



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Airway segmentation based on topological structure enhancement using multi-task learning

Xuan Yang¹, Lingyu Chen², Yuchao Zheng³, Longfei Ma³, Fang Chen⁴, Guochen Ning⁵, Hongen Liao^{3,4} (✉)

¹ Shenzhen International Graduate School, Tsinghua University, Shenzhen, China

² College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China

³ School of Biomedical Engineering, Tsinghua University, Beijing, China
liao@tsinghua.edu.cn

⁴ School of Biomedical Engineering and the Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China

⁵ School of Clinical Medicine, Tsinghua University, Beijing, China

Abstract. Airway segmentation in chest computed tomography (CT) images is critical for tracheal disease diagnosis and surgical navigation. However, airway segmentation is challenging due to complex tree structures and branches of different sizes. To enhance airway integrity and reduce fractures during bronchus segmentation, we propose a novel network for airway segmentation, using centerline detection as an auxiliary task to enhance topology awareness. The network introduces a topology embedding interactive module to emphasize the geometric properties of tracheal connections and reduce bronchial breakage. In addition, the proposed topology-enhanced attention module captures contextual and spatial information to improve bronchioles segmentation. In this paper, we conduct qualitative and quantitative experiments on two public datasets. Compared to several state-of-the-art algorithms, our method outperforms in detecting terminal bronchi and ensuring the continuity of the entire trachea while maintaining comparable segmentation accuracy. Our code is available at https://github.com/xyang-11/airway_seg.

Keywords: Airway segmentation, Topology embedding, Topology enhancement.

1 Introduction

Segmenting airways in the chest CT scans for the diagnosis and surgical treatment of lung disease is crucial. Accurately extracting pulmonary airways from CT and quantifying morphological changes in the airways can assist doctors in diagnosing diseases such as idiopathic pulmonary fibrosis [1], coronavirus disease (COVID-19), and chronic obstructive pulmonary disease (COPD) [2]. On the other hand, accurate airway modeling is of great significance in bronchoscopy navigation. By modeling the airway

before surgery, doctors can simulate and plan the surgical route, reduce surgical risks, and maximize the protection of surrounding normal tissues.

As convolutional neural networks (CNNs) are widely used in various medical image processing tasks, many related research studies have also emerged in the airway segmentation task. Juarez et al. [3] learned topological information by stacking multiple graph convolutional layers to reduce false negatives in the trachea. Wang et al. [4] designed a radial distance loss for tubular structures to measure the topological error of segmentation. Qin et al. [5] introduced anatomical priors based on feature recalibration and knowledge distillation to improve the network's sensitivity to the bronchi. Wang et al. [6] designed a bronchioles-sensitive loss function and trained a deep learning network using an iterative training strategy. However, the work mentioned above partly added topological priors but are still insufficient for bronchioles.

Due to the complex tree structure of airways and the specificity between different individuals, it is difficult to obtain complete fine-grained segmentation results. Currently, the task of accurate airway segmentation still faces the following challenges. 1) There is a significant difference in the total volume of the main trachea and the distal peripheral bronchi, resulting in a serious imbalance within the class. Fine features of the peripheral trachea may disappear with down-sampling and cannot be recaptured in the decoder part, making accurate segmentation difficult. 2) Complex tree topology, with terminal bronchi forming many thin and divergent branches. During the segmentation of peripheral small airways, there is a recurrent phenomenon of breakage or leakage. 3D-UNet [7] serves as the backbone for [3-6], while the segmentation results of 3D-UNet typically exist the airway topology breakage problem. Figure 1(b) shows the airway segmentation result using 3D-UNet. It produces a lot of fractures compared to the ground truth in Figure 1(a).

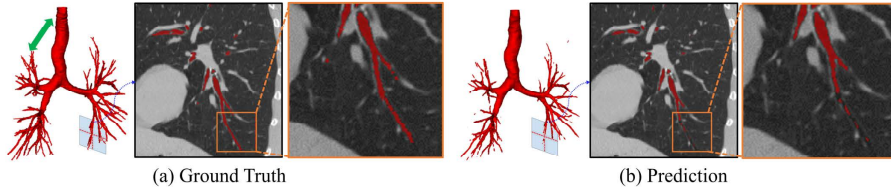


Fig. 1. (a) Ground Truth of the airway. (b) Prediction of the airway using 3D-UNet trained on the public dataset. The blue box is the sagittal section. The red-marked area depicts the airway. The green arrow shows the scale difference between the trachea and the peripheral bronchi.

