Deep Learning Methods for Coronary Artery Segmentation from Computed Tomography Angiograms

Tushar Nayak, Jared Scott, Pranav Kini

Abstract

Accurate 3D reconstruction of coronary arteries from Computed Tomography Angiography (CTA) is critical for diagnosing stenosis, planning stent interventions, and guiding robotic surgery. However, manual segmentation of these intricate, narrow vascular structures is labor-intensive and prone to inter-observer variability. This paper presents an end-to-end deep learning framework for the automated segmentation and geometric reconstruction of the coronary tree. We employ a ResUNet3D architecture, optimized via mixed-precision training to handle high-dimensional volumetric data efficiently. To address the severe class imbalance inherent in vascular imaging, we utilize a hybrid loss function combining Dice Similarity Coefficient and Binary Cross Entropy (BCE). Beyond voxel-wise prediction, our pipeline integrates a robust post-processing stage involving morphological operations, Marching Cubes surface extraction, and Laplacian smoothing. We demonstrate that this approach not only identifies vascular structures with high fidelity but also generates clinically actionable 3D meshes and point clouds. An ablation study confirms that the hybrid loss function significantly reduces surface artifacts compared to Dice loss alone, bridging the gap between raw neural network predictions and surgical usability.

INTRODUCTION

Coronary Computed Tomography Angiography (CCTA) has established itself as the cornerstone of non-invasive cardiovascular diagnosis, providing rich volumetric data essential for assessing coronary artery disease (CAD). However, the clinical utility of these scans is frequently bottlenecked by the dimensionality. While modern scanners capture the coronary tree in sub-millimeter detail, translating these dense volumetric arrays into actionable geometric models for hemodynamic simulation or robotic navigation remains a laborious manual process. The coronary arteries characterized by their tortuous topology, thin diameters, and variable contrast intensity, present a unique challenge where human segmentation is not only time-prohibitive but highly operator-dependent.

The advent of deep learning, particularly Fully Convolutional Networks (FCNs), has promised to automate this workflow. Yet, a significant gap remains between achieving high pixel-wise accuracy metrics and producing anatomically coherent models suitable for surgical planning. Raw network predictions often suffer from disconnected components, staircase artifacts, and surface noise that render them unusable for downstream computational fluid dynamics (CFD) or intraoperative guidance.

To address these limitations, we propose a comprehensive, end-to-end framework for coronary artery segmentation and reconstruction. Our approach integrates a ResUNet3D architecture—leveraging residual pathways to stabilize gradient flow across deep volumetric layers, with a specialized geometric post-processing pipeline. We introduce a hybrid loss function designed to maximize overlap while enforcing boundary precision, and we detail a computationally efficient workflow utilizing mixed-precision training. Crucially, this work moves beyond simple voxel classification; we focus on the transformation of probability maps into watertight 3D meshes and high-density point clouds. This pipeline effectively bridges the divide between deep learning abstraction and the rigorous geometric requirements of modern interventional cardiology.

1 BACKGROUND

1.1 Patch-based Segmentation

In 2018, a paper by Huang et al. proposed an approach to artery segmentation from CTCAs using a 3-D U-Net convolutional neural network (CNN) framework. They devised a U-Net structure consisting of 4 CNNs for encoding and 4 for decoding, sandwiching several additional layers for max pooling

and upsampling; the full network architecture is detailed in Figure 1. Taking an original dataset of 34 volumetric images, they divided these scans into 12,364 32 × 32 × 32 "patches" for training and testing the U-Net (Huang et al., 2018).

Layer (Type)	Input Size	Output Size	Kernel Size	Filter Number
Conv3D_1	1x64x64x16	32x64x64x16	3x3x3	32
Conv3D_1	32x64x64x16	64x64x64x16	3x3x3	64
MaxPooing1	64x64x64x16	64x32x32x8	2x2x2	-
Conv3D_2	64x32x32x8	64x32x32x8	3x3x3	64
Conv3D_2	64x32x32x8	128x32x32x8	3x3x3	128
MaxPooling2	128x32x32x8	128x16x16x4	2x2x2	-
Conv3D_3	128x16x16x4	128x16x16x4	3x3x3	128
Conv3D_3	128x16x16x4	256x16x16x4	3x3x3	256
MaxPooling3	256x16x16x4	256x8x8x2	2x2x2	-
Conv3D_4	256x8x8x2	256x8x8x2	3x3x3	256
Conv3D_4	256x8x8x2	512x8x8x2	3x3x3	512
UpSampling1	512x8x8x2	512x16x16x4	2x2x2	-
Concatenate1	512x16x16x4	768x16x16x4	-	+256
Conv3D_5	768x16x16x4	256x16x16x4	3x3x3	256
Conv3D_5	256x16x16x4	256x16x16x4	3x3x3	256
UpSampling2	256x16x16x4	256x32x32x8	2x2x2	-
Concatenate2	256x32x32x8	384x32x32x8	-	+128
Conv3D_6	384x32x32x8	128x32x32x8	3x3x3	128
Conv3D_6	128x32x32x8	128x32x32x8	3x3x3	128
UpSampling3	128x32x32x8	128x64x64x16	2x2x2	-
Concatenate3	128x64x64x16	192x64x64x16	-	+64
Conv3D_7	192x64x64x16	64x64x64x16	3x3x3	64
Conv3D_7	64x64x64x16	64x64x64x16	3x3x3	64
Conv3D_8	64x64x64x16	1x64x64x16	1x1x1	1
Activation	1x64x64x16	1x64x64x16	-	-

Figure 1. U-Net architecture; adapted from Huang et al. (2018).

