FDRAS: Failure Diagnosis and Repair for Airway Segmentation

Francis Xiatian Zhang[1,2][0000−0003−0228−6359], Kevin Dhaliwal[1][0000−0002−3925−3174], and Mohsen Khadem[1,2][0000−0002−6873−273X]

[1] Baillie Gifford Pandemic Science Hub, Institute of Regeneration and Repair, University of Edinburgh, Edinburgh, United Kingdom
[2] School of Informatics, University of Edinburgh, Edinburgh, United Kingdom
{francis.zhang, kev.dhaliwal, mohsen.khadem}@ed.ac.uk

Abstract. Airway segmentation from chest CT is essential for safe navigation of the bronchoscopy and the targeting of distal lesions. However, segmentation models frequently exhibit structural failures under domain shift, including disconnected distal branches and boundary leakage. Existing post-hoc refinement methods either apply heuristic topology operations or directly predict corrected masks, implicitly coupling error localization and correction, and limiting controllability. We propose **FDRAS** (Failure Diagnosis and Repair for Airway Segmentation), a two-stage reliability framework that explicitly decouples failure localization from conservative correction. Given a CT volume and an initial segmentation, FDRAS performs hypothesis verification by predicting two complementary error maps: (1) a Signed Distance Function (SDF)-based geometric error map capturing boundary discrepancies, and (2) a tree-level connectivity error map localizing disconnected branches via skeleton analysis. These error maps condition a mask-only refinement network through cross-attention, enabling targeted updates while preventing topology drift. By externalizing structural inconsistencies and anchoring correction to predicted failure fields, FDRAS enables topology-aware repair rather than direct mask transformation. Under strict zero-shot evaluation across two external datasets and multiple backbones, FDRAS improves branch-level detection and connectivity by up to +4.8 percentage points (pp) compared to the strongest competing refinement, while maintaining competitive voxel-level accuracy, demonstrating its effectiveness as a post-hoc structural reliability layer for deployment.

Keywords: Airway segmentation · Segmentation refinement · Domain shift

1 Introduction

Accurate airway segmentation from chest CT is critical for bronchoscopy navigation and surgical planning [4], where structural completeness and connectivity directly affect downstream anatomical reasoning. Although recent architectures report strong benchmark performance and incorporate topology-aware

constraints [21, 12, 20, 23], topology consistency can still degrade under domain shift arising from scanner variability and population heterogeneity [6], and voxel-wise overlap metrics may fail to reveal these structural failures [13]. Therefore, reliable clinical deployment requires an additional layer that explicitly localizes structural violations and enables controlled correction.

Current efforts toward improving segmentation reliability generally fall into two categories: post-hoc quality diagnosis and mask refinement. Quality control methods estimate global scores or voxel-wise error maps [5, 1, 14], but they stop at detection and do not perform correction. Refinement approaches directly modify the predicted mask, either via endpoint-based geometric linking [19], universal topology-aware transformations [8], or CT-guided adversarial refinement [16]. Although these methods can improve overlap or topological metrics, detection and correction are either implicitly coupled within a single mask transformation [16, 8] or implemented as a learnable topology-focused detection module followed by heuristic, non-learnable repair operations [19].

This reveals a key limitation: diagnosis methods detect errors but do not correct them, while refinement methods modify the mask without explicitly separating failure localization from correction within a jointly learned framework. As a result, repair lacks controllable, diagnosis-conditioned guidance under domain shift. To address this gap, we propose FDRAS, a failure diagnosis and repair framework that first predicts where structural errors occur and then conditions refinement on these localized failure maps. Given a predicted mask and CT volume, FDRAS estimates (1) an SDF-based boundary discrepancy field and (2) a tree-level connectivity map derived from skeleton analysis. These maps guide a mask-only refinement network via cross-attention, enabling targeted correction while preserving reliable regions. By separating error detection from correction, FDRAS performs controlled, topology-aware repair under domain shift.

