Reducing Expert Annotation Effort for CTA Vessel Delineation with Synthetic Pretraining

Jacopo Bracci[1,2][0009−0006−9760−7619], Alexander Katzmann[2][0000−0003−1958−1549], Tri-Thien Nguyen[1][0009−0007−0974−3299], Michael Sühling[2], and Andreas Maier[1][0009−0006−7583−9144]

[1] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen–Nürnberg, Erlangen, Germany
[2] Computed Tomography, Siemens Healthineers AG, Forchheim, Germany
jacopo.bracci@fau.de

Abstract. Computed tomography angiography is widely used to visualize vascular anatomy. Accurately distinguishing the inner lumen from the outer vessel wall is essential for many clinical and research tasks, yet in routine practice this delineation is typically performed manually by radiologists. Recent deep learning approaches aim to automate this process, but they generally rely on large, and often private, manually annotated datasets, incurring substantial labeling cost and observer-dependent variability. To address these limitations, we propose a pipeline that is pre-trained on a large set of synthetically generated contour annotations (requiring no manual labeling) and then fine-tuned using a small set of manually corrected data. We evaluate the method with a multi-annotator study to benchmark performance against human variability. After fine-tuning with n=400 cases, our method achieves a mean MAE of 0.493 mm (vs. Reader 1) and 0.643 mm (vs. Reader 2), closely matching inter-annotator variability (0.500 mm). We also report labeling cost, where n=400 corresponds to ≈6 hours of correction time in our setting. We also analyze performance as a function of variable n manual supervision. In a blinded preference test, an independent radiologist selected the model prediction over the human annotation in 56.8% (vs. Reader 1) and 67.6% (vs. Reader 2) of comparisons. Overall, these results indicate that the proposed approach substantially reduces annotation effort while approaching observed inter-reader variability, enabling scalable and reproducible lumen and vessel-wall delineation for downstream analysis with limited manual supervision.

Keywords: Computed Tomography Angiography · Vessel Delineation · Label-efficient Learning · Synthetic Pretraining.

1 Introduction

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, accounting for roughly 32 % of all deaths (≈ 19.8 million annually) [19]. Quantitative vessel assessment (e.g., lumen diameter, wall thickness, plaque burden)

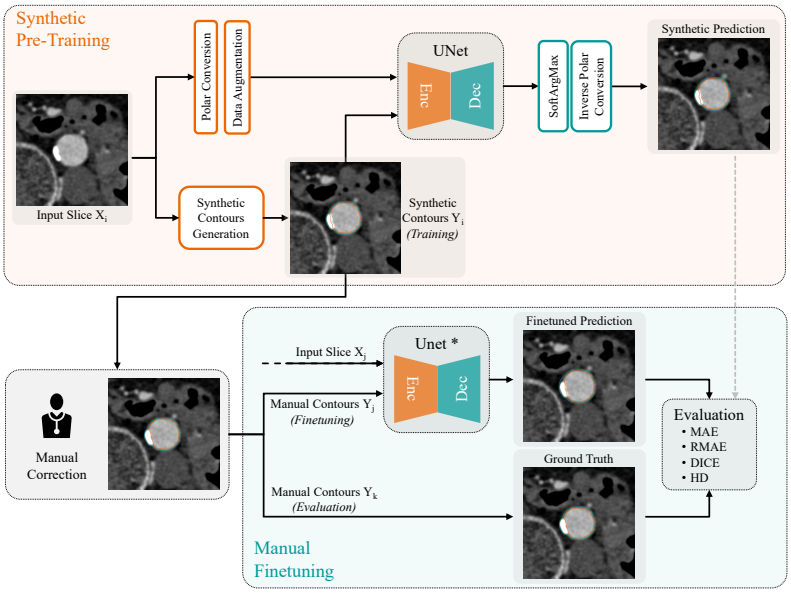

Fig. 1. Overview of the proposed two-stage pipeline. A U-Net is first pre-trained on automatically generated (synthetic) contour annotations and subsequently fine-tuned (U-Net*) on a small set of manually corrected contours.

is central to diagnosis and therapy planning, and relies on accurate delineation of both the inner lumen and the outer vessel wall. Computed tomography angiography (CTA) and black-blood MRI are widely used for this purpose due to their availability and favorable contrast.