After testing the U-Net model trained on the CTCA patches over 15 epochs, Huang et al. yielded mean Dice similarity coefficients ranging from 0.71 to 0.78 (using labels from human experts as ground truth) (Huang et al., 2018). They compared these results with those from 4 other CNN architectures, as shown in Figure 2; the novel U-Net method provided mean Dice scores that were approximately 0.10 higher than those yielded by the comparative methods.

Our Method (DL, 3D, with or without centerline)	[7] DL, 3 views of 2D CTCA (no centerline)	[9] DL, DeepMedic, 3D (no centerline)	[11] Regression, graph-cut, (with centerline)	[15] Tissue probability, level-set (with centerline)
0.71-0.78	0.65	0.60	0.65-0.68	0.69-0.73

Figure 2. U-Net results; adapted from Huang et al. (2018).

The new patching + U-Net method proposed by Huang et al. brings the advantages of being trainable on small datasets, and resulting in higher accuracy for such datasets than other CNNs. Among its disadvantages, however, are that it can be prone to overfitting and requires a considerable amount of pre-processing to divide volumes into thousands of normalized patches.

One year after Huang et al. published their method and findings, Chen et al. proposed a new, more accurate segmentation method that similarly utilized a combination of patching + U-Net architecture. However, in addition to passing 32 × 32 × 32 segments of the the original CTA images as inputs to the U-Net, Chen et al. also used a Frangi vascular enhancement filter (Frangi et al., 2000) to extract a 32 × 32 × 32 vesselness map from each segment, and passed this in as a second input to the U-Net (Chen et al., 2019). This multichannel approach is depicted in Figure 3.

This multichannel patch-based approach achieved a Dice similarity coefficient of 0.8 on a testing set of CTA images; similar CNN approaches cited in the article achieved scores of only 0.5975 and 0.66 (Chen et al., 2019). Furthermore, the multichannel U-Net was found to outperform segmentation results from a set of human observers (of which the highest scored 0.79) (Chen et al., 2019).

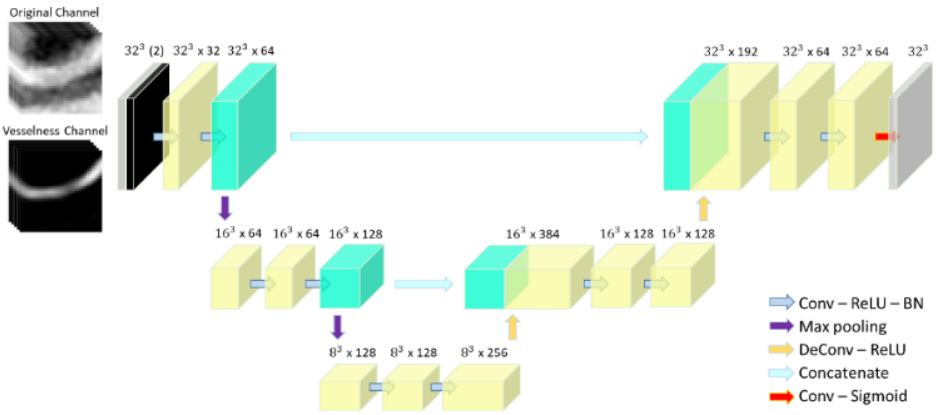


Figure 3. Multichannel U-Net segmentation; adapted from Chen et al. (2019).

The method proposed by Chen et al. appears to outperform the prior one by Huang et al., likely because the multichannel method places additional emphasis on the shape of the artery during the segmentation process via vascular enhancement filtering. However, the multichannel U-Net was trained on a relatively small data set (11 images) and remains to be benchmarked on a wide variety of CTA scans (Chen et al., 2019). Moreover, like Huang et al.'s method, it requires significant pre-processing to create the volumetric input patches, and additional pre-processing to create the Frangi filtered inputs. It also requires several post-processing steps to reconstruct the artery structure from the output of the U-Net (Chen et al., 2019).

1.2 Direct Segmentation and Hybrid Refinement

Moving beyond the constraints of small patches, Shen et al. (2019) proposed a "Direct Segmentation" framework designed to process larger volumetric contexts. They argued that while deep learning extracts semantic features well, standard CNN outputs often suffer from blurry, non-smooth boundaries. To address this, they introduced a hybrid architecture combining a 3D Fully Convolutional Network (FCN) with two critical innovations: an Attention Gate (AG) mechanism and a traditional Level Set function.