We train FDRAS on ATM [22] and evaluate robustness under strict zero-shot settings on two unseen external datasets, AirRC [9] and AeroPath [15]. Experimental results demonstrate consistent improvements in structural validity and branch-level detection metrics while maintaining competitive voxel-level accuracy. To further assess generalization beyond public datasets, we perform additional validation on an ex-vivo human lung with corresponding CT imaging. Without retraining or fine-tuning, FDRAS restores distal branch continuity and improves structural coherence under this in-vivo to ex-vivo distribution shift, providing further evidence of robustness in anatomically and acquisition-diverse settings. The code, ex-vivo dataset, and implementation and evaluation utilities are available at `https://github.com/SIRGLab/FDRAS`. To summarize, our contributions are:

1. We propose **FDRAS**, a backbone-agnostic framework that decouples structural failure diagnosis from segmentation refinement.
2. We introduce a structural error-map formulation consisting of an SDF-based boundary discrepancy field and a tree-level connectivity map.
3. We design a lightweight two-stage cross-attention architecture that performs error-conditioned, spatially targeted refinement.

2 Methodology

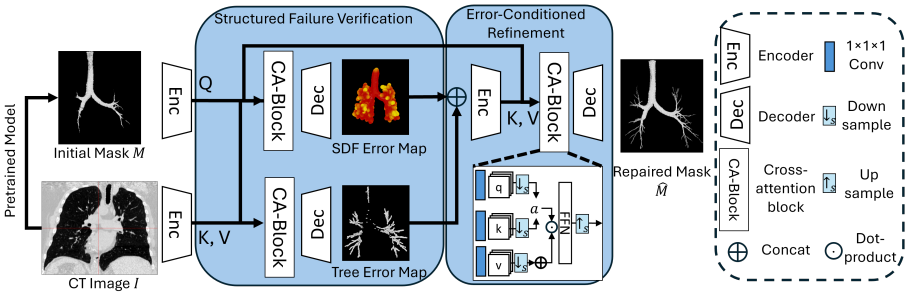

Fig. 1. Overview of FDRAS. The framework consists of (1) structural error-map prediction from the initial segmentation and CT volume, and (2) error-conditioned refinement guided by spatially reduced cross-attention.

As illustrated in Fig. 1, given an initial segmentation $M \in \mathbb{R}^{1\times D\times H\times W}$ and its corresponding CT volume $I \in \mathbb{R}^{1\times D\times H\times W}$, FDRAS operates in two explicitly decoupled stages: (1) structural failure verification and (2) error-conditioned conservative refinement. First, a diagnosis network predicts two structural failure fields from (M, I): a boundary-aware map E_{SDF} capturing geometric discrepancies, and a topology-aware map E_{tree} localizing connectivity violations. Next, a refinement network takes $(M, E_{\text{SDF}}, E_{\text{tree}})$ as input and produces a corrected segmentation $\hat{M}$. Cross-attention conditions the refinement on encoded mask features and the predicted error maps, focusing updates on localized structural defects while preserving reliable regions.

2.1 Structural Error Representation

Existing topology-repair methods [19, 8] mainly target connectivity violations. However, connectivity alone does not fully characterize structural failure: boundary leakage, over-segmentation, and geometric inconsistencies may persist even when global topology appears intact. Reliable structural diagnosis therefore requires modeling both geometric and topological discrepancies.

To address this gap, we explicitly model structural errors using two complementary voxel-wise maps that identify failure regions and guide targeted refinement. While prior works employ SDFs as segmentation representations [2], we instead construct an *SDF error map* that quantifies boundary discrepancy. This is paired with a *tree error map* that isolates root-connected branch violations. The formulation is detailed below.

SDF Error Map. The SDF error map E_{SDF} captures dense boundary discrepancies using continuous geometric representations. Let $\phi(\cdot)$ denote the signed

distance transform [10], which encodes the distance of each voxel to the nearest boundary with sign indicating inside or outside. We define

$$E_{\mathrm{SDF}} = \left|\phi(M) - \phi(M_{\mathrm{gt}})\right|, \tag{1}$$

which measures voxel-wise boundary deviation magnitude and provides a spatially dense characterization of geometric mismatch.