To address the challenges of intra-class imbalance and bronchi breakage, we propose a multi-task learning approach for airway segmentation using an encoder-decoder architecture. First, to fully use spatial perception and context awareness information, we introduce a topology-enhanced attention module in skip connections. The module adaptively fuses high-level and low-level features and makes full use of spatial information and semantic information to capture and locate features that are easy to disappear. In order to learn the global topology of the airway tree, we introduce centerline detection as an auxiliary task, construct a multi-task learning architecture, and further design a feature fusion module. The features extracted from the centerline detection task interact

with the airway segmentation features. Learning topological perception by introducing anatomical prior knowledge naturally promotes the continuity of airway segmentation, and reduces peripheral bronchi breakage.

2 Methods

The overview of the proposed method is depicted in Figure 2. We innovatively perform a multi-task learning network with embedded topological priors. It incorporates centerline detection as an auxiliary task to preserve the overall connectivity of the segmentation. First, we design an interactive module to exchange features between semantic segmentation and centerline detection. Second, we embed an attention module within each skip connection to amalgamate segmentation outputs with encoder features, selecting saliency for enhanced refinement.

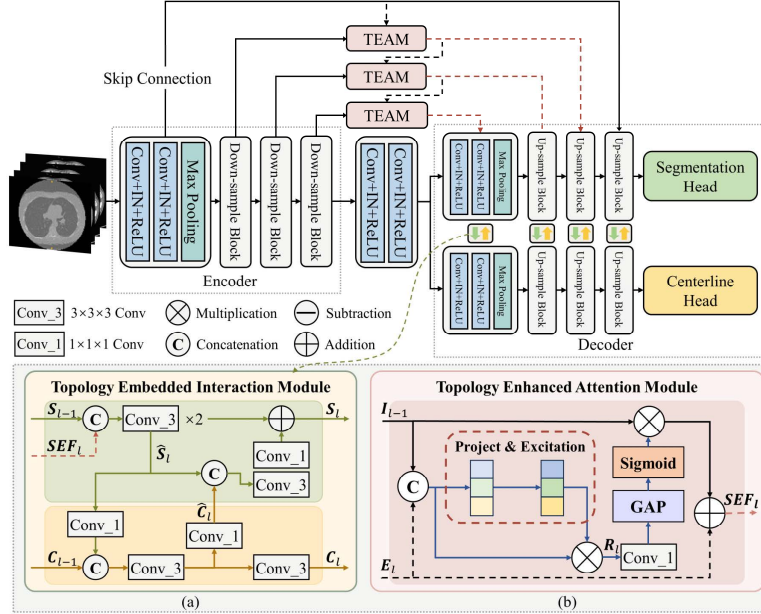


Fig. 2. The proposed network framework diagram. (a) Topology Embedded Interaction Module. The green branch is the airway segmentation task (S), the yellow branch is the centerline detection task (C). (b) Topology Enhanced Attention Module. GAP is global average pooling.

2.1 Topology Embedded Interaction Module

Inspired by [8], we devise a Topology Embedding Interactive Module (TEIM) to capture the interplay of features between segmentation and centerline detection. Figure 2(a) depicts the architecture of TEIM. Here, $S_l \in R^{C_l \times D_l \times H_l \times W_l}$ and $C_l \in R^{C_l \times D_l \times H_l \times W_l}$ denote the learned features within the segmentation task and centerline detection task, respectively, where l indicates the layer number. Notations C_l , D_l , H_l , and W_l are

referring to the number of channels, depths, heights, and widths of the l -th layer. $SEF_l \in R^{C_l \times D_l \times H_l \times W_l}$ refers to attention features from TEIM.