In current clinical workflows, this delineation is largely manual. Manual contouring is time-consuming and difficult to scale, especially over long vessel segments or multiple vessels. In our annotation setup, correcting a cross-sectional slice took 53-58 s on average across readers (Table 2), implying that annotating thousands of slices would require tens of hours of time. Moreover, substantial inter-reader variability has been reported [16], reducing reproducibility and complicating downstream analysis. These challenges motivate automated methods that can provide consistent delineations while reducing workload.

Unsupervised pipelines typically combine hand-crafted vessel heuristics, such as vesselness filtering, region growing, morphological operations, and contour evolution, to estimate lumen and wall boundaries [8, 5]. While they minimize labeling, their performance often depends on hand-tuned assumptions about vessel appearance and acquisition protocols, limiting robustness.

Fully supervised deep learning methods have achieved strong performance across modalities. A common strategy, particularly in black-blood MRI, is to sample cross-sectional slices along a vessel centerline and perform 2D segmentation, where lumen and wall are most separable [1, 4, 20]. Learning-based strate-

Table 1. Training set sizes reported in related work. For 2D methods we list the number of annotated slices/cross-sections; for 3D methods we report the number of training volumes when slice counts are not reported.

Study	Mod.	Dataset	Train slices	Train volumes
Alblas et al. [1]	MRI	Challenge, CARE-II subset	2655	26
Rahlfs et al. [12]	MRI	Custom [16]	2654	202
Chen et al. [4]	MRI	CARE-II [23], KOWA [17, 21]	26008	925
Xu et al. [20]	MRI	Custom	9073	115
Serrano et al. [14]	CTA	Custom [15]	—	22
Liu et al. [9]	CTA	MR CLEAN subset	—	10

gies have also been explored in CTA in both 3D and 2D formulations [13, 9, 14]. However, these methods are data-hungry: training typically requires large, carefully annotated datasets, which are costly to produce and often private, hindering reuse and benchmarking. Table 1 summarizes the scale of annotated training data reported in representative vessel-wall and CTA segmentation studies, motivating the need for label-efficient alternatives.

In this work, we reduce the reliance on expert annotation by leveraging synthetically generated contour annotations (i.e., produced lumen and wall contours, not synthetic images), and then using limited manual correction to align predictions with reader delineations. Specifically, we propose a two-stage pipeline (Fig. 1) in which a model is first pre-trained on a large corpus of synthetic contours generated without human intervention, and then fine-tuned on a small set of manually corrected contours.

Our main contributions are: (i) We propose a label-efficient two-stage training strategy for CTA lumen and vessel-wall delineation, combining large-scale pre-training on generated contour annotations with low-data fine-tuning on manually corrected contours. (ii) We benchmark performance against human variability via a multi-annotator study, enabling direct comparison of model–reader and inter-reader disagreement. (iii) We quantify performance as a function of the number of manual fine-tuning samples, demonstrating competitive accuracy under limited supervision.

2 Material and Methods

2.1 Dataset

A dataset of 3D CT scans was acquired at a single site using a NAEOTOM Alpha photon-counting scanner (Siemens Healthineers AG, Forchheim, Germany). Using manually annotated vessel centerlines, we sampled axial cross-sectional slices along each centerline at intervals of 0.5 mm. Each slice spans $60 \times 60\,\mathrm{mm}^2$ and is represented at a resolution of $128 \times 128\,\mathrm{px}^2$, corresponding to an isotropic spacing of approximately 0.47 mm. The dataset covers multiple vascular territories across the body, including head/neck, thoracoabdominal, and peripheral lower-limb vessels (Table 2).

Table 2. Distribution of cross-sectional slices across splits, stratified by vessel type. Manual slices are split into development and evaluation; annotation time reports per-slice correction time in seconds (mean ± std) for each reader. Left/right (and similar variants) are aggregated within each vessel category.