Tree Error Map. The tree error map E_{tree} localizes branch-level connectivity violations using airway centerline analysis [7]. Let M_{gt} denote the ground-truth airway mask, and let $S(\cdot)$ extract its centerline representation. Let $C_{\mathrm{root}}(M)$ denote the connected component of the predicted mask M that remains attached to the proximal airway root. We define

$$E_{\mathrm{tree}} = S(M_{\mathrm{gt}}) \wedge \neg C_{\mathrm{root}}(M), \tag{2}$$

where $\wedge$ and $\neg$ denote voxel-wise logical AND and NOT operations, respectively. Thus, E_{tree} highlights centerline voxels in the ground truth that are not connected to the root in the prediction. This sparse map isolates important connectivity failures that are not detectable from boundary discrepancy alone.

Together, the SDF and tree error maps jointly encode dense geometric deviations and sparse connectivity violations, forming a structural error representation for controlled refinement. Although defined using ground-truth discrepancy for supervision during training, inference relies solely on the predicted mask and CT volume, without access to ground truth. During training, error maps are constructed from ground truth for supervision, while at inference the diagnosis network predicts these maps directly from (M, I).

2.2 Failure-Aware Diagnosis and Refinement

To enable diagnosis-conditioned repair, we adopt a two-stage design that first verifies structural consistency using CT evidence and then performs refinement conditioned on the predicted error maps. Unlike prior approaches that directly modify the predicted mask [19, 8, 16], we explicitly separate (i) error localization and (ii) error correction. This separation ensures that updates are driven by localized structural failures and minimizes unintended changes in regions that are already correct.

Failure Verification Stage. In the first stage, we verify the predicted segmentation rather than re-segmenting the CT. The predicted mask is treated as a hypothesis, and the CT volume is used only to assess whether it is anatomically consistent. During inference, the model receives only the predicted mask and CT image and outputs structural error maps. To ensure that CT evidence does not overwrite the mask, we use asymmetric cross-attention, i.e, mask features act as queries, while CT features act as keys and values. This design allows CT information to highlight inconsistencies in the existing segmentation, but prevents it from reconstructing structures directly from appearance.

Let $F_M = f_M(M)$ and $F_I = f_I(I)$ denote mask and CT feature embeddings extracted by lightweight 3D convolutional blocks. Each block consists of two

$3 \times 3 \times 3$ convolutions with instance normalization and ReLU, while decoder upsampling uses 3D transposed convolutions with skip connections. We employ a *spatially reduced cross-attention* operator in which query, key, and value embeddings are projected and uniformly downsampled using strided 3D average pooling with reduction factor $s = 4$. Scaled dot-product attention is computed on the reduced grid, and the upsampled attended features are added residually to the mask features before task-specific decoding:

$$\hat{E}_k = g_k(F_M + \mathrm{Up}(\mathrm{Attn}(F_M, F_I, F_I))), \quad k \in \{\mathrm{tree}, \mathrm{SDF}\}. \tag{3}$$

By computing attention on a reduced-resolution grid and adding updates residually to mask features, the module evaluates structural consistency regionally. CT evidence modulates the existing segmentation without reconstructing it, highlighting inconsistent regions while preserving reliable structures.

Error-Conditioned Refinement. In the refinement stage, the predicted mask remains the primary representation, while structural error maps provide localized correction cues. We perform cross-attention with mask features as queries and encoded error-map features as keys and values, ensuring that updates are applied relative to the current segmentation state and only where structural errors are indicated. Unlike approaches that directly fuse CT and mask features during repair [16], refinement is conditioned exclusively on the predicted failure fields, preventing appearance-induced topology drift under domain shift.

Let $\hat{E} = \{\hat{E}_{\mathrm{tree}}, \hat{E}_{\mathrm{SDF}}\}$ denote the predicted error maps from the diagnosis stage, and let $F_E = f_E(\hat{E})$ denote their encoded representation. Let Z denote decoder features extracted from the initial mask. We adopt the same spatially reduced cross-attention design as in the diagnosis stage, replacing CT features with encoded error-map features as keys and values. The voxel-wise residual correction is predicted as

$$R = g_{\mathrm{dec}}(\mathrm{Attn}(Z, F_E, F_E)). \tag{4}$$

The refined probability map is obtained as $\hat{P} = \mathrm{clip}(M + R, 0, 1)$, and the final segmentation is defined as $\hat{M} = \mathbf{1}(\hat{P} > 0.5)$. By modeling refinement as residual correction rather than full regeneration, the model performs spatially localized, structurally guided updates conditioned on the predicted failure maps.