The design of TEIM is as follows: within the segmentation part, the refined features from TEIM are connected simply to the previous layer's S features $S_{l-1} \in R^{C_{l-1} \times D_{l-1} \times H_{l-1} \times W_{l-1}}$. Following this, two convolutional layers are employed to update the learned features \hat{S}_l . These updated features are then transmitted to the centerline detection pathway. $\text{Conv}_3(\cdot)$ denotes $3 \times 3 \times 3$ convolution, $\text{Conv}_1(\cdot)$ denotes $1 \times 1 \times 1$ convolution, and $[\dots]$ is the concatenation operation.

$$\hat{S}_l = \text{Conv}_3(\text{Conv}_3[S_{l-1}, SEF_l]) \quad (1)$$

The \hat{S}_l concatenates the previous features $C_{l-1} \in R^{C_{l-1} \times D_{l-1} \times H_{l-1} \times W_{l-1}}$ and performs feature fusion through convolutional operations. Subsequently, the fused features from C are sent back to S. They are added to \hat{S}_l , forming the final segmentation features S_l .

$$\hat{C}_l = \text{Conv}_1(\text{Conv}_3[C_{l-1}, \text{Conv}_1(\hat{S}_l)]) \quad (2)$$

$$S_l = \text{Conv}_1(\text{Conv}_3[\hat{C}_l, \hat{S}_l]) + \hat{S}_l \quad (3)$$

Taking centerline detection as an auxiliary task, TEIM transfers features from centerline detection to segmentation branches, thus aiming at forming a weak supervision signal that maintains the global topology.

2.2 Topology Enhanced Attention Module

The remarkable performance of U-Net heavily relies on the skip connections, which amalgamate low-level features from the encoder with high-level features from the decoder. In order to better process the complex anatomical semantic information and enhance the network's responsiveness to different scales of airway, we proposed a topology enhanced attention module (TEAM) to optimize the original skip connections. TEAM first fuses feature maps of different scales at adjacent levels, and then uses feature recalibration to build cross-scale dependencies. By merging low-level spatial and high-level semantic feature maps, TEAM is used to build cross-scale dependencies and enhance the network's response to various airway scales.

Figure 2(b) illustrates the architecture of TEAM. First, the encoder feature maps of adjacent levels are connected and the correlation of the fused feature channels is modeled. The coding features of this level are expressed as $E_l \in R^{C_l \times D_l \times H_l \times W_l}$. The upper layer encoder feature is expressed as $E_{l-1} \in R^{C_{l-1} \times D_{l-1} \times H_{l-1} \times W_{l-1}}$. For the first level of TEAM, the input I_0 is defined as the down-sampling result of E_0 . For the other levels, the input I_{l-1} is the down-sampling result of SEF_{l-1} (the previous level's TEAM output).

$$R_l = P\&E(\text{Concat}[E_l + I_{l-1}]) \quad (4)$$

The module first connects high-level features with rich semantic information and low-level features with spatial information. In order to make the network to adaptively extract important features at different levels, we use feature recalibration for feature

response adjustment. $P\&E(\cdot)$ stands for Project and Excite module, which shares a similar structure with [9] and completes feature recalibration through two steps of projection and excitation. It aims to represent the features of smaller structures better.

In order to further extract global context information, we use global average pooling and Sigmoid activation operations to process R_l . The generated attention weight mask is then multiplied by the low-level features, and finally added to the high-level features to get SEF_l .

$$SEF_l = E_l + (Concat[E_l + I_{l-1}]) \times Sigmoid(GAP(Conv_1(R_l))) \quad (5)$$

Ultimately, the results from both branches are merged. Guided by TEAM, E_l is enhanced to refine features SEF_l , subsequently connected to the decoder for improved reconstruction.

2.3 Loss Functions

Our network has two decoders, so we choose different loss functions for the two tasks during training. A combination of dice coefficient loss [10] and binary cross entropy loss is used for the segmentation task. To maintain the topological completeness of the predicted airway labels, the cl-Dice loss [11] is applied to supervise the centerline in the topology decoder.