The authors utilized a 3D FCN with an encoder-decoder structure similar to U-Net to learn the 3D shape prior of the coronary arteries. However, acknowledging that small vessels are easily confused with background noise in CTA scans, they integrated Attention Gates into the skip connections. These gates automatically learn to suppress feature activation in irrelevant background regions while highlighting vessel-like structures, eliminating the need for a separate organ localization model.

To fix the rough boundaries typical of deep learning predictions, the method employs a coarse-to-fine strategy. The probability map output by the FCN-AG serves as the initialization for a traditional Level Set function, which iteratively evolves the contour to minimize energy and smooth the vessel edges.

Evaluated on a dataset of 70 patients (11,200 slices), this hybrid approach achieved a Mean Dice similarity coefficient of 0.9005, significantly outperforming the vanilla 3D FCN (0.8842). The results demonstrated that while deep learning provides excellent object recognition, the integration of traditional mathematical models (Level Sets) remains a powerful tool to achieve anatomical precision in segmentation tasks.

1.3 Tree-based Segmentation

While Chen et al. encoded the arterial structure as inputs to their deep learning model, Kong et al. (2020) proposed a method encoding the shape of the arteries within the model itself, using a tree-structured convolutional gated recurrent unit (TreeConvGRU). This model drew inspiration from long short-term memory (LSTM) neural networks commonly used to model semantic relatedness in natural language processing applications, but with additional local spatial correlations to "explicitly model the topological structure of the coronary artery" (Fig. 6) (Kong et al., 2020).

Kong et al. compared their TreeConvGRU architecture with a conventional, sequential ConvGRU as well as a DenseVox deep learning model given the task of segmenting arteries from a set of 916 CTA images. They found that the TreeConvGRU outperformed both models in terms of speed (25s as opposed to 26-58s) and accuracy (Dice score of 0.85 compared to 0.78-82) (Fig. 7) (Kong et al., 2020).

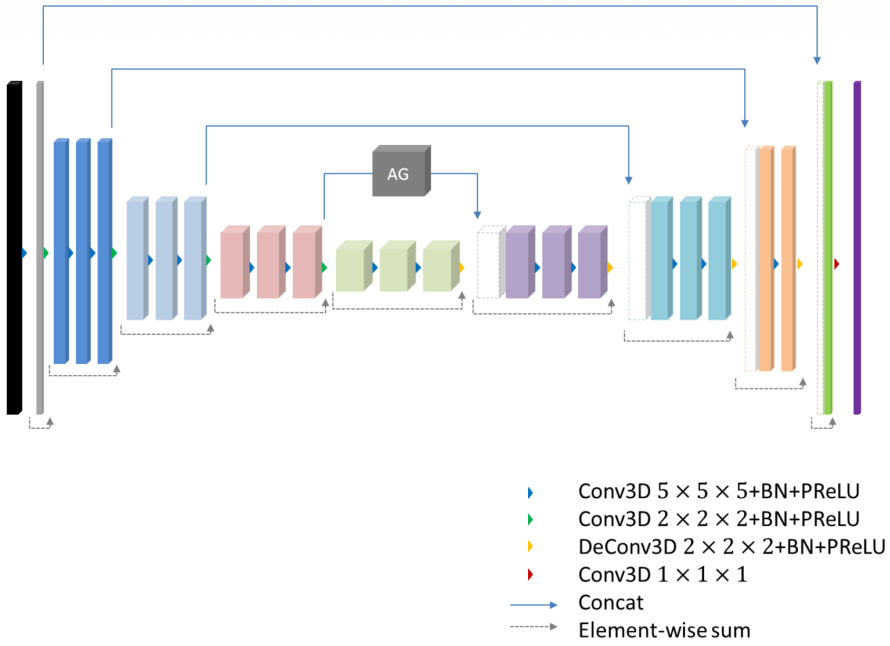


Figure 4. The joint framework combining 3D FCN, Attention Gates, and Level Set refinement; adapted from **?**.

Figure 5. Visual comparison showing the refinement from FCN to FCN+AG+LevelSet; adapted from **?**.

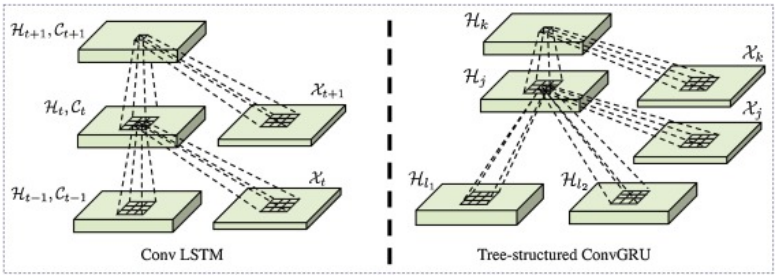


Figure 6. LSTM vs. TreeConvGRU structure; adapted from Kong et al. (2020).

