



AirwayFormer: Structure-Aware Boundary-Adaptive Transformers for Airway Anatomical Labeling

Weihao Yu¹, Hao Zheng^{1,3}, Yun Gu^{1(✉)}, Fangfang Xie², Jiayuan Sun²,
and Jie Yang^{1(✉)}

¹ Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China
{geron762, jieyang}@sjtu.edu.cn

² Shanghai Chest Hospital, Shanghai, China

³ Tencent Jarvis Lab, Shenzhen, China

Abstract. Pulmonary airway labeling identifies anatomical names for branches in bronchial trees. These fine-grained labels are critical for disease diagnosis and intra-operative navigation. Recently, various methods have been proposed for this task. However, accurate labeling of each bronchus is challenging due to the fine-grained categories and inter-individual variations. On the one hand, training a network with limited data to recognize multitudinous classes sets an obstacle to the design of algorithms. We propose to maximize the use of latent relationships by a transformer-based network. Neighborhood information is properly integrated to capture the priors in the tree structure, while a U-shape layout is introduced to exploit the correspondence between different nomenclature levels. On the other hand, individual variations cause the distribution overlapping of adjacent classes in feature space. To resolve the confusion between sibling categories, we present a novel generator that predicts the weight matrix of the classifier to produce dynamic decision boundaries between subsegmental classes. Extensive experiments performed on publicly available datasets demonstrate that our method can perform better than state-of-the-art methods. The code is publicly available at <https://github.com/EndoluminalSurgicalVision-IMR/AirwayFormer>.

Keywords: Airway anatomical labeling · Structural prior · Dynamic decision boundary

This work was partly supported by National Key R&D Program of China (2019YFB1311503, 2017YFC0112700), Committee of Science and Technology, Shanghai, China (19510711200), Shanghai Sailing Program (20YF1420800), NSFC (61661010, 61977046, 62003208), SJTU Trans-med Awards Research (20210101), Science and Technology Commission of Shanghai Municipality (20DZ2220400).

W. Yu and H. Zheng—Equal contribution.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43990-2_37.

1 Introduction

Automatic airway labeling aims to assign the corresponding anatomical names to the branches in airway trees. The identification of peripheral branches plays an essential role in bronchoscopic navigation. Figure 1 illustrates the hierarchical nomenclature [11] from lobar to subsegmental levels in the bronchial tree. In the first level, the naming is based on the five lung lobes, and similarly, the subtrees are further divided according to the 18 lung segments. The nomenclature of subsegmental bronchi is more complex, which contains six classes (a, b, c, a+b, a+c, and b+c) for the subtrees and their common branches.

Various methods have been proposed for airway labeling [2, 6, 7, 9, 10, 14, 16, 21, 22] in recent years. Several works [2, 6, 16] adopted graph-matching-based algorithms to clarify the candidate trees based on a reference tree. However, the performance is limited at the subsegmental level due to individual variation. Deep learning methods [14, 21, 22] are also developed for this task. Tan et al. [14] proposed a multi-task U-Net [12] with a structure-aware graph convolutional network (GCN) [5] to segment the airway tree semantically. Yu et al. [21] converted the airway tree from the image to the graph space and designed a multi-stage framework with hypergraph neural networks (HGNN) for node classification. There are two main challenges for these learning-based methods. First, 127 classes are included in the nomenclature up to subsegmental level. The annotation is quite time-consuming and labor-intensive. Classification for over one hundred categories with limited training data is an extremely hard problem. Second, the inter-individual differences regarding the location, direction, length, and diameter of branches become increasingly significant from the lobar order to the subsegmental level. Figure 1(d) demonstrates the t-distributed stochastic neighbor embedding (t-sne) results of the learned features of RB10 branches in two examples. The distribution of RB10b in the first case is overlapped with the distribution of RB10c in the second example. Such individual variation dramatically affects the generalization ability of models.