Vessel	Synthetic	Manual		Annotation Time [s]	
		Develop.	Eval.	Reader 1	Reader 2
Runoff	14729	124	342	61.4 ± 45.1	55.3 ± 45.2
Aorta	11442	94	287	64.4 ± 51.8	65.1 ± 65.7
Carotis	8197	88	243	55.2 ± 39.4	49.2 ± 72.4
Tibialis	3768	27	88	47.1 ± 31.1	35.8 ± 24.6
Vertebralis	3264	34	109	49.3 ± 46.4	40.1 ± 38.4
Femoralis	2293	18	84	62.6 ± 71.2	45.2 ± 28.6
Iliaca	1697	39	103	59.8 ± 39.8	56.2 ± 31.0
Fibularis	1274	13	24	44.0 ± 26.1	49.1 ± 70.1
Mesenterica	996	13	42	49.9 ± 42.2	62.4 ± 81.0
Poplitea	703	7	40	45.5 ± 36.4	41.1 ± 21.5
Renalis	663	15	50	50.6 ± 42.6	54.8 ± 41.1
Basilaris	20	2	3	39.0 ± 11.5	29.0 ± 11.5
Total	**49046**	**474**	**1415**	**57.9 ± 46.3**	**52.9 ± 54.7**

We randomly drew two non-overlapping subsets: a *synthetic* split and a *manual* split. The manual split was further subdivided into *development* and *evaluation*. The development split was used only for hyperparameter selection and method development, while the evaluation split was reserved exclusively for final evaluation. There is no slice overlap between splits. We initially sampled 50,000 synthetic and 2,000 manual slices; after filtering failed automatic proposals and a small number of quality-flagged cases, the final counts are 49,046 synthetic slices and 1,889 manual slices (Table 2).

Inner-lumen and outer-wall contours were initially generated using a fully automatic approach [2]. For the *synthetic* split, these contours were used as-is, without human intervention. For the *manual* split, corrections were performed independently by two medical students (denoted *Reader 1* and *Reader 2* throughout the paper) with prior experience in vascular image annotation. Both used identical initialization from the automatic contour proposals.

The annotation tool logged per-slice correction time, which averaged 57.9±46.3 s (Reader 1) and 52.9±54.7 s (Reader 2) over all manually corrected slices (Table 2). This provides a concrete estimate of labeling effort and enables reporting label efficiency in terms of annotation time.

2.2 Pipeline

Figure 1 summarizes the proposed two-stage training setup. We first pre-train on the large *synthetic* split using fully automatic contour annotations, and then fine-tune on the small (≈1%) *manual-development* split using human-corrected contours. All results are reported on the held-out *manual-evaluation* split.

Task formulation in polar coordinates. Following prior work [4, 1], we convert each Cartesian cross-section into a polar image centered at the vessel centerline.

In polar coordinates, the inner-lumen and outer-wall boundaries are represented as radius functions of angle, $r_{\text{in}}(\theta)$ and $r_{\text{out}}(\theta)$. We discretize the angular dimension into 128 steps ($\approx 2.8^\circ$ per step) and cast learning as regression: given a polar image, the network predicts the two radius profiles as continuous arrays. For visualization, predicted profiles are mapped back to Cartesian contours, while all evaluation is performed in polar coordinates.

Pre-training on synthetic annotations. We use a plain 2D U-Net backbone [7]. Radii are estimated with a differentiable softargmax-style operator along the radial dimension [11, 10]. Inputs are normalized to zero mean and unit variance (computed on the training split). We apply random rotations, additive Gaussian noise, and HU intensity shifts for augmentation. Since automatic contours may be noisy, we optimize the robust T-loss [6]. We train with Adam (lr = 10^{-4}, weight decay = 0.1) and reduce the learning rate on plateau (monitoring validation loss; patience = 3, factor = 0.1). Pre-training runs for up to 300 epochs with batch size 32.