Methods	DenseVox	ConvGRU	TreeConvGRU
Average dice	0.7806	0.8223	0.8537
Time (s)	58	26	**25**

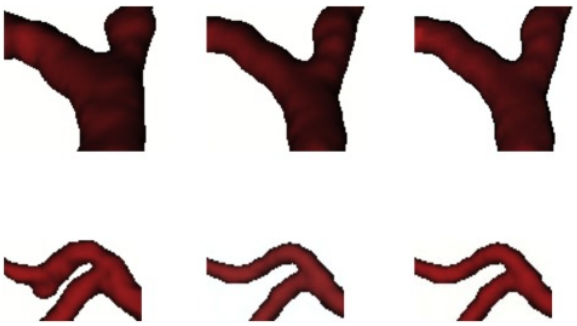


Figure 7. TreeConvGRU evaulation; adapted from Kong et al. (2020).

TreeConvGRU also achieved higher accuracy and speeds than purported by the previous patch-based methods, bringing the additional advantages of context propagation among nodes in order to preserve the topological structure of arteries within the model itself, not just the inputs/output. However, its generalizability across datasets has yet to be fully proven, as it relies on specific tree reconstruction data from a 2019 study to produce outputs (Kong et al., 2020). Furthermore, while the model itself runs quite quickly, it does require a somewhat time-consuming pre-segmentation step to prepare the data for processing.

1.4 Graph-based Segmentation

The final deep learning segmentation method surveyed for this project is a graph convolutional network (GCN) method proposed by Wolterink et al. in 2019. Like the tree-based method, it leverages the structure of coronary arteries within the predictive model architecture, but using a graph structure to aggregate information about each node relative to its neighbors (as shown in Figure 8) (Wolterink et al., 2019).

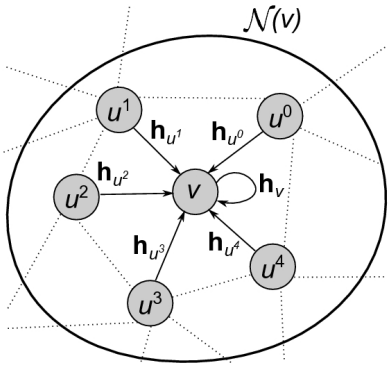


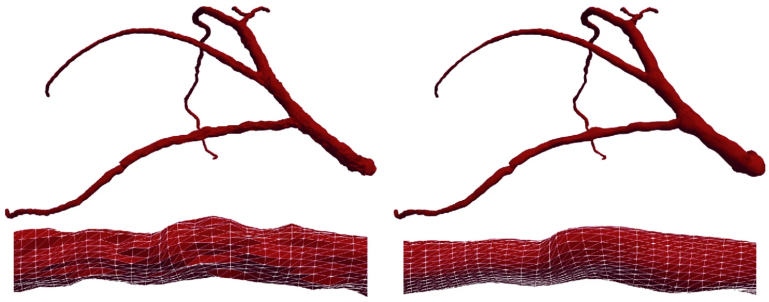
Figure 8. Graph convolutional network layer; adapted from Wolterink et al. (2019).

Wolterink et al. evaluated their GCN on 78 artery segments, and compared the results of reconstruction with those produced by 3 human experts and 4 other computational methods, including traditional fully-connected networks (FCN) and a multi-layer perceptron (MLP) model (Wolterink et al., 2019). the GCN produced a Dice similarity coefficient of 0.75 for healthy artery segments, and 0.73 for diseased ones, achieving higher accuracy than experts in about half of cases and than the computational methods in most instances (Fig. 9) (Wolterink et al., 2019). When compared to the MLP, the GCN method produced far smoother meshes in addition to more accurate segmentations, as shown in Figure 10 (Wolterink et al., 2019).

The GCN method requires no post-processing or regularization for smoothing (Wolterink et al., 2019). However, it uses a mean aggregator to concatenate GCN layers, meaning that the spatial relationships between nodes are not preserved between the layers (Wolterink et al., 2019); this loss of information could be mitigated by adding features between nodes in future work, and may account for the relatively low Dice accuracy compared to the previously discussed TreeConvGRU method.

Method	DSC		MSD (mm)		HD (mm)	
	Healthy	Diseased	Healthy	Diseased	Healthy	Diseased
Expert 1	0.74	0.79	0.26	0.26	3.61	3.29
Expert 2	0.66	0.73	0.25	0.31	3.00	2.70
Expert 3	0.80	0.76	0.23	0.24	3.25	3.07
Lugauer et al. [10]	0.77	0.75	0.32	0.27	2.79	1.96
Freiman et al. [2]	0.69	0.74	0.49	0.28	1.69	1.22
Mohr et al. [5]	0.75	0.73	0.45	0.29	3.73	1.87
Graph convolutional network (GCN)	0.75	0.73	0.25	0.28	1.53	1.86
Multi-layer perceptron (MLP)	0.67	0.69	0.32	0.31	1.59	1.84

Figure 9. GCN results and comparison; adapted from Wolterink et al. (2019).



(a) MLP (b) GCN

Figure 10. GCN vs MLP segmentation; adapted from Wolterink et al. (2019).