To resolve the first problem, in this work, we propose to fully use the latent relationships within the tree structure and airway nomenclature by a novel network named AirwayFormer. The tree structure provides inherent message transmission roads, while both global and local information is critical for anatomical labeling. To this end, we first adopt a neighborhood information encoding module based on the transformer block to aggregate both global and local features within the tree structure. Another structural relationship appears in the nomenclature of bronchial trees. As shown in Fig. 1(a), (b), and (c), the subsegmental bronchi RB10a belong to RB10 in the segmental level and the right lower lobe bronchi (RLL) in the lobar level, respectively. The consistency between classes at different levels is an important cue in this task. In previous works, Yu et al. [21] introduced the classification results of the current stage to the next stage as extra constraints. However, the inference is only single-directional, where the prediction is not used to refine the classification in the upper level. To achieve a bi-directional refinement, a U-shape layout is designed to explore the correspondence between different nomenclature levels.

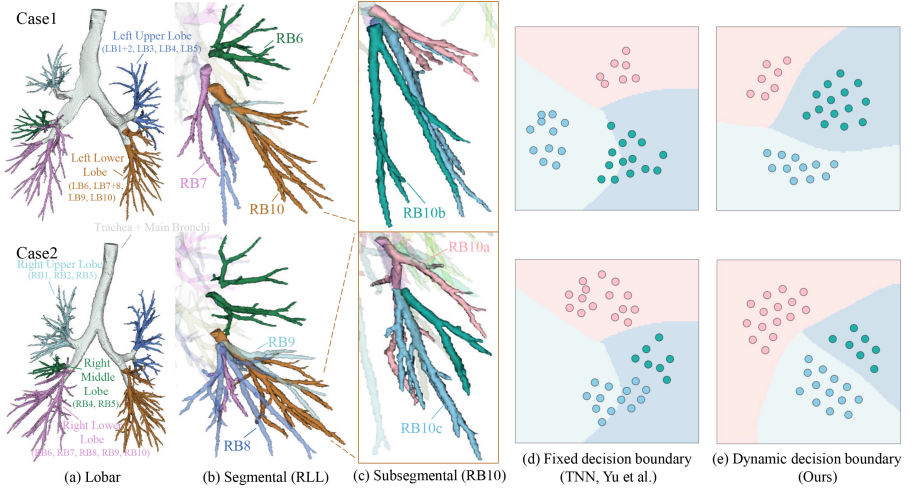


Fig. 1. Three-level nomenclature of the bronchial tree, and visualization of branch features and decision boundaries of different methods. The text color is consistent with the classes. Colors in t-sne figures are consistent with those in the airway tree. RLL is short for the right lower lobe.

For the second challenge, an ideal solution is that the model can adaptively adjust itself to deal with individual variation. Especially in subsegmental level, the same category in different cases may have distinct features. Actually, the nomenclature of subsegmental bronchi depends on their relative positions to the sibling branches within a segment. Based on this prior knowledge, we design a novel generator that learns to capture the relative relationships by predicting the weights of the last fully connected layer. More specifically, the weight for a subsegmental class is generated based on the feature representation of the corresponding segmental class. As illustrated in Fig. 1, compared with using fixed weights in the classifier, the predicted weights dynamically adjust the decision boundaries according to the characteristics of the segmental bronchi in each case, alleviating the overfitting problem caused by individual variation.

Our main contributions can be summarized as follows: (1) AirwayFormer is proposed for accurate airway labeling up to subsegmental level by exploiting the latent structural relationships. (2) A weight generator is designed to mitigate the overfitting caused by individual variation via adaptive decision boundary adjustment. (3) With extensive experiments on the public dataset, our method achieves state-of-the-art results in lobar, segmental, and subsegmental levels.