Fine-tuning on manual annotations. We fine-tune on the manually corrected contours and randomly sample the target contour from the two readers during training. Restricting fine-tuning to a single reader did not meaningfully change performance. To mitigate overfitting under limited supervision, we freeze the encoder (as early features tend to transfer better [22]) and progressively freeze decoder blocks over training, inspired by FreezeOut [3]. Specifically, we freeze one additional decoder block every five epochs while keeping the final layer trainable. Fine-tuning runs for up to 50 epochs with batch size 4 using Adam (lr = 10^{-4}, weight decay = 10^{-4}) with the same reducing schedule.

3 Experiment Setup

We evaluate the proposed two-stage pipeline with three goals: (i) quantify the benefit of fine-tuning on n manually corrected contours over synthetic-only training, (ii) compare model-reader disagreement to inter-reader variability, and (iii) assess whether the fine-tuned model is preferred over humans in a blinded study.

All models are pre-trained on the *synthetic* split and evaluated only on the held-out *manual-evaluation* split. Manual supervision is introduced by fine-tuning on n corrected cross-sections from the *manual-development* split (n=0: synthetic-only; n=400: full fine-tuning). For each n, we fine-tune on 100 different random subsets and report aggregate performance. Inter-reader agreement is computed on the same evaluation set using identical metrics. For the blinded preference test, a board-certified radiologist (> 5 years experience) reviewed 200 randomly sampled cases, each shown as two randomized contour overlays (A/B) on the original image, and chose *A better*, *B better*, or *Tie*.

Metrics. Each cross-section is represented in polar coordinates by two 1D radius profiles (inner lumen and outer wall, in mm) sampled at 128 fixed angles.

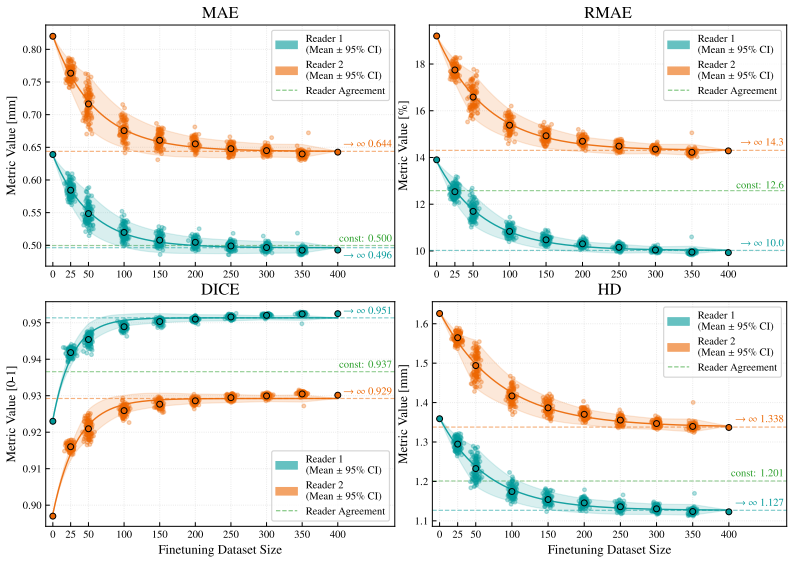

Fig. 2. Fine-tuning data efficiency. Performance on the held-out manual-evaluation split as a function of the number of manually corrected fine-tuning samples n (n=0: synthetic-only). Dots denote individual runs with different random subsets; large markers show the mean and shaded bands indicate 95% confidence intervals. Curves are exponential fits used to estimate asymptotic performance. Dashed green lines report inter-reader agreement on the same set, providing a human-variability reference.

Metrics are computed per slice (inner/outer averaged) and then averaged over the evaluation set.

We report: **MAE (mm)**: Mean Absolute Error between the predicted and reference radius profiles. **RMAE (%)**: Relative MAE, computed as the MAE normalized at each angle by the reference radius and then averaged; angles with zero reference radius contribute 0 (lower is better). **DICE (0–1)**: 2×overlap/union, where overlap and union are computed from the areas under the two radius profiles (areas approximated via trapezoidal integration; higher is better). **HD (mm)**: Undirected Hausdorff distance, defined as the maximum of the two directed Hausdorff distances [18] between the radius profiles (lower is better).