Training Objective. We jointly optimize the diagnosis and refinement stages in an end-to-end manner:

$$\mathcal{L} = \mathcal{L}_{\mathrm{err}} + \mathcal{L}_{\mathrm{ref}}. \tag{5}$$

The error-map supervision term is defined as:

$$\mathcal{L}_{\mathrm{err}} = \lambda_{\mathrm{err}}^{\ell_1} \|\hat{E} - E\|_1 + \lambda_{\mathrm{err}}^{\mathrm{bce}} \mathcal{L}_{\mathrm{BCE}}(\hat{E}, E), \tag{6}$$

where $\|\cdot\|_1$ denotes the voxel-wise ℓ_1 loss and $\mathcal{L}_{\mathrm{BCE}}$ denotes binary cross-entropy. $E = [E_{\mathrm{tree}}, E_{\mathrm{SDF}}]$ and $\hat{E} = [\hat{E}_{\mathrm{tree}}, \hat{E}_{\mathrm{SDF}}]$ denote the concatenated ground-truth and predicted error maps derived from M_{gt} and M, respectively.

The refinement supervision term is:

$$\mathcal{L}_{\mathrm{ref}} = \lambda_{\ell_1} \|\hat{P} - M_{\mathrm{gt}}\|_1 + \lambda_{\mathrm{dice}} \mathcal{L}_{\mathrm{Dice}}(\hat{P}, M_{\mathrm{gt}}) + \lambda_{\mathrm{bce}} \mathcal{L}_{\mathrm{BCE}}(\hat{P}, M_{\mathrm{gt}}), \tag{7}$$

where $\mathcal{L}_{\text{Dice}}$ denotes the soft Dice loss [11], $\hat{P}$ is the refined probability map, and M_{gt} is the ground-truth mask.

3 Experiments

3.1 Experimental Setting

To evaluate FDRAS as a backbone-agnostic post-hoc reliability layer, we train it on the 300 publicly released ATM cases [22], split at the patient level into 240 training and 60 validation scans. Initial airway segmentations are generated using nnUNet from TotalSegmentator [18], and FDRAS is trained on frozen backbone predictions without updating backbone weights. All volumes are resampled to a unified spatial resolution and intensity-normalized. The model is optimized with AdamW (learning rate 1×10^{-3}) for 20,000 iterations with batch size 1 and standard augmentation (random flipping and rotation). Loss weights are $\lambda_{\ell_1} = 0.1$, $\lambda_{\text{dice}} = 2$, $\lambda_{\text{bce}} = 1$, $\lambda_{\text{err}}^{\ell_1} = 1$, and $\lambda_{\text{err}}^{\text{bce}} = 1$ without scenario-specific tuning for individual datasets or backbones. Experiments are conducted on a NVIDIA RTX 4090 GPU, with an average inference time of 25 s per case.

For evaluation, the same FDRAS checkpoint is applied to initial segmentations produced by nnUNet [18] and TfeNet [20], without TfeNet-specific retraining. For cross-dataset robustness, the trained model is directly applied to two external datasets under strict zero-shot settings: AirRC [9] (254 CT scans with pulmonary nodules) and AeroPath [15] (27 CT scans with airway pathologies). We compare against representative post-processing methods for airway topology repair, including PTR [19], UTR [8], and ATRNet [16]. As pretrained weights are unavailable, all baselines are retrained on ATM using identical backbone predictions, following the original papers, with hyperparameters tuned on the validation split. Following prior work [20], we report Dice (DSC), tree detection rate (TD), branch detection rate (BD), and precision (Pre) as case-wise means.