2 Method

An overview of our proposed method is illustrated in Fig. 2. Following previous work [21], each branch is regarded as a node in a graph space with well-designed 20-dimensional features. Five transformer blocks are used to classify these nodes hierarchically. The weight matrix of the subsegmental classifier is dynamically updated according to the segmental features. We denote the function of the

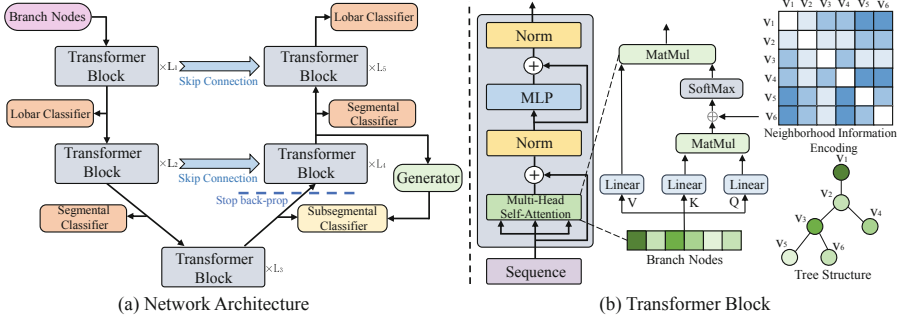


Fig. 2. Overview of the proposed method for bronchi anatomical labeling.

m -th transformer blocks as $\mathcal{F}_m(\cdot)$, $1 \leq m \leq M = 5$. $X_m \in \mathbb{R}^{n \times d}$ denotes the input feature of the corresponding transformers. The node number and feature dimension are denoted by n and d , respectively. The number of categories of the m -th classifier is c_m . More details about our method are introduced in the following sections.

2.1 Structure-Aware Transformers

We propose a transformer-based model named AirwayFormer to exploit the structural relationships in airway labeling. The tree structure is introduced into the self-attention calculation, while the class consistency in the hierarchical nomenclature is encoded in the U-shape layout.

Self-attention module in transformers enables nodes to aggregate messages from a global scale while neglecting the local structure prior. Recent studies [1, 8, 20] showed that incorporating graph information into vanilla transformers can achieve competitive performance. As the labeling of peripheral branches needs both global and local information, we adopt a neighborhood information encoding (NIE) module, which integrates the structural prior into the self-attention mechanism. Given an airway tree with n branches, each branch can be seen as a node v . The parent bronchus and children bronchus is adjacent. Then we adopt a distance function $\psi(v_i, v_j)$ to measure the spatial relation between nodes v_i and v_j . Here, $\psi(v_i, v_j)$ is defined as the shortest path distance (SPD) between v_i and v_j . Finally, a codebook \mathcal{C}_m is used to convert the SPD matrix $D \in \mathbb{R}^{n \times n}$ to learnable scale parameters, which can serve as a graph bias term of the attention map $A \in \mathbb{R}^{n \times n}$:

$$A = \frac{X_m Q (X_m K)^T}{\sqrt{d}} + \mathcal{C}_m(D). \quad (1)$$

Here, $Q \in \mathbb{R}^{d \times d}$ and $K \in \mathbb{R}^{d \times d}$ are trainable parameters. Matrix D encodes the neighborhood information of the tree structure, and the network can adaptively concentrate on the local structure by optimizing \mathcal{C}_m .

To simultaneously utilize the bi-directional correspondence between different levels, AirwayFormer adopts a U-shape layout. Concretely, the network performs

nodes classification hierarchically from lobar level to subsegmental level and then back to lobar level:

$$G_m = \mathcal{F}_m(X_m). \quad (2)$$

$$X_{m+1} = \begin{cases} G_m, & 1 \leq m < \frac{M}{2}, \\ \mathcal{CAT}(G_m, G_{M-m}), & \frac{M}{2} \leq m < M. \end{cases} \quad (3)$$