4 Results and Discussion

Fig. 2 summarizes how performance changes as the number of manually corrected fine-tuning samples n increases. Across all metrics, fine-tuning yields consistent gains over synthetic-only training, with improvements that flatten as n grows. The small spread across random subsets indicates that performance is not strongly driven by the specific slices selected for fine-tuning.

Table 3. Synthetic-only (n=0) vs. fully fine-tuned (n=400) performance on the manual-evaluation split, measured against each reader. Metrics are computed per slice (inner and outer averaged in polar coordinates) and summarized over the evaluation set. **Bold** indicates the better value for each metric within a reader.

Metric	Model	Reader 1					Reader 2				
		Mean	Std ↓	Median	Min	Max	Mean	Std ↓	Median	Min	Max
MAE	Synthetic	0.639	0.475	0.492	0.083	4.842	0.820	0.382	0.731	0.162	2.413
[mm] ↓	Finetuned	**0.493**	**0.424**	**0.324**	**0.082**	**4.724**	**0.643**	**0.344**	**0.557**	**0.108**	**2.207**
RMAE	Synthetic	13.90	10.30	10.10	3.70	65.30	19.20	12.40	15.30	3.60	**98.20**
[%] ↓	Finetuned	**9.90**	**6.00**	**8.10**	**2.70**	**52.80**	**14.30**	**9.20**	**12.10**	**3.30**	102.40
DICE	Synthetic	0.923	0.069	0.948	0.460	0.981	0.897	0.074	0.919	0.422	0.980
[0–1] ↑	Finetuned	**0.952**	**0.026**	**0.960**	**0.791**	**0.986**	**0.930**	**0.046**	**0.941**	**0.480**	**0.983**
HD	Synthetic	1.359	1.067	1.047	0.196	14.188	1.626	0.919	1.383	0.314	13.764
[mm] ↓	Finetuned	**1.123**	**0.995**	**0.779**	**0.218**	**10.943**	**1.337**	**0.814**	**1.147**	**0.263**	**10.440**

Table 4. Blind evaluation results. Each row corresponds to one head-to-head comparison. Numbers indicate how often the expert selected the method named in the column as better than its opponent in that row; the last column reports ties. **Bold** indicates the higher (preferred) percentage within each comparison.

	Reader 1	**Reader 2**	**Ours**	**Tie**
Reader 1 vs. Reader 2	**62.7%**	30.5%	—	6.8%
Ours vs. Reader 1	37.8%	—	**56.8%**	5.4%
Ours vs. Reader 2	—	30.9%	**67.6%**	1.5%

Table 3 compares the synthetic-only model (n=0) to the fully fine-tuned model (n=400). Fine-tuning improves all metrics for both readers, i.e. reducing mean MAE from 0.639→0.493 mm (Reader 1) and 0.820→0.643 mm (Reader 2), while also lowering variability (smaller standard deviations), indicating fewer large-error cases.

Performance approaches inter-reader variability with limited supervision. The green dashed lines in Fig. 2 report inter-reader agreement on the evaluation set (MAE = 0.500 mm, RMAE = 12.6%, DICE = 0.937, HD = 1.201 mm). As n increases, model–Reader 1 disagreement converges to this level (e.g., MAE → 0.496 mm and DICE → 0.951), suggesting that remaining differences are comparable to typical reader-to-reader variability. In contrast, model–Reader 2 metrics plateau at a higher error level (e.g., MAE → 0.644 mm, RMAE → 14.3%). Together, these trends support the intended role of fine-tuning: synthetic pretraining captures broadly valid vessel geometry, while a small amount of manual correction calibrates the prediction to reader delineations.