1.5 Large-Scale Benchmarking and Hybrid Fusion

While the patch, tree, and graph-based innovations described above marked significant theoretical progress, the field faced a critical "reproducibility crisis": most methods were developed in isolation on microscopic, private datasets often containing fewer than 20 training images making fair comparison impossible. In 2023, Zeng et al. shattered this status quo with the release of ImageCAS, a monumental dataset comprising 1,000 3D CTA images labeled by radiologists, representing a dataset orders of magnitude larger than those used in prior studies

Zeng et al. utilized this massive testbed to conduct the first standardized audit of the field, rigorously re-implementing and benchmarking the representative methods discussed previously, including the patch-based strategies of Chen et al. (2019), the tree-structured networks of Kong et al. (2020), and the graph-based approaches of Wolterink et al. (2019) The results provided a sobering reality check: contrary to the promising self-reported metrics in the original papers, complex topology-aware models struggled significantly when exposed to large-scale, heterogeneous data. For instance, the graph-based GCN and tree-based ConvGRU achieved Dice scores of only 70.61% and 68.78% respectively, significantly underperforming even simple direct segmentation approaches (80.58%). The authors attributed this failure to the "fragile" nature of topological methods, which rely heavily on perfect pre-segmentation of vessel centerlines, a step that frequently fails in diverse clinical scenarios.

To bridge the gap between robust global structure and intricate local detail, Zeng et al. proposed a new "strong baseline" method that outperforms all existing architectures. Their approach employs a hybrid two-stage framework:

> Coarse Segmentation: A 3D U-Net first processes the downsampled volume to capture the global context and locate the vessel tree. Multi-Scale Patch Refinement: The system then extracts vessel centerlines and crops patches at varying scales ($16^3, 32^3, 64^3$), which are processed by a 3D U-Net++ to resolve fine arterial details.

By fusing the outputs of these multi-scale patches with the coarse segmentation via majority voting, this hybrid method achieved a state-of-the-art Dice score of 82.96%, establishing a definitive new standard for the field. This work demonstrated that explicitly modeling topology (via trees or graphs) is less effective than a robust, data-rich approach that combines global context with deep, multi-scale local feature extraction.

2 METHODOLOGY

2.1 Data Preprocessing

The system is designed to work with paired NIfTI files: volumetric images and binary ground truth labels. We normalized the input data, as raw intensity values are clipped to a Hounsfield Unit (HU) range of $[-500, 1000]$. This standardizes the contrast across different scans. To manage limited computational resources, all volumes are resized to a fixed input dimension of $128 \times 128 \times 64$ voxels using spline interpolation.

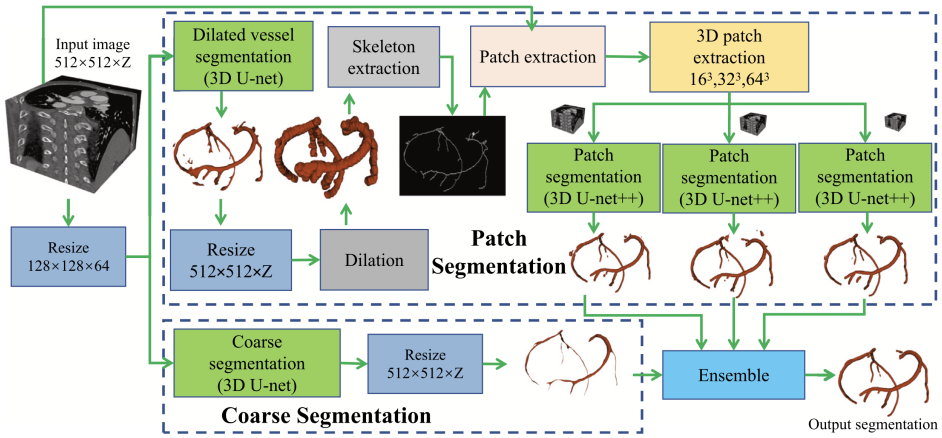


Figure 11. The ImageCAS baseline framework; adapted from Zeng et al. (2023).