$$P_m = \mathcal{Z}_m(G_m). \quad (4)$$

Here, $G_m \in \mathbb{R}^{n \times d}$ is the output feature of the m -th transformer blocks, and $\mathcal{CAT}(\cdot)$ denotes the concatenate operation. $\mathcal{Z}_m(\cdot)$ is the m -th classifier consisting of a simple linear layer while $P_m \in \mathbb{R}^{n \times c_m}$ is the prediction result. The advantages of this design are two-fold. First, the output of coarser level is used as the input of finer level, which is helpful for the network to learn the corresponding relationship of the categories from coarse to fine. Second, the results of coarser level can be directly deduced from the predictions of finer level. Fine-grained features feedback to coarser level reduces the classification difficulty of coarser level and promotes the labeling performance. Besides, transformers conducting labeling of the same level are connected directly using highways. The same-level features can prevent performance degradation and ameliorate training stability. We further stop gradient back-propagation from segmental level to subsegmental level. The reason is that the severe distribution overlap in subsegments makes the feature learning in this level markedly different from others. This is further discussed in the supplementary materials.

2.2 Boundary-Adaptive Classifier

Individual variations cause the distribution of adjacent classes to intersect, especially in subsegmental level. Although most subsegmental bronchi in the same tree are separable according to their relative positions, the inter-individual differences seriously affect the generalization performance of the model. To introduce case-specific information, we propose a boundary-adaptive classifier that adopts a novel generator to predict case-specific weights for each subsegmental category. Let $G_k \in \mathbb{R}^{n \times d}$ denote the subsegmental output feature. We first use the next segmental level feature $G_{k+1} \in \mathbb{R}^{n \times d}$ and prediction $P_{k+1} \in \mathbb{R}^{n \times c_{k+1}}$ to obtain coarser representation for each segmental class:

$$S_{k+1} = \text{Softmax}(P_{k+1}). \quad (5)$$

$$H = (S_{k+1}^\alpha)^T G_{k+1}. \quad (6)$$

Here, $S_{k+1} \in \mathbb{R}^{n \times c_{k+1}}$ denotes the probability matrix of the segmental level while α is a learnable cluster parameter. $H = (h_1, h_2, \dots, h_{c_{k+1}})^T \in \mathbb{R}^{c_{k+1} \times d}$ are representations for c_{k+1} segmental categories. However, the segmental prediction P_{k+1} is not always perfect and may contain some classification errors. To avoid potential error propagation and obtain better class representations, these cluster

centers are refined using a vanilla transformer. Specifically, we take H as query vectors and G_{k+1} as key and value vectors:

$$\tilde{H} = \text{transformer}(q = H, k = v = G_{k+1}). \quad (7)$$

After getting the enhanced segmental class representations, a generator network $\mathcal{T}(\cdot)$ is used to produce classifier weights for subsegmental categories. Since each segment contains six subsegmental bronchi, $\mathcal{T}(\cdot)$ transforms each segmental center into the seven (including itself) classifier weights:

$$w_i = \mathcal{T}(\tilde{h}_i) = U(\text{Gelu}(V(\tilde{h}_i))), \quad 1 \leq i \leq c_{k+1}. \quad (8)$$

Here, $\text{Gelu}(\cdot)$ denotes the GELU activation function [3]. A bottleneck architecture is used to limit the number of parameters: $V \in \mathbb{R}^{d \times b}$ and $U \in \mathbb{R}^{b \times 7 \times d}$ are down-projection and up-projection learnable matrices respectively where $b \ll d$. Finally, the weight matrix of the subsegmental classifier can be obtained by stacking $w_i \in \mathbb{R}^{7 \times d}$.

We use cross-entropy loss with labeling smoothing [13] as the loss function for each task. Then the total loss function can be formulated as

$$\mathcal{L}_{total} = \sum_{m=1}^M \gamma_m \mathcal{L}_{ce}, \quad (9)$$

where γ_m is the weight of the m -th loss function.