We use multi-scale loss for training. After each output feature map in the decoder, a $1 \times 1 \times 1$ convolutional layer with sigmoid activation is added to generate the prediction map of each layer. The output prediction map of each layer is up-sampled to the original image size by bilinear interpolation, and then the loss is calculated using the ground truth. The multi-scale loss is calculated as follows.

$$L^{(m)} = L^{(m)}_{dice} + L^{(m)}_{bce} + L^{(m)}_{clDice} \quad (6)$$

$$L_{total} = \sum_{m=1}^M L^{(m)} \quad (7)$$

For each layer of the decoder, the loss $L^{(m)}$. And α represents the weights corresponding to the loss of the two tasks. M is the number of decoder layers, $L^{(m)}$ represents the loss function of the m -th layer.

3 Experiments and Results

Datasets and Evaluation Metrics. We evaluated the proposed method in the BAS dataset [12]. The dataset contains 90 CT scans, 70 CT scans were sourced from the LIDC [13], while the remaining 20 were obtained from the EXACT'09 [14]. It was divided randomly into training set of 50 cases, validation set of 20 cases, and testing set of 20 cases after merging all 90 CT scans. The slice thickness of these scans varies from 0.5mm to 1mm, and the voxel spacing is in the range of 0.5-0.821 mm. For topological accuracy, we choose pixel-level evaluation metrics: the Dice Similarity Coefficient (DSC), Precision (Pre), and Sensitivity (Sen). For topology completeness, we adopt Branches detected (BD) and Tree-length detected (TD) [15].

Implementation details. We truncated the Hounsfield unit (HU) values of all scans to the range of $[-1000, 600\text{HU}]$, and then normalized to $[0, 1]$. During the training phase, elastic deformation, rotation, gaussian smoothing, and gamma transformation were applied as dynamic data augmentation techniques. The proposed network was implemented in PyTorch 2.1 and trained on a single NVIDIA RTX 3060. During training, the CT cubes were cropped to $128 \times 128 \times 128$ as the input, and then were trained with a mini-batch size of 2. The optimizer used for the network is Adam, with an initial learning rate of $1e-3$ and a weight decay of $5e-4$. The maximum epoch was 300. In addition, we produced skeleton ground truth using the robust and widely used algorithm in [16]. Two authoritative doctors confirmed the accuracy of the skeleton labels. In inference, we obtained predictions only from the semantic segmentation head.