3.2 Results

Quantitative Results. Under zero-shot evaluation, FDRAS consistently achieves the best structural performance across both datasets (Table 1). Compared with the strongest competing refinement, TD/BD improve by +2.0/+1.6 pp (AirRC) and +2.4/+3.1 pp (AeroPath) with nnUNet, and by +4.0/+4.8 pp (AirRC) and +0.9/+1.2 pp (AeroPath) with TfeNet. Although precision decreases in some settings—most notably on AeroPath with TfeNet—the gains reflect improved recovery of distal branches rather than spurious over-segmentation (TD improved in 25/27 cases and BD in 19/27 cases). Overall, FDRAS enhances structural completeness and connectivity under domain shift. By restoring disconnected distal branches, FDRAS improves global tree integrity and anatomical coverage, which is clinically important for bronchoscopy navigation and surgical planning where missing peripheral branches can compromise path planning and lesion localization [21]. The consistent gains in TD and BD therefore indicate improved functional reliability under domain shift.

Table 1. Quantitative comparison on the AirRC [9] and AeroPath [15] datasets under strict zero-shot cross-dataset evaluation. All metrics are reported in percentage (%). Best results are highlighted in bold.

Method	AirRC				AeroPath			
	DSC	TD	BD	Pre	DSC	TD	BD	Pre
nnUNet [18]	68.3	34.9	32.6	**98.1**	82.5	61.7	51.3	**96.7**
nnUNet [18]+PTR [19]	72.9	46.5	42.3	85.7	84.1	72.5	60.8	91.3
nnUNet [18]+ATRNet [16]	68.3	35.3	32.9	**98.1**	82.5	60.7	51.3	96.6
nnUNet [18]+UTR [8]	74.1	46.4	42.3	88.6	84.4	72.6	60.8	92.3
nnUNet [18]+Ours	**76.8**	**48.5**	**43.9**	88.8	**85.2**	**75.0**	**63.9**	88.6
TfeNet [20]	80.9	61.9	54.4	96.7	**87.2**	90.5	87.4	85.4
TfeNet [20]+PTR [19]	81.8	62.7	55.1	93.8	86.5	91.4	88.5	87.9
TfeNet [20]+ATRNet [16]	82.1	62.8	55.2	**97.9**	86.3	91.4	88.5	**89.0**
TfeNet [20]+UTR [8]	81.5	62.7	55.1	92.7	86.3	91.4	88.6	87.3
TfeNet [20]+Ours	**84.8**	**66.8**	**60.0**	88.2	83.7	**92.3**	**89.8**	78.8

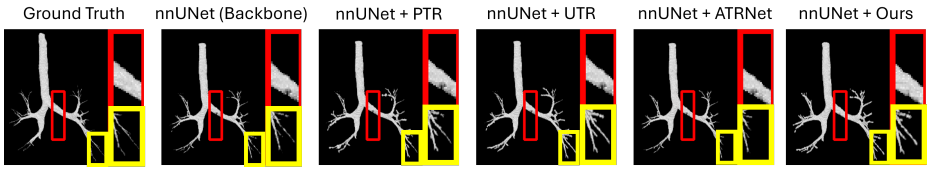

Fig. 2. Qualitative comparison on an AeroPath case under strict zero-shot evaluation. From left to right: ground truth, nnUNet backbone, nnUNet+PTR [19], nnUNet+UTR [8], nnUNet+ATRNet [16], and nnUNet+Ours. Red boxes highlight larger proximal airway regions, while yellow boxes highlight distal airway regions. Our method produces a more structurally solid and hierarchically coherent airway tree, with improved branch continuity.

Qualitative Results. Figure 2 shows a representative AeroPath case under strict zero-shot evaluation. The backbone exhibits distal boundary irregularities and structural thinning, partially mitigated by PTR and UTR. In contrast, FDRAS restores branch continuity and yields a more coherent airway tree with smoother boundaries. Some peripheral branches appear slightly thicker than the annotation, likely due to CT resolution limits and thin-airway labeling uncertainty, explaining the precision decrease while improving structural completeness without introducing anatomically implausible structures.

Ablation Results. Table 2 demonstrates that both structural error conditioning and cross-attention are critical for topology-aware repair. Removing cross-attention causes the largest degradation, with TD dropping by over 7% on both datasets, while the precision increase reflects conservative yet incomplete predictions. Removing explicit error conditioning further reduces performance, including a notable DSC drop (–7.7% on AirRC) and decreased BD, indicating insufficient guidance under domain shift. Overall, explicit failure localization and attention-based conditioning are jointly necessary for effective airway repair.

