization, rather than to claim a validated stenosis detection system.

Pipeline scheme. Figure 4 summarizes the process. We interpret the detector outputs as *candidate regions* that would require subsequent 3D refinement and rigorous evaluation on datasets with standardized stenosis annotations.

Qualitative visualization. Figure 5 provides an example panel of the three projections with predicted candidate boxes. In this work we restrict evaluation to qualitative inspection, as quantitative detection metrics would require consistent lesion-level ground truth.

7 Discussion

Our segmentation experiments indicate that a residual CNN design (SegResNet) provides the best Dice score among the tested backbones under a unified prepro-

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

Fig. 3. Example 3D rendering of a segmented coronary tree with extracted centerline (red) and sampling markers (green spheres) placed along the centerline (10+ markers).

cessing and patch-based training setup. While SwinUNETR offers transformer-based contextual modeling, its Dice score in our setting was lower than SegResNet. The baseline 3D U-Net lags behind both newer architectures, consistent with broader trends in volumetric segmentation benchmarks. The paired per-case statistics and geometry-aware metrics follow the same ranking, with SegResNet achieving the lowest HD95, the smallest centerline-length error, and the lowest connected-component count error, which suggests better preservation of vessel continuity.

Beyond segmentation, we demonstrated two practical downstream modules. Experiment 1 showed that centerline-based diameter profiling can be built directly from predicted masks, enabling interpretable measurements (diameters in millimeters) and distribution-based comparisons between prediction and ground truth. The observed mean/median discrepancy in MSE suggests that robust handling of rare degenerate cases is important for reliable reporting.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

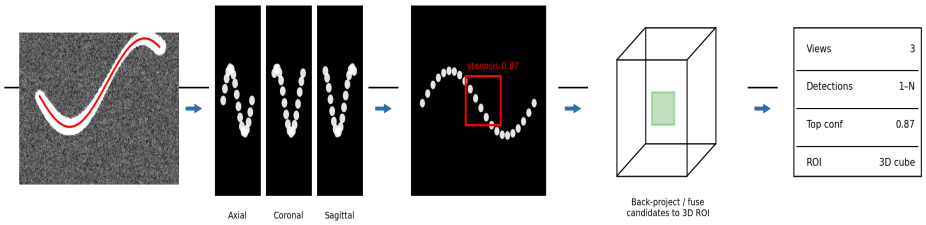

Fig. 4. Experiment 2 scheme: 3D mask → 3 orthogonal 2D projections → YOLOv8 detections → candidate ROIs.

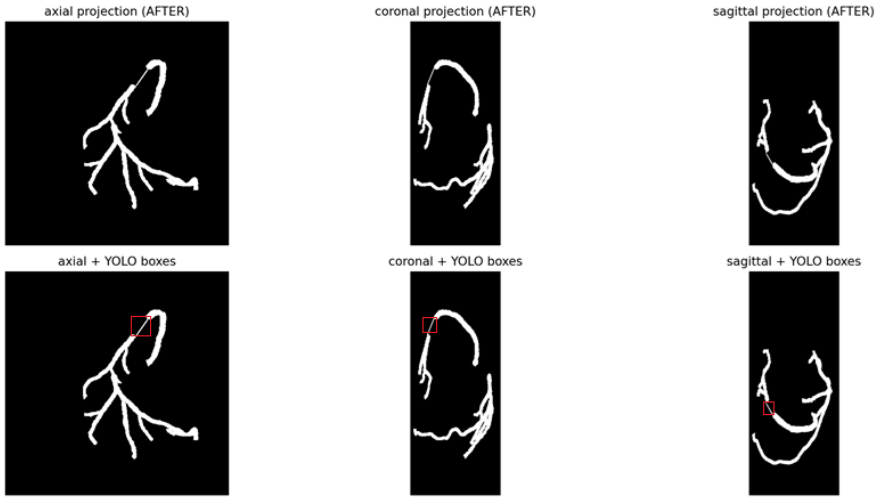

Fig. 5. Experiment 2 qualitative example: orthogonal projections (axial/coronal/sagittal) and YOLOv8 stenosis candidate bounding boxes.