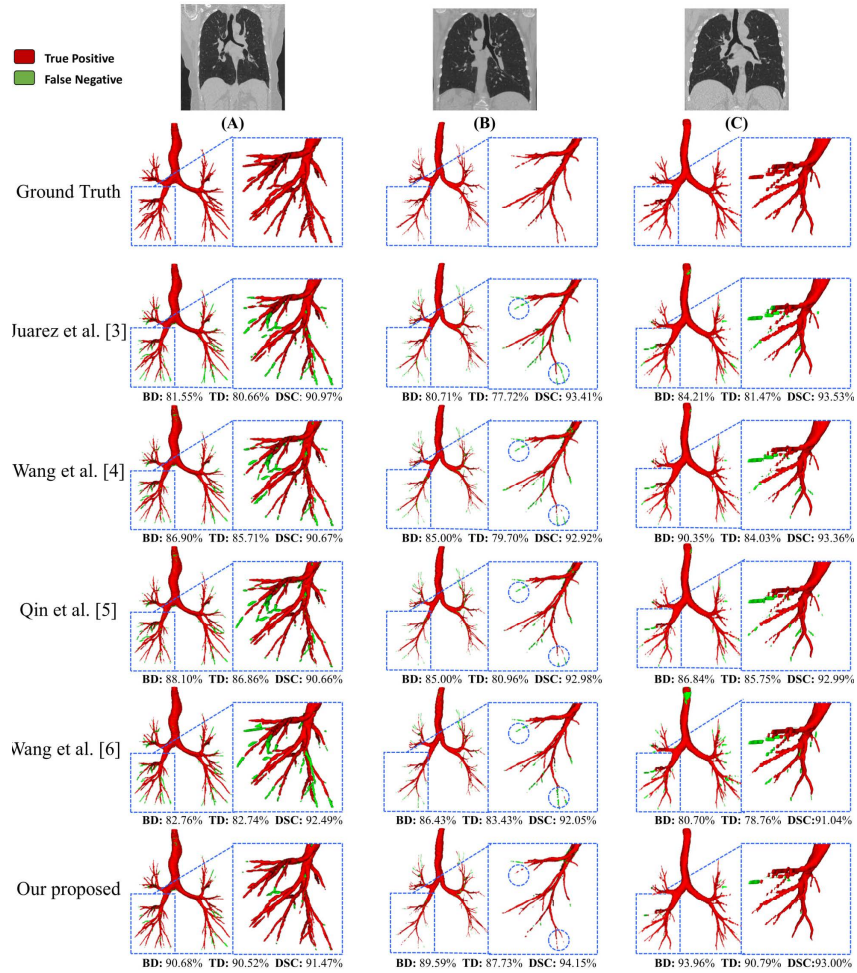


Fig. 3. Results of airway segmentation using different methods. True positive voxels are depicted in red color. False positive voxels are depicted in green color. Blue circles highlight where our method eliminates breakage.

Comparative results. Apart from U-Net, we compare the original 3D U-Net [7] and its variants, including V-Net [10] and VoxResNet [17]. In addition, we compare with four relevant state-of-the-art methods in recent years: Juarez et al. [3], Wang et al. [4], Qin et al. [5], Wang et al. [6]. Table 1 presents the quantitative results of the comparisons. Our method outperforms others in DSC, BD, and TD metrics, exhibiting competitive results in other metrics. The best DSC means that our network is more robust than other networks. Although the accuracy decreases slightly, this might be attributed to the model successfully detecting some real bronchi that are not correctly annotated, which are considered false positives during metric evaluation, resulting in the precision decline.

Table 1. Results (%) of the proposed method compared to state-of-the-art methods (Mean \pm Standard deviation)

Method	DSC	Pre	Sen	TD	BD
3D-U-Net [7]	93.88(2.35)	<u>97.79(0.63)</u>	90.38(4.60)	77.08(5.78)	78.35(6.88)
V-Net [10]	92.77(2.92)	97.80(0.76)	88.42(5.65)	73.28(7.76)	75.94(7.96)
VoxResNet [17]	<u>94.35(1.96)</u>	96.56(1.05)	92.33(3.95)	82.69(6.02)	84.26(5.59)
Juarez et al. [3]	93.87(1.48)	93.85(1.20)	93.97(3.21)	83.94(5.22)	85.48(4.70)
Wang et al. [4]	93.92(1.97)	95.49(1.25)	92.47(3.79)	86.29(5.56)	87.85(4.15)
Qin et al. [5]	93.92(1.89)	95.50(1.22)	92.46(3.48)	<u>86.95(5.06)</u>	<u>88.73 (3.91)</u>
Wang et al. [6]	91.63(1.35)	88.21(2.43)	95.43(2.47)	86.01(5.72)	87.78(5.23)
Our proposed	94.41(1.78)	93.59(1.76)	<u>95.30(3.58)</u>	90.19(3.81)	91.96(3.31)

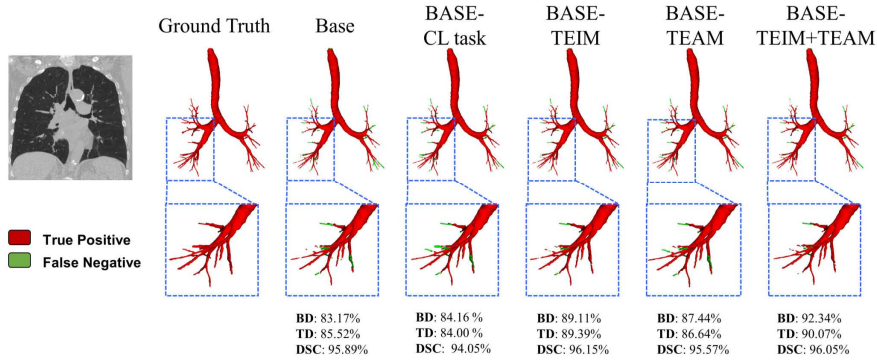
Compared to other methods, our approach increases BD and TD by approximately 4%. Since these two indicators represent the topological integrity of the airway, the results show that our method has good performance in the detection of peripheral bronchioles. The visualization results in Figure 3 further validate our method's capability to reconstruct more bronchi and improve the continuity of the peripheral bronchi. It reduces the occurrence of bronchial breakage, greatly improves topological integrity, and provides a more accurate fine-grained map for subsequent bronchoscopy navigation.