3 Experiments and Results

3.1 Dataset and Implementation Details

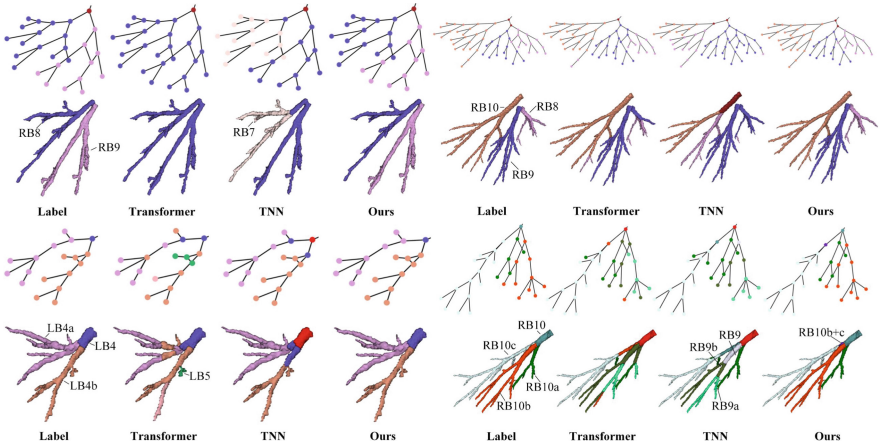
We evaluated our method on the public Airway Tree Labeling (ATL) Dataset¹ [21]. The dataset contains 104 labeled bronchial trees from CT scans whose slice thickness ≤ 0.67 mm and spatial resolution ranges from 0.78 mm to 0.82 mm. The annotation is three-level, including six lobar bronchi, 19 segmental bronchi, and 127 subsegmental bronchi. We conducted 4-fold cross-validation on the dataset.

We set the label smoothing hyperparameter σ of the loss function to 0.02. The weight of loss function γ_m was set to 1. More experiments about hyperparameters can be found in the supplementary materials. We stacked two transformer blocks for airway labeling of each level. The model was trained using Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$) with a learning rate of $5e^{-4}$ for 800 epoches. Predictions of the last lobar and segmental classifiers were used to be the final results of these levels. Four evaluation metrics were used: (a) accuracy (ACC), (b) precision (PR), (c) recall (RC), and (d) F1 score (F1). All the networks were implemented in PyTorch framework with a GeForce GTX TITAN XP GPU.

¹ <https://github.com/yuyouxixi/airway-labeling/tree/main/dataset>.

Table 1. Comparison with different methods in airway anatomical labeling (%).

Method	Lobar				Segmental				Subsegmental			
	ACC	PR	RC	F1	ACC	PR	RC	F1	ACC	PR	RC	F1
<i>(A) GNNs</i>												
GCN [5]	98.0	96.8	96.1	96.5	85.4	86.3	84.0	85.2	60.2	59.5	62.2	60.8
GAT [18]	98.6	98.0	97.0	97.5	88.6	89.6	88.7	89.2	70.7	63.7	66.6	65.1
DAGNN [15]	98.5	97.7	97.2	97.4	89.1	88.6	88.1	88.3	74.0	67.8	71.1	69.4
<i>(B) HGNNs</i>												
HyperGCN [19]	98.4	98.1	96.5	97.3	87.6	83.9	85.1	84.5	61.8	48.2	52.5	50.3
UniSAGE [4]	98.6	98.0	97.0	97.5	89.0	89.6	88.6	89.1	70.9	66.5	68.7	67.6
TNN [21]	98.6	98.4	96.4	97.4	93.6	93.0	93.6	93.3	82.0	76.9	79.7	78.2
<i>(C) CNNs</i>												
SGNet [14]	96.8	96.5	95.8	96.1	83.2	76.6	75.1	74.8	—	—	—	—
<i>(D) Conventional</i>												
Kitaoka et al. [6]	92.5	92.2	91.3	91.8	78.7	77.3	77.4	77.3	55.9	54.2	53.6	53.9
Tschirren et al. [16]	92.8	92.6	92.2	92.4	79.5	78.6	77.5	78.0	57.1	54.5	56.1	55.4
Feragen et al. [2]	94.1	94.1	93.6	93.8	81.8	80.1	83.2	81.7	62.8	60.2	56.2	58.1
<i>(E) Transformers</i>												
Transformer [17]	98.8	98.5	97.6	98.0	93.1	93.0	92.8	92.9	82.5	77.8	80.8	79.3
Ours	99.2	98.8	98.2	98.5	95.7	95.9	95.9	95.9	86.0	83.5	84.5	84.0

**Fig. 3.** The results of airway labeling are displayed from the perspective of graph and 3D. The first (last) two rows are the results of segmental (subsegmental) bronchi.