Table 2. Key ablation study on the AirRC [9] and AeroPath [15] datasets using nnUNet as the backbone. All metrics are reported in percentage (%). CAT: cross-attention modules; EM: structural error-map conditioning.

Method	AirRC				AeroPath			
	DSC	TD	BD	Pre	DSC	TD	BD	Pre
nnUNet+Ours (Full)	**76.8**	**48.5**	**43.9**	88.8	**85.2**	**75.0**	**63.9**	88.6
nnUNet+Ours w/o CAT	75.0	41.1	38.5	**93.3**	83.7	66.7	56.8	**89.9**
nnUNet+Ours w/o EM	69.1	45.9	41.2	75.3	78.7	73.8	61.8	77.3

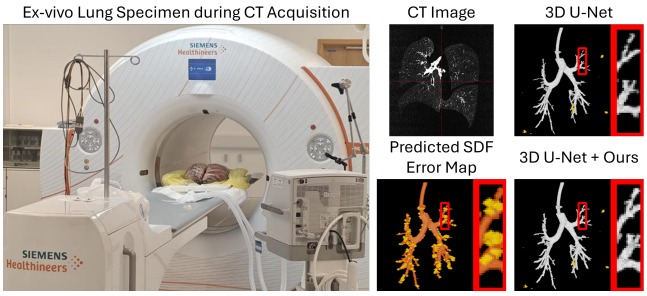

Fig. 3. Validation on an ex-vivo human lung specimen (ethics approval obtained). The backbone segmentation is generated by a 3D U-Net [3] trained on separate private ex-vivo data. The figure shows the ex-vivo lung specimen in the CT scanner, the corresponding CT slice, backbone prediction with zoomed-in view, FDRAS refinement, and the predicted SDF-based error map (warmer colors indicate larger geometric discrepancy). Yellow regions in the U-Net and FDRAS outputs denote nodule tissue. Red boxes highlight distal airway discontinuities in the backbone prediction.

Ex-vivo Validation. To evaluate robustness under a more significant domain shift, we further applied FDRAS without retraining to repair predictions obtained from an ex-vivo human lung (Fig. 3). This setting introduces substantial anatomical and acquisition differences compared to the in-vivo training data. The backbone output exhibits distal branch discontinuities and structural thinning (red boxes), while the predicted SDF error map highlights geometric inconsistencies that guide localized correction. FDRAS restores branch continuity and produces a more coherent airway trajectory extending closer to the nodule region, providing qualitative evidence of generalization across a pronounced in-vivo to ex-vivo distribution shift.

4 Conclusion

We presented FDRAS, a backbone-agnostic failure diagnosis and repair framework that externalizes geometric and topological inconsistencies as structural error maps and performs error-conditioned refinement. By decoupling failure localization from correction, FDRAS consistently improves branch detection and

connectivity under strict zero-shot evaluation, with additional qualitative robustness observed in an ex-vivo setting. As a post-hoc reliability layer, its corrective capacity remains bounded by the quality of the initial segmentation, and branches entirely absent from the initial mask may not be recoverable. The current tree-error formulation relies on skeleton-based comparison and may be sensitive to voxel resolution and resampling. Future work will strengthen geometry- and topology-aware modeling through distance-transform–based geometric decoupling [17], sensitivity analysis of architectural and loss settings, and broader validation on diverse clinical datasets.

Acknowledgments. This work was supported by the UKRI Medical Research Council (MR/Y011694/1), Baillie Gifford Pandemic Science Hub, and the Engineering and Physical Sciences Research Council (EPSRC) 'U-care' Programme Grant (EP/T020903/1).

Disclosure of Interests. The authors declare no conflicts of interest in this paper.