Furthermore, in order to verify the robustness of the model, we conducted model fine-tuning and validation on AeroPath [18], an airway dataset with pathological symptoms. The qualitative and quantitative results are shown in Figure A1 and Table A1 in the Appendix.

Ablation study. Table 2 displays the quantitative results of the ablation experiments. To assess the performance of individual components, we combine the encoder with the single segmentation decoder as the baseline network, denoted as **Base**. In order to prove that the centerline detection task is effective in maintaining the topology of the airway, we add a parallel decoder for centerline detection based on **Base** and represented the model as **Base-CLtask**. We add TEIM to each pair of (segmentation, centerline detection) decoders and denote the model as **Base-TEIM**. To verify the performance of TEIM, we integrate it into Base without adding a centerline detection task or TEIM.

Table 2. Comparisons (%) of ablation experiment results (Mean \pm Standard deviation)

Method	DSC	Pre	Sen	TD	BD
Base	93.31(2.46)	97.54(0.88)	89.55(4.67)	82.49(4.73)	84.30(5.06)
Base-CLtask	94.38(1.59)	95.28(1.40)	<u>93.58(3.54)</u>	86.01(4.80)	87.62(4.68)
Base-TEIM	94.57(1.73)	95.96(1.02)	93.30(3.49)	<u>88.59(4.48)</u>	<u>90.27(3.82)</u>
Base-TEAM	94.16(1.60)	<u>96.25(1.31)</u>	92.19(2.91)	86.98(2.99)	88.22(2.75)
Base-TEIM+TEAM	<u>94.41(1.78)</u>	93.59(1.76)	95.30(3.58)	90.19(3.81)	91.96(3.31)

**Fig. 4.** Comparisons of ablation experiments visualization results. True positive voxels are depicted in red color. False positive voxels are depicted in green color.

Analyzing the results in Table 2, we observe that compared to **Base**, most indicators of **Base-CLtask** have improved. It can be inferred that adding the auxiliary task of centerline detection to the model will help improve model performance. **Base-TEIM** demonstrates outstanding performance in topological completeness metrics, exhibiting an approximate 6% improvement over the Base model. As seen from the example in Figure 4, **Base-CLtask** can learn some new connections during segmentation, and **Base-TEIM** restores more small branches, eliminates some breaks, and obtains a more complete topology. **Base-TEAM** notably outperforms other models in the sensitivity index, showcasing a 4.5% and 3.9% improvement over **Base** in TD and BD, respectively. It means that **Base-TEAM** can detect important branches and bronchioles more correctly. By learning more local and global semantic information, the network is able to solve the problem of intra-class imbalance in the airway. The ablation experiment results verified the effectiveness of the proposed method.

4 Conclusion

In this paper, we presented a multi-task learning network to deal with bronchi breakage based on U-Net for airway segmentation. Centerline detection, as an auxiliary task, was conducted simultaneously with the segmentation task to aid the airway prediction. To comprehensively extract the topological information from the auxiliary tasks, the TEIM

was proposed to facilitate interactive feature exchange between the two tasks. Additionally, the TEAM attempted to enable the network to focus more broadly on finer airway branches by enhancing and refining saliency. Extensive experimental evaluations demonstrated that our framework could extract a more significant number of branches, while maintaining competitive segmentation performance compared to other state-of-the-art methods. In the future, we will conduct experiments on datasets with abnormal tracheal structures, combining doctors' subjective evaluations and further verifying the effectiveness of our proposed method.

Acknowledgments. This work was supported National Key Research and Development Program of China (2022YFC2405200), National Natural Science Foundation of China (82027807, 62201315, U22A2051), Institute for Intelligent Healthcare, Tsinghua University (2022ZLB001), Tsinghua-Foshan Innovation Special Fund (2021THFS0104) and Young Elite Scientists Sponsorship Program by CAST (2023QNRC001).