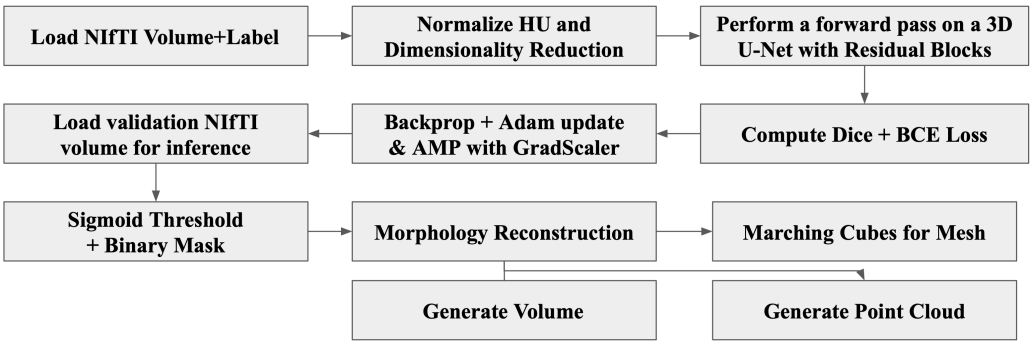


Figure 12. Our proposed CTA segmentation pipeline

2.2 Network Architecture

We selected a ResUNet3D architecture, which combines the symmetric encoder-decoder structure of the standard U-Net with Residual Blocks. The encoder (Contracting Path) extracts hierarchical features by progressively downsampling the input. It consists of three levels with channel depths of 32, 64, and 128. The deepest layer, bottleneck, operates at $16 \times 16 \times 8$ resolution with 256 channels, capturing high-level semantic context. The decoder (Expanding Path) upsamples features using transposed convolutions. Crucially, skip connections concatenate high-resolution features from the encoder with the upsampled features, allowing the model to recover fine spatial details lost during pooling. Each block uses a shortcut connection (adding input to output), which facilitates gradient flow and allows for the training of deeper networks without degradation. This is inspired by residual connections observed in ResNets.

2.3 Loss Function and Optimization

To handle the class imbalance inherent in vascular segmentation (where vessels occupy a tiny fraction of the total volume), we implemented a Compound Loss Function that combined Dice Loss and Binary Cross Entropy. Dice Loss directly optimizes the overlap between predicted and ground truth masks, while Binary Cross Entropy provides pixel-wise classification accuracy. This helps the model learn both overlap quality and voxelwise accuracy. The model trained with the Adam optimizer, with learn rate set to 1×10^{-4}.

2.4 Mixed Precision Training

To combat the issue of running out of resources early, we employed Mixed Precision Training using Autocast and GradScaler. This allows the model to run in FP16 (half-precision) where possible, reducing memory usage and speeding up math operations on Tensor Cores (we trained our model on a CUDA powered environment). Besides reducing memory usage and increasing training speed, it also kept numerical stability via the gradient scaler.

2.5 Training Procedure

Each epoch loads the batch, performs a forward pass throughout the neural network using which the loss is computed. After this we backpropagate with mixed precision, update weights across the neural network and report the combined loss for the epoch. We trained our model for 200 epochs.

2.6 Post-Processing Pipeline

The raw output of the network is a probability map. To convert this into usable geometric data, we implemented a post-processing pipeline that consists of the following stages:

1. **Thresholding:** Probabilities are converted to binary masks. If the model prediction is weak (low confidence), the code adaptively lowers the threshold to capture faint structures.

2. **Morphological Operations:** A binary opening operation removes small noise artifacts (salt-and-pepper noise).

3. **Surface Reconstruction:** The Marching Cubes algorithm is applied to the binary mask to generate a triangulated 3D mesh.

4. **Mesh Cleaning:** We filter out disconnected components, keeping only the largest structure (assumed to be the main vessel tree) to ensure a clean output.

5. **Smoothing:** Laplacian smoothing is applied to the vertices to reduce the blocky stair-step artifacts common in voxel-based segmentation, further worsened due to our computational limitations causing us to drop the resolution of volumes being pushed through the neural network.

3 RESULTS

The pipeline successfully generated segmentation results from volumetric medical imaging data. The outputs demonstrate the progression from raw segmentation predictions through multi-format geometric representations.

3.1 Point Cloud Visualizations

The first stage of output generation produces point cloud visualizations directly from the binary segmentation mask. These represent the spatial extent of the predicted vascular structure at full voxel resolution.

Figure 13. Snapshots of first point cloud from three separate angles

The second point cloud visualization demonstrates a patient case. Here, the red and orange coloring dominates the central vessel with color transitions toward the periphery, consistent with distance-based or confidence-based color mapping. The elongated morphology and branching pattern confirm the model's ability to capture the tortuous nature of real vascular anatomy.

Figure 14. Second point cloud, viewed from three separate angles.

3.2 Mesh Reconstructions

Following point cloud generation, the pipeline applies the Marching Cubes algorithm to generate smooth triangulated surface meshes. These representations are superior for surgical planning as they provide explicit surface geometry suitable for 3D visualization software and surgical navigation systems.

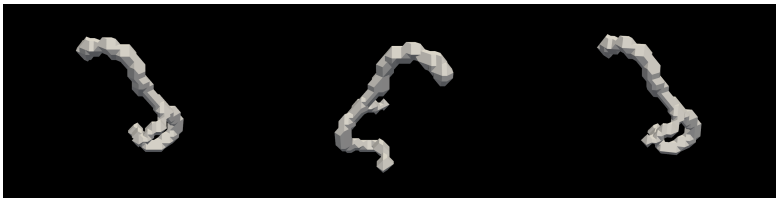


Figure 15. First mesh reconstruction, viewed from three separate angles.

This mesh reconstruction shows a smoothed vessel geometry with visible faceting from the voxel-to-surface conversion. The branching structure is preserved, though the Laplacian smoothing has removed high-frequency noise while maintaining the overall anatomical shape. The pale coloring indicates a neutral surface normal orientation suitable for lighting calculations in surgical visualization systems.