References

1. Audelan, B., Delingette, H.: Unsupervised quality control of segmentations based on a smoothness and intensity probabilistic model. Medical Image Analysis **68**, 101895 (2021)
2. Bogensperger, L., Narnhofer, D., Falk, A., Schindler, K., Pock, T.: Flowsdf: Flow matching for medical image segmentation using distance transforms. International Journal of Computer Vision **133**(7), 4864–4876 (2025)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
4. Clipp, R.B., Vicory, J., Horvath, S., Mitran, S., Kimbell, J.S., Rhee, J.S., Enquobahrie, A.: An interactive, patient-specific virtual surgical planning system for upper airway obstruction treatments. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC). pp. 5802–5805. IEEE (2018)
5. Fournel, J., Bartoli, A., Bendahan, D., Guye, M., Bernard, M., Rauseo, E., Khanji, M.Y., Petersen, S.E., Jacquier, A., Ghattas, B.: Medical image segmentation automatic quality control: A multi-dimensional approach. Medical Image Analysis **74**, 102213 (2021)
6. Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. IEEE Transactions on Biomedical Engineering **69**(3), 1173–1185 (2021)
7. 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)
8. Li, L., Wang, H., Baugh, M., Ma, Q., Zhang, W., Ouyang, C., Rueckert, D., Kainz, B.: Universal topology refinement for medical image segmentation with polynomial feature synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 670–680. Springer (2024)
9. Liu, J., Zhang, Z., Niu, B., Kang, S., Ren, J., Wang, L., Xu, K.: a custom annotated dataset for segmentation of pulmonary veins, arteries, and airways. Scientific Data **12**(1), 1806 (2025)

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 Transactions on Pattern Analysis and Machine Intelligence **25**(2), 265–270 (2003)
11. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
12. Nan, Y., Del Ser, J., Tang, Z., Tang, P., Xing, X., Fang, Y., Herrera, F., Pedrycz, W., Walsh, S., Yang, G.: Fuzzy attention neural network to tackle discontinuity in airway segmentation. IEEE transactions on neural networks and learning systems **35**(6), 7391–7404 (2023)
13. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H.: cldice-a novel topology-preserving loss function for tubular structure segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16560–16569 (2021)
14. Specktor-Fadida, B., Ben-Sira, L., Ben-Bashat, D., Joskowicz, L.: Segqc: a segmentation network-based framework for multi-metric segmentation quality control and segmentation error detection in volumetric medical images. Medical Image Analysis **103**, 103638 (2025)
15. Støverud, K.H., Bouget, D., Pedersen, A., Leira, H.O., Amundsen, T., Langø, T., Hofstad, E.F.: Aeropath: An airway segmentation benchmark dataset with challenging pathology and baseline method. Plos one **19**(10), e0311416 (2024)
16. Tang, Z., Nan, Y., Walsh, S., Yang, G.: Adversarial transformer for repairing human airway segmentation. IEEE Journal of Biomedical and Health Informatics **27**(10), 5015–5022 (2023)
17. Wang, K., Zhang, X., Lu, Y., Zhang, W., Huang, S., Yang, D.: Gsal: Geometric structure adversarial learning for robust medical image segmentation. Pattern Recognition **140**, 109596 (2023)
18. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023)
19. Weng, Z., Yang, J., Liu, D., Cai, W.: Topology repairing of disconnected pulmonary airways and vessels: baselines and a dataset. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 382–392. Springer (2023)
20. Wu, Q., Wang, Y., Zhang, Q.: Direction-aware convolution for airway tubular feature enhancement network. Medical Image Analysis p. 103882 (2025)
21. Zhang, M., Gu, Y.: Towards connectivity-aware pulmonary airway segmentation. IEEE Journal of Biomedical and Health Informatics **28**(1), 321–332 (2023)
22. Zhang, M., Wu, Y., Zhang, H., Qin, Y., Zheng, H., Tang, W., Arnold, C., Pei, C., Yu, P., Nan, Y., et al.: Multi-site, multi-domain airway tree modeling. Medical image analysis **90**, 102957 (2023)
23. Zhang, S., Nan, Y., Fang, Y., Wang, S., Liu, Y., Papanastasiou, G., Gao, Z., Li, S., Walsh, S., Yang, G.: Dynamical multi-order responses and global semantic-infused adversarial learning: A robust airway segmentation method. Medical Image Analysis p. 103867 (2025)