3.2 Evaluations

Table 1 compares the quantitative results of our proposed method with other methods. GNNs methods such as GCN [5] and GAT [18] only aggregate information from neighborhoods, while DAGNN [15] gradually performs message passing and updating from root nodes to the end. These methods merely encode the local structure of airway trees and are seriously disturbed by distribution overlapping in subsegments. Thus, they fail to achieve satisfactory performance, especially in subsegmental level. In HGNN methods, HyperGCN [19] and UniSAGE [4] use hyperedges to represent subtrees. TNN [21] further proposes a sub-network for subtrees to exchange information and outperforms other HGNNs methods. SGNet [14] uses CNNs to conduct semantic segmentation directly on CT images, which can be easily affected by class imbalance. Conventional methods [2, 6, 16] usually adopt graph-matching-based algorithms, but the reference tree cannot meet all situations due to individual variations. Thus, their performance drops a lot in the subsegmental level. Vanilla Transformer [17] achieves the same effect as TNN owing to the global-scale self-attention module while lacking the utilization of structural prior. Compared to all the above methods, our approach improves the accuracy by more than 2% and 4% respectively in segmental and subsegmental levels, demonstrating the effectiveness of the proposed method.

Qualitative results are displayed in Fig. 3 to demonstrate the superior performance of our model at different levels. In segmental labeling, transformer lacks the local structural prior, in which way the sibling categories cannot be classified well. The problem is more severe in subsegments. TNN adopts a subtree interaction module to learn relative information, but the correspondence between different nomenclature level is not fully utilized. By contrast, our method can perform accurate airway labeling both in segmental and subsegmental levels.

Table 2. Ablation study of key components (%). U denotes the U-shape layout, NIE is neighborhood information encoding module, and G is the weight generator network.

Components			Lobar				Segmental				Subsegmental			
U	NIE	G	ACC	PR	RC	F1	ACC	PR	RC	F1	ACC	PR	RC	F1
×	×	×	98.8	98.5	97.6	98.0	93.1	93.0	92.8	92.9	82.5	77.8	80.8	79.3
✓	×	×	99.1	98.5	98.0	98.2	94.7	94.8	94.7	94.7	83.6	80.7	82.2	81.4
×	✓	×	99.0	98.2	98.0	98.1	94.6	94.5	94.1	94.3	84.8	81.6	82.7	82.2
✓	✓	×	99.0	98.6	97.9	98.3	95.4	95.6	95.6	95.6	85.4	83.0	84.1	83.6
✓	×	✓	99.1	98.6	98.0	98.3	95.0	95.2	94.9	95.0	84.3	81.8	83.0	82.4
✓	✓	✓	99.2	98.8	98.2	98.5	95.7	95.9	95.9	95.9	86.0	83.5	84.5	84.0

We further conducted ablation studies to verify the effectiveness of each component of AirwayFormer. Table 2 shows the results. Vanilla transformer performing bronchi labeling for each level independently is used as the baseline. The U-shape layout encodes the correspondence of nomenclature and improves the

results in all three levels. NIE module introduces the local structure prior to self-attention calculation and also promotes the labeling performance, especially in the subsegmental level. Combining the two modules, the performance is further ameliorated. The weight generator network dynamically adjusts the classification weights and improves the subsegmental results. When applying the three modules simultaneously, the network demonstrates the most powerful ability for airway labeling. The experimental results indicate that these proposed modules do contribute to the satisfactory performance of our method.