Experiment 2 demonstrated an initial stenosis candidate localization approach using YOLOv8 on orthogonal projections. This approach is attractive because it is lightweight and provides intuitive 2D evidence for clinicians and researchers. However, rigorous detection evaluation requires standardized stenosis annotations and consistent projection conventions; we plan to finalize this protocol and report detection metrics in subsequent work.

8 Conclusion

We presented an experiment-driven study of 3D coronary artery segmentation on the ImageCAS CTA dataset, comparing a 3D U-Net baseline, SegResNet, and SwinUNETR under a shared training and evaluation protocol. SegResNet achieved the best Dice score (0.780) in our experiments. We additionally extended the study with two downstream experiments: centerline-based diameter

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

estimation with quantile-profile MSE comparison, and a prototype projection-based stenosis candidate detection module using YOLOv8. These additions move toward an end-to-end analysis pipeline that combines strong segmentation with interpretable geometry and practical visualization. In future work, we plan to validate both the segmentation models and the geometry-aware metrics on an additional external coronary CTA dataset.

Limitations

Our study has several limitations:

- **Single-dataset evaluation.** All quantitative segmentation results are obtained on ImageCAS only; no external validation is performed, which limits conclusions about generalization. We therefore plan a future validation study on an additional external coronary CTA dataset.
- **Sensitivity to preprocessing choices.** Performance may depend on resampling to isotropic spacing and on the ROI sampling strategy; these choices can affect thin structures and boundary consistency.
- **Thin-vessel failure modes.** Like many CTA segmentation systems, errors are more frequent in distal branches due to low contrast, motion artifacts, and extreme class imbalance.
- **Detection module is qualitative.** The projection-based YOLOv8 module is presented as a proof-of-concept candidate visualization step; without standardized stenosis annotations, we do not report quantitative detection metrics.

Code Availability

The code and experiment scripts are available in our public repository: `https://github.com/slikyumsh/Stenosis3D`

Acknowledgments

The research was supported by MEDVISION LLC (ITN 7804703197).

References

1. Evtyukhov, D., Kopanitsa, G., Metsker, O., Mogilevskii, A., Yakovlev, A., Kovalchuk, S.: Automatic Detection and Segmentation of Coronary Artery Stenosis in Coronary Angiography Images. Computational Science – ICCS 2025 Workshops. Lecture Notes in Computer Science, vol. 15908, pp. 128–140. Springer, Cham (2025).
2. Zeng, X., et al.: ImageCAS: A Large-Scale Dataset and Benchmark for Coronary Artery Segmentation Based on CT Angiography. arXiv preprint arXiv:2211.01607 (2022)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36

3. Cardoso, M.J., et al.: MONAI: An Open-Source Framework for Deep Learning in Healthcare. arXiv preprint arXiv:2211.02701 (2022)
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: MICCAI (2016)
5. Milletari, F., Navab, N., Ahmadi, S.-A.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: 3DV (2016)
6. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. Nature Methods **18**, 203–211 (2021)
7. Myronenko, A.: 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization. arXiv preprint arXiv:1810.11654 (2018)
8. Hatamizadeh, A., et al.: Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. arXiv preprint arXiv:2201.01266 (2022)
9. Lee, T.-C., Kashyap, R.L., Chu, C.-N.: Building Skeleton Models via 3-D Medial Surface/Axis Thinning Algorithms. CVGIP: Graphical Models and Image Processing **56**(6), 462–478 (1994)
10. Maurer, C.R., Qi, R., Raghavan, V.: A Linear Time Algorithm for Computing Exact Euclidean Distance Transforms of Binary Images in Arbitrary Dimensions. IEEE Trans. Pattern Anal. Mach. Intell. **25**(2), 265–270 (2003)
11. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLOv8 (software), 2023. Available: `https://docs.ultralytics.com/models/yolov8/`
12. Dice, L.R.: Measures of the Amount of Ecologic Association Between Species. Ecology **26**(3), 297–302 (1945)
13. Sørensen, T.: A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content. Biologiske Skrifter **5**, 1–34 (1948)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_36