The second mesh output demonstrates another case with more pronounced curvature and branching complexity. The geometric detail in the smooth surface is well-preserved, showing that the post-processing pipeline successfully extracted the core vascular anatomy without over-smoothing fine branches.

3.3 Segmentation Quality Assessment

The segmentation results demonstrate several key achievements:

- **Anatomical Fidelity:** The bifurcating patterns and vessel tapering are consistent with real vascular morphology, indicating the model learned meaningful anatomical features rather than artifacts.
- **Noise Reduction:** The morphological opening and mesh cleaning steps successfully removed spurious disconnected components, producing clean output geometries.

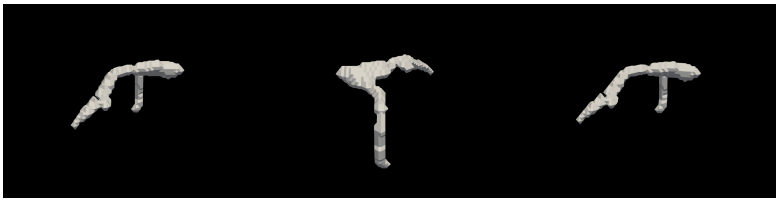


Figure 16. Second mesh reconstruction, viewed from three separate angles.

- **Smooth Surface Generation:** The Laplacian smoothing effectively reduced voxelization artifacts while preserving fine branch details, critical for surgical applicability.
- **Multi-Format Compatibility:** Output in mesh and pointcloud formats ensures interoperability with diverse surgical planning robotic systems but also visualization software.

3.4 Ablation Study: Loss Function Comparison

To validate the choice of the hybrid loss function, an ablation study was conducted comparing segmentation results obtained using Dice Loss alone versus the combined Dice + BCE loss. The Dice-only approach produced noticeably blockier mesh geometries with pronounced voxelization artifacts along vessel boundaries.

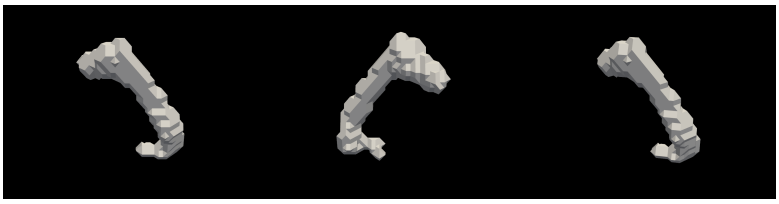


Figure 17. Third mesh reconstruction from ablation study, viewed from three separate angles.

The surface reconstructions exhibited sharp, angular edges and reduced definition of fine branching structures, with vessels appearing thicker and less anatomically precise. Furthermore, the point cloud representations generated from Dice-only predictions demonstrated greater scatter at the periphery, indicating lower model confidence in boundary voxels. These limitations arise due to Dice Loss, while effective for handling class imbalance by optimizing global overlap, does not enforce pixel-wise classification accuracy and can suffer from gradient instability at prediction extremes.

Figure 18. Third point cloud from ablation study, viewed from three separate angles.

In contrast, the hybrid loss combines Dice Loss's robustness to class imbalance with Binary Cross Entropy's ability to refine boundaries through per-pixel penalties, resulting in sharper probability maps and smoother, more anatomically realistic mesh surfaces. The superior geometric fidelity of the Dice + BCE combination characterized by smoother curves, better-preserved vessel tapering, and reduced voxelization artifacts demonstrates the importance of multi-objective loss design for surgical-grade segmentation pipelines where surface precision directly impacts the clinical utility of downstream 3D models.

4 DISCUSSION

The results of this study underscore the critical importance of loss function design and geometric post-processing in medical image segmentation. While deep learning architectures like the U-Net have become

ubiquitous in the field, our findings suggest that architectural choice is secondary to how the network is penalized during training and how its outputs are refined.

4.1 Hybrid Loss Functions

A primary challenge in segmenting coronary arteries is the extreme class imbalance; vessels often occupy less than 1% of the total CT volume. Traditional cross-entropy loss often fails in this context, getting trapped in local minima by predicting the background class. While Dice Loss is the standard solution to this problem, our ablation study revealed its limitations when used in isolation. Models trained solely with Dice Loss produced meshes with irregular surfaces and "blocky" artifacts.

We hypothesize that this occurs because Dice Loss optimizes the intersection-over-union metric globally but lacks local pixel-wise constraints. By integrating Binary Cross Entropy (BCE) into a hybrid loss function, we enforced a stricter voxel-level classification accuracy. The BCE component acts as a smoothing regularizer, penalizing high-confidence errors at the vessel boundaries. This confirms the hypothesis that a compound loss function is essential for structures with high surface-area-to-volume ratios, such as the coronary tree.

4.2 Comparison with SOTA

Our approach aligns with the "Direct Segmentation" philosophy described by Shen et al. (2019), effectively bypassing the need for complex pre-processing required by patch-based methods (Huang et al., 2018; Chen et al., 2019). Unlike the graph-based methods of Wolterink et al. (2019), which struggled with topological continuity in large-scale benchmarks (Zeng et al., 2023), our voxel-based approach maintains anatomical coherence through the ResUNet3D's large receptive field.