4 Conclusion

This paper presented a transformer-based method named AirwayFormer for airway anatomical labeling. A U-shape layout integrating graph information is used to exploit the latent relationships fully. Meanwhile, a weight generator that produces the dynamic decision boundaries is designed to capture the relative relationships between sibling categories. Extensive experiments showed that our proposed method achieved superior performance in the bronchi labeling of all three levels, leading to an efficient clinical tool for intra-operative navigation.

References

1. Dwivedi, V.P., Bresson, X.: A generalization of transformer networks to graphs. arXiv preprint [arXiv:2012.09699](https://arxiv.org/abs/2012.09699) (2020)
2. Feragen, A., et al.: Geodesic atlas-based labeling of anatomical trees: application and evaluation on airways extracted from CT. *IEEE Trans. Med. Imaging* **34**(6), 1212–1226 (2014)
3. Hendrycks, D., Gimpel, K.: Gaussian error linear units (GELUs). arXiv preprint [arXiv:1606.08415](https://arxiv.org/abs/1606.08415) (2016)
4. Huang, J., Yang, J.: Unignn: a unified framework for graph and hypergraph neural networks. arXiv preprint [arXiv:2105.00956](https://arxiv.org/abs/2105.00956) (2021)
5. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint [arXiv:1609.02907](https://arxiv.org/abs/1609.02907) (2016)
6. Kitaoka, H., et al.: Automated nomenclature labeling of the bronchial tree in 3D-CT lung images. In: Dohi, T., Kikinis, R. (eds.) *MICCAI 2002*. LNCS, vol. 2489, pp. 1–11. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45787-9_1
7. Lo, P., van Rikxoort, E.M., Goldin, J., Abtin, F., de Bruijne, M., Brown, M.: A bottom-up approach for labeling of human airway trees. *MICCAI Int. WS. Pulm. Im. Anal.* (2011)
8. Min, E., et al.: Transformer for graphs: an overview from architecture perspective. arXiv preprint [arXiv:2202.08455](https://arxiv.org/abs/2202.08455) (2022)
9. Mori, K., Hasegawa, J.I., Suenaga, Y., Toriwaki, J.I.: Automated anatomical labeling of the bronchial branch and its application to the virtual bronchoscopy system. *IEEE Trans. Med. Imaging* **19**(2), 103–114 (2000)
10. Mori, K., et al.: Automated anatomical labeling of bronchial branches extracted from CT datasets based on machine learning and combination optimization and its application to bronchoscope guidance. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 707–714. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04271-3_86

11. Netter, F.H.: Atlas of human anatomy, Professional Edition E-Book: including NetterReference. com Access with full downloadable image Bank. Elsevier health sciences (2014)
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
13. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
14. Tan, Z., Feng, J., Zhou, J.: SGNet: structure-aware graph-based network for airway semantic segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 153–163. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_15
15. Thost, V., Chen, J.: Directed acyclic graph neural networks. arXiv preprint [arXiv:2101.07965](https://arxiv.org/abs/2101.07965) (2021)
16. Tschirren, J., McLennan, G., Palágyi, K., Hoffman, E.A., Sonka, M.: Matching and anatomical labeling of human airway tree. IEEE Trans. Med. Imaging **24**(12), 1540–1547 (2005)
17. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
18. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint [arXiv:1710.10903](https://arxiv.org/abs/1710.10903) (2017)
19. Yadati, N., Nimishakavi, M., Yadav, P., Nitin, V., Louis, A., Talukdar, P.: Hyper-gcn: a new method for training graph convolutional networks on hypergraphs. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
20. Ying, C., et al.: Do transformers really perform badly for graph representation? Adv. Neural. Inf. Process. Syst. **34**, 28877–28888 (2021)
21. Yu, W., et al.: TNN: tree neural network for airway anatomical labeling. IEEE Trans. Med. Imaging **42**(1), 103–118 (2022)
22. Zhao, T., Yin, Z., Wang, J., Gao, D., Chen, Y., Mao, Y.: Bronchus segmentation and classification by neural networks and linear programming. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 230–239. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_26