Blind preference test favors the fine-tuned model. Table 4 summarizes the blinded preference study. Overall, the radiologist often preferred the fine-tuned model over either human annotation, while also expressing a preference for Reader 1 over Reader 2. In the accompanying qualitative feedback, the radiologist emphasized that the inner lumen contour was weighted more heavily during judging

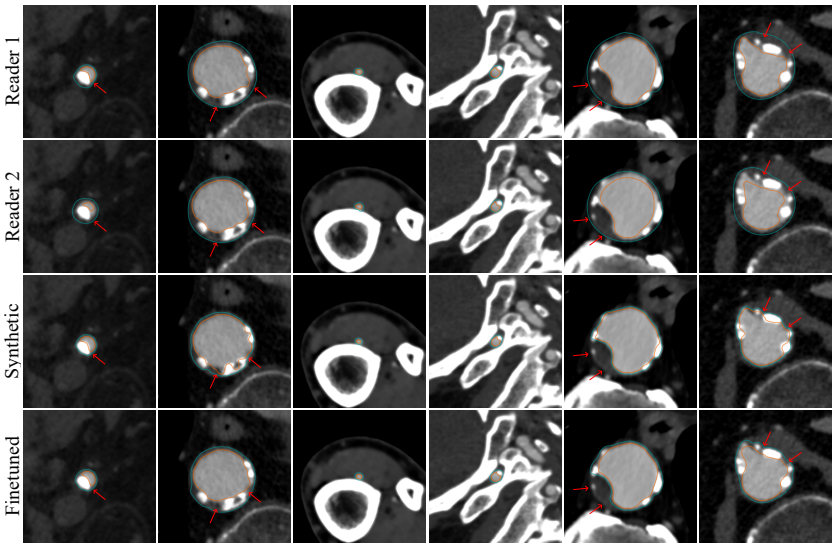

Fig. 3. Qualitative examples on the manual-evaluation split. Rows (top to bottom) show Reader 1, Reader 2, synthetic-only (n=0), and fine-tuned (n=400) contours. Red arrows highlight regions with the most visible differences.

because it is more clinically relevant, whereas the outer wall is not always clearly visualized on CTA. When contours were otherwise comparable, smoother delineations were typically preferred, as they better match angiographic experience and appear more natural. Major differences were rare: the largest outer-wall discrepancies occurred in two cases with mural thrombus (shown in Fig. 3), where the estimated vessel boundary differed substantially. Additional disagreements were noted near branch points and under beam-hardening artifacts. Overall, however, the radiologist judged the observed differences to be small and not clinically significant. Fig. 3 shows qualitative examples showing how fine-tuning generally improves contour consistency and reduces local irregularities.

Limitations. This study focuses specifically on label-efficient contour delineation and should be interpreted within that scope. First, we did not include a real-only baseline or direct comparisons with other label-efficient strategies; these remain important directions for future work once comparable CTA lumen/wall benchmarks become available. Second, because the quantitative references were corrected by trained student annotators rather than expert-consensus radiologists, our blinded radiologist study should be viewed as a complementary, preference-based validation rather than an absolute, gold-standard quantitative reference. Finally, while fine-tuning consistently outperformed synthetic-only training, robustness to automatic centerline errors and potential residual bias from automatically generated pretraining contours remain open questions.