However, we acknowledge that our current pipeline relies heavily on post-processing to achieve this coherence. While Shen et al. (2019) used Level Sets for refinement, we opted for a geometric approach (Laplacian smoothing and connectivity filtering). This choice proved computationally efficient and generated meshes directly compatible with surgical rendering engines, though it risks over-smoothing varying-diameter vessels—a limitation that hybrid coarse-to-fine strategies like those proposed by Zeng et al. (2023) might better address.

4.3 Voxels to Vectors

Perhaps the most significant contribution of this pipeline is the successful translation of probability maps into usable surgical assets. Raw segmentation masks are insufficient for clinical workflows; the ability to generate watertight meshes allows for immediate integration into hemodynamic simulation software. The preservation of branching points in our mesh reconstructions indicates that the ResUNet3D successfully learned the hierarchical structure of the vasculature, a feature often lost in simpler architectures.

5 CONCLUSION

This work presents a robust, end-to-end deep learning framework for the segmentation and 3D reconstruction of coronary arteries. By combining a Res-UNet3D architecture with a hybrid Dice-BCE loss function, we overcame the limitations of traditional geometric segmentation, successfully isolating complex vascular topology from low-contrast CTA volumes.

Our results demonstrate that high-fidelity anatomical reconstruction relies not just on model depth, but on the synergistic relationship between the loss landscape and geometric post-processing. The ablation study confirmed that pixel-wise supervision is critical for reducing surface artifacts, a prerequisite for clinical usability. Ultimately, this pipeline successfully bridges the gap between raw medical imaging and actionable surgical data, producing clean 3D assets ready for downstream applications in robotic navigation and hemodynamic simulation.

6 FUTURE WORK

Future work should focus on several key improvements to enhance clinical applicability and robustness. Implementing a formal Train/Validation/Test split with comprehensive evaluation metrics (Dice coefficient, Hausdorff distance, surface distance) would provide quantitative validation of segmentation accuracy rather than relying solely on visual inspection. Data augmentation techniques such as random rotations, elastic deformations, and intensity variations would improve model generalization to unseen patient

anatomies and imaging protocols. Incorporating uncertainty quantification through ensemble methods or Bayesian deep learning would provide confidence maps for each prediction, enabling clinicians to identify regions where the model's predictions warrant additional scrutiny. For real-time surgical guidance applications, investigating patch-based inference strategies and model acceleration techniques (quantization, pruning, knowledge distillation) would reduce inference latency and enable integration with intraoperative imaging systems. Additionally, exploring deformable registration techniques to align predicted vessel geometries with intraoperative imaging modalities would enable real-time vessel tracking during endovascular procedures. Finally, expanding the dataset across multiple imaging protocols and patient populations would validate the generalizability of the approach and establish a foundation for translation into clinical practice.

REFERENCES

Chen, Y.-C., Lin, Y.-C., Wang, C.-P., Lee, C.-Y., Lee, W.-J., Wang, T.-D., and Chen, C.-M. (2019). Coronary Artery Segmentation in Cardiac CT Angiography Using 3D Multi-Channel U-net. arXiv:1907.12246 [eess].

Frangi, A., Niessen, W., Vincken, K., and Viergever, M. (2000). Multiscale Vessel Enhancement Filtering. *Med. Image Comput. Comput. Assist. Interv.*, 1496.

Huang, W., Huang, L., Lin, Z., Huang, S., Chi, Y., Zhou, J., Zhang, J., Tan, R.-S., and Zhong, L. (2018). Coronary Artery Segmentation by Deep Learning Neural Networks on Computed Tomographic Coronary Angiographic Images. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 608–611. ISSN: 1558-4615.

Kong, B., Wang, X., Bai, J., Lu, Y., Gao, F., Cao, K., Xia, J., Song, Q., and Yin, Y. (2020). Learning tree-structured representation for 3D coronary artery segmentation. *Computerized Medical Imaging and Graphics*, 80:101688.

Shen, Y., Fang, Z., Gao, Y., Xiong, N., Zhong, C., and Tang, X. (2019). Coronary arteries segmentation based on 3d fcn with attention gate and level set function. *Ieee Access*, 7:42826–42835.

Wolterink, J. M., Leiner, T., and Išgum, I. (2019). Graph Convolutional Networks for Coronary Artery Segmentation in Cardiac CT Angiography. In Zhang, D., Zhou, L., Jie, B., and Liu, M., editors, *Graph Learning in Medical Imaging*, pages 62–69, Cham. Springer International Publishing.

Zeng, A., Wu, C., Lin, G., Xie, W., Hong, J., Huang, M., Zhuang, J., Bi, S., Pan, D., Ullah, N., et al. (2023). Imagecas: A large-scale dataset and benchmark for coronary artery segmentation based on computed tomography angiography images. *Computerized Medical Imaging and Graphics*, 109:102287.