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

References

1. Wu, X., Kim, G.H., Salisbury, M.L., Barber, D., Bartholmai, B.J., Brown, K.K., Conoscenti, C.S., De Backer, J., Flaherty, K.R., Gruden, J.F.: Computed tomographic biomarkers in idiopathic pulmonary fibrosis. The future of quantitative analysis. *American journal of respiratory and critical care medicine* 199, 12-21 (2019)
2. Luo, F., Darwiche, K., Singh, S., Torrego, A., Steinfert, D.P., Gasparini, S., Liu, D., Zhang, W., Fernandez-Bussy, S., Herth, F.J.: Performing bronchoscopy in times of the COVID-19 pandemic: practice statement from an international expert panel. *Respiration* 99, 417-422 (2020)
3. Garcia-Uceda Juarez, A., Selvan, R., Saghir, Z., de Bruijne, M.: A joint 3D UNet-graph neural network-based method for airway segmentation from chest CTs. In: *Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10*, pp. 583-591. Springer (2019)
4. Wang, C., Hayashi, Y., Oda, M., Itoh, H., Kitasaka, T., Frangi, A.F., Mori, K.: Tubular structure segmentation using spatial fully connected network with radial distance loss for 3D medical images. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*, pp. 348-356. Springer (2019)
5. Qin, Y., Zheng, H., Gu, Y., Huang, X., Yang, J., Wang, L., Yao, F., Zhu, Y.-M., Yang, G.-Z.: Learning tubule-sensitive CNNs for pulmonary airway and artery-vein segmentation in CT. *IEEE transactions on medical imaging* 40, 1603-1617 (2021)
6. Wang, A., Tam, T.C.C., Poon, H.M., Yu, K.-C., Lee, W.-N.: Naviairway: a bronchiole-sensitive deep learning-based airway segmentation pipeline. *arXiv preprint arXiv:2203.04294* (2022)
7. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: *Medical Image Computing and*

- Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19, pp. 424-432. Springer (2016)
8. Cheng, M., Zhao, K., Guo, X., Xu, Y., Guo, J.: Joint topology-preserving and feature-refinement network for curvilinear structure segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7147-7156. (2021)
 9. Rickmann, A.-M., Roy, A.G., Sarasua, I., Navab, N., Wachinger, C.: ‘Project & excite’ modules for segmentation of volumetric medical scans. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22, pp. 39-47. Springer (2019)
 10. 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)
 11. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylyka, 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)
 12. Qin, Y., Gu, Y., Zheng, H., Chen, M., Yang, J., Zhu, Y.-M.: AirwayNet-SE: A simple-yet-effective approach to improve airway segmentation using context scale fusion. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 809-813. IEEE, (2020)
 13. Qin, Y., Chen, M., Zheng, H., Gu, Y., Shen, M., Yang, J., Huang, X., Zhu, Y.-M., Yang, G.-Z.: Airwaynet: a voxel-connectivity aware approach for accurate airway segmentation using convolutional neural networks. In: International conference on medical image computing and computer-assisted intervention, pp. 212-220. Springer, (2019)
 14. Lo, P., Van Ginneken, B., Reinhardt, J.M., Yavarna, T., De Jong, P.A., Irving, B., Fetita, C., Ortner, M., Pinho, R., Sijbers, J.: Extraction of airways from CT (EXACT’09). IEEE Transactions on Medical Imaging 31, 2093-2107 (2012)
 15. Zhang, M., Wu, Y., Zhang, H., Qin, Y., Zheng, H., Tang, W., Arnold, C., Pei, C., Yu, P., Nan, Y.: Multi-site, multi-domain airway tree modeling. Medical Image Analysis 90, 102957 (2023)
 16. 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, 462-478 (1994)
 17. Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.-A.: VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. NeuroImage 170, 446-455 (2018)
 18. Støverud, K.-H., Bouget, D., Pedersen, A., Leira, H.O., Lango, T., Hofstad, E.F.: AeroPath: An airway segmentation benchmark dataset with challenging pathology. arXiv preprint arXiv:2311.01138 (2023)