5 Conclusion

We presented a two-stage deep learning pipeline for lumen and vessel-wall delineation in CTA that reduces the need for expert annotation. A U-Net is first pre-trained on a large set of fully automatic (synthetic) contour labels and then fine-tuned on a small set of manually corrected cross-sections, yielding accurate and stable predictions with minimal supervision ($\approx$ 6 hours of correction time for n=400 in our setting). Fine-tuning consistently improved over synthetic-only training across all metrics, reducing both average error and the frequency of large outliers. Model–reader disagreement approached observed inter-reader variability. Our data-efficiency analysis showed strong diminishing returns: most gains were achieved with relatively few fine-tuning cases, after which improvements saturated. In a blinded preference test, an independent radiologist preferred the fine-tuned model over both annotators in the majority of comparisons, suggesting improved consistency. Future work will evaluate generalization across scanners and institutions, extend to additional vessels and pathologies, and add uncertainty-aware outputs to support clinical editing and downstream analysis.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alblas, D., Brune, C., Wolterink, J.M.: Deep Learning-Based Carotid Artery Vessel Wall Segmentation in Black-Blood MRI Using Anatomical Priors (Dec 2021). https://doi.org/10.48550/arXiv.2112.01137
2. Bracci, J., Katzmann, A., Rist, L., Vorberg, L., Sühling, M., Maier, A.: Hybrid vessel wall segmentation for assisted annotation in ct angiography. In: Handels, H., Breininger, K., Deserno, T., Maier, A., Maier-Hein, K., Palm, C., Tolxdorff, T. (eds.) Bildverarbeitung für die Medizin 2026. pp. 285–291. Springer Fachmedien Wiesbaden, Wiesbaden (2026). https://doi.org/10.1007/978-3-658-51100-5_57
3. Brock, A., Lim, T., Ritchie, J.M., Weston, N.: Freezeout: Accelerate training by progressively freezing layers (2017). https://doi.org/10.48550/arXiv.1706.04983, https://arxiv.org/abs/1706.04983
4. Chen, L., Sun, J., Canton, G., Balu, N., Hippe, D.S., Zhao, X., Li, R., Hatsukami, T.S., Hwang, J.N., Yuan, C.: Automated Artery Localization and Vessel Wall Segmentation Using Tracklet Refinement and Polar Conversion. IEEE Access **8**, 217603–217614 (2020). https://doi.org/10.1109/ACCESS.2020.3040616, https://ieeexplore.ieee.org/document/9269957/
5. Ghanem, A.M., Hamimi, A.H., Matta, J.R., Carass, A., Elgarf, R.M., Gharib, A.M., Abd-Elmoniem, K.Z.: Automatic Coronary Wall and Atherosclerotic Plaque Segmentation from 3D Coronary CT Angiography. Scientific Reports **9**(1), 47 (Jan 2019). https://doi.org/10.1038/s41598-018-37168-4, https://www.nature.com/articles/s41598-018-37168-4
6. Gonzalez-Jimenez, A., Lionetti, S., Gottfrois, P., Gröger, F., Pouly, M., Navarini, A.A.: Robust T-Loss for Medical Image Segmentation. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention

– MICCAI 2023, vol. 14222, pp. 714–724. Springer Nature Switzerland (2023). https://doi.org/10.1007/978-3-031-43898-1_68

7. 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**(2), 203–211 (Feb 2021). https://doi.org/10.1038/s41592-020-01008-z, https://www.nature.com/articles/s41592-020-01008-z
8. Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G.: A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes. Med Image Anal **13**(6), 819–845 (Dec 2009). https://doi.org/10.1016/j.media.2009.07.011, https://linkinghub.elsevier.com/retrieve/pii/S136184150900067X
9. Liu, S., Su, R., Su, J., Van Zwam, W.H., Van Doormaal, P.J., Van Der Lugt, A., Niessen, W.J., Van Walsum, T.: Segmentation-assisted vessel centerline extraction from cerebral CT Angiography. Medical Physics **52**(7), e17855 (Jul 2025). https://doi.org/10.1002/mp.17855, https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.17855
10. Luvizon, D.C., Tabia, H., Picard, D.: Human pose regression by combining indirect part detection and contextual information. Computers & Graphics **85**, 15–22 (2019). https://doi.org/10.1016/j.cag.2019.09.002, https://www.sciencedirect.com/science/article/pii/S0097849319301475
11. Nibali, A., He, Z., Morgan, S., Prendergast, L.: Numerical coordinate regression with convolutional neural networks (2018). https://doi.org/10.48550/arXiv.1801.07372, https://arxiv.org/abs/1801.07372
12. Rahlfs, H., Hüllebrand, M., Schmitter, S., Strecker, C., Harloff, A., Hennemuth, A.: Learning carotid vessel wall segmentation in black-blood MRI using sparsely sampled cross-sections from 3D data. Journal of Medical Imaging **11**(04) (Jul 2024). https://doi.org/10.1117/1.jmi.11.4.044503
13. Schaap, M., Van Walsum, T., Neefjes, L., Metz, C., Capuano, E., De Bruijne, M., Niessen, W.: Robust Shape Regression for Supervised Vessel Segmentation and its Application to Coronary Segmentation in CTA. IEEE Transactions on Medical Imaging **30**(11), 1974–1986 (Nov 2011). https://doi.org/10.1109/TMI.2011.2160556, http://ieeexplore.ieee.org/document/5929564/
14. Serrano-Antón, B., Insúa Villa, M., Pendón-Minguillón, S., Paramés-Estévez, S., Otero-Cacho, A., López-Otero, D., Díaz-Fernández, B., Bastos-Fernández, M., González-Juanatey, J.R., P. Muñuzuri, A.: Unsupervised clustering based coronary artery segmentation. BioData Mining **18**(1), 21 (Mar 2025). https://doi.org/10.1186/s13040-025-00435-y, https://biodatamining.biomedcentral.com/articles/10.1186/s13040-025-00435-y
15. Serrano-Antón, B., Otero-Cacho, A., López-Otero, D., Díaz-Fernández, B., Bastos-Fernández, M., Pérez-Muñuzuri, V., González-Juanatey, J.R., Muñuzuri, A.P.: Coronary artery segmentation based on transfer learning and unet architecture on computed tomography coronary angiography images. IEEE Access **11**, 75484–75496 (2023). https://doi.org/10.1109/ACCESS.2023.3293090
16. Strecker, C., Krafft, A.J., Kaufhold, L., Hüllebrandt, M., Treppner, M., Ludwig, U., Köber, G., Hennemuth, A., Hennig, J., Harloff, A.: Carotid Geometry and Wall Shear Stress Independently Predict Increased Wall Thickness—A Longitudinal 3D MRI Study in High-Risk Patients. Frontiers in Cardiovascular Medicine **8**, 723860 (Oct 2021). https://doi.org/10.3389/fcvm.2021.723860

17. Sun, J., Balu, N., Hippe, D.S., Xue, Y., Dong, L., Zhao, X., Li, F., Xu, D., Hatsukami, T.S., Yuan, C.: Subclinical carotid atherosclerosis: Short-term natural history of lipid-rich necrotic core—a multicenter study with mr imaging. Radiology **268**(1), 61–68 (2013). https://doi.org/10.1148/radiol.13121702, https://doi.org/10.1148/radiol.13121702, pMID: 23513240
18. Taha, A.A., Hanbury, A.: An efficient algorithm for calculating the exact hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence **37**(11), 2153–2163 (2015). https://doi.org/10.1109/TPAMI.2015.2408351
19. World Health Organization: Cardiovascular diseases (CVDs). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (2025), accessed: October 10, 2025
20. Xu, W., Yang, X., Li, Y., Jiang, G., Jia, S., Gong, Z., Mao, Y., Zhang, S., Teng, Y., Zhu, J., He, Q., Wan, L., Liang, D., Li, Y., Hu, Z., Zheng, H., Liu, X., Zhang, N.: Deep Learning-Based Automated Detection of Arterial Vessel Wall and Plaque on Magnetic Resonance Vessel Wall Images. Frontiers in Neuroscience **16**, 888814 (Jun 2022). https://doi.org/10.3389/fnins.2022.888814, https://www.frontiersin.org/articles/10.3389/fnins.2022.888814/full
21. Yoneyama, T., Sun, J., Hippe, D.S., Balu, N., Xu, D., Kerwin, W.S., Hatsukami, T.S., Yuan, C.: In vivo semi-automatic segmentation of multicontrast cardiovascular magnetic resonance for prospective cohort studies on plaque tissue composition: Initial experience. The International Journal of Cardiovascular Imaging **32**(1), 73–81 (Jan 2016). https://doi.org/10.1007/s10554-015-0704-0, https://doi.org/10.1007/s10554-015-0704-0
22. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? Advances in neural information processing systems **27** (2014)
23. Zhao, X., Li, R., Hippe, D.S., Hatsukami, T.S., Yuan, C., CARE-II Investigators: Chinese Atherosclerosis Risk Evaluation (CARE II) study: A novel cross-sectional, multicentre study of the prevalence of high-risk atherosclerotic carotid plaque in Chinese patients with ischaemic cerebrovascular events—design and rationale. Stroke and Vascular Neurology **2**(1), 15–20 (Mar 2017). https://doi.org/10.1136/svn-2016-000053