AIRWAYNET-SE: A SIMPLE-YET-EFFECTIVE APPROACH TO IMPROVE AIRWAY SEGMENTATION USING CONTEXT SCALE FUSION

Yulei Qin, Yun Gu, Hao Zheng, Mingjian Chen, Jie Yang, Yue-Min Zhu

AIRWAYNET-SE: A SIMPLE-YET-EFFECTIVE APPROACH TO IMPROVE AIRWAY SEGMENTATION USING CONTEXT SCALE FUSION

Yulei Qin[1,2,3], *Yun Gu*[1,2*], *Hao Zheng*[1,2], *Mingjian Chen*[1,2], *Jie Yang*[1,2*], *Yue-Min Zhu*[3]

[1]Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China
[2]Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China
[3]Université de Lyon, INSA Lyon, CNRS, INSERM, CREATIS UMR 5220, U1206, Lyon, France

ABSTRACT

Accurate segmentation of airways from chest CT scans is crucial for pulmonary disease diagnosis and surgical navigation. However, the intra-class variety of airways and their intrinsic tree-like structure pose challenges to the development of automatic segmentation methods. Previous work that exploits convolutional neural networks (CNNs) does not take context scales into consideration, leading to performance degradation on peripheral bronchiole. We propose the two-step AirwayNet-SE, a Simple-yet-Effective CNNs-based approach to improve airway segmentation. The first step is to adopt connectivity modeling to transform the binary segmentation task into 26-connectivity prediction task, facilitating the model's comprehension of airway anatomy. The second step is to predict connectivity with a two-stage CNNs-based approach. In the first stage, a Deep-yet-Narrow Network (DNN) and a Shallow-yet-Wide Network (SWN) are respectively utilized to learn features with large-scale and small-scale context knowledge. These two features are fused in the second stage to predict each voxel's probability of being airway and its connectivity relationship between neighbors. We trained our model on 50 CT scans from public datasets and tested on another 20 scans. Compared with state-of-the-art airway segmentation methods, the robustness and superiority of the AirwayNet-SE confirmed the effectiveness of large-scale and small-scale context fusion. In addition, we released our manual airway annotations of 60 CT scans from public datasets for supervised airway segmentation study.

Index Terms— Airway segmentation, context scale, voxel connectivity, convolutional neural networks

1. INTRODUCTION

Segmentation of airway tree from chest computed tomography (CT) scans is indispensable for both pulmonary disease diagnosis and endobronchial navigation. Manual segmentation of airway tree is time-consuming due to its complex tree-like structure with varying airway lumen size and lumen-wall intensity contrast. To relieve radiologists from such strenuous work, airway segmentation methods have been proposed [1, 2, 3, 4]. Most of them used region growing or filtering-based techniques for tubular structure enhancement. The evaluation results in [4] revealed that the performance of these methods degraded drastically on detecting small and thin peripheral bronchi. Besides, a large amount of leakage was observed when the intensity contrast between airway wall and lumen was low. More recently, convolutional neural networks (CNNs)-based methods have been explored for airway extraction [5, 6, 7, 8, 9]. They adopted U-Net [10] or 3D U-Net [11] as their backbone for segmentation in either a volume-of-interest (VOI) tracking way or a sliding window manner. Moreover, graph-based methods such as fuzzy connectedness segmentation (FCS) [7] and graph neural networks (GNNs) [12] have been combined with CNNs for post-refinement.

Although deep learning approaches achieved superior performance, there still remain challenges to be solved. First, the intensity distribution of airways is quite different among trachea, primary bronchus, secondary bronchus, and peripheral bronchiole (see Fig. 1). The intensity contrast between lumen and wall is clear at trachea regions, but becomes weaker as the airway bifurcates into smaller branches. The airway wall is much thinner and darker at bronchiole regions. Second, the scale of context is dissimilar for segmentation on large and small bronchi. The context refers to the feature information that describes the mutual relationship between airways and background. To extract trachea and main bronchus, large-scale context is preferred for the model to perceive the main body with large field-of-view. On the contrary, for segmenting bronchiole, context of close neighborhood is enough. Third, for CNNs architecture, the number of pooling layers requires careful design. Features of thin bronchi, whose diameters are usually only 2-3 voxels, are prone to vanish after three times of pooling, making it difficult to reconstruct and recover. However, for large bronchi, multiple pooling layers are necessary to extract effective context. Furthermore, pub-

---

lic datasets with airway annotations are unavailable for model training and fair comparison between different methods.

(a) (b) (c) (d)
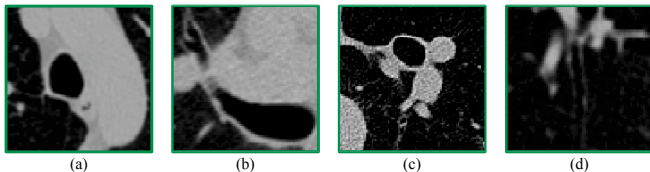
Fig. 1. The intensity distribution of trachea (a), primary (b) and secondary (c) bronchus, and peripheral bronchiole (d). The scale of contexts needed for airway segmentation on (a)-(d) is decreasing from large to small.

To address these gaps, we propose the two-step AirwayNet-SE, a **S**imple-yet-**E**ffective approach that incorporates two different scales of context to comprehend large and small airways, respectively. With 3-D connectivity modeling [9] in the first step, the networks are trained to predict whether a voxel is connected to its neighbors instead of directly classifying airway voxels. The AirwayNet-SE consists of one **d**eep-yet-**n**arrow **n**etwork (DNN) and one **s**hallow-yet-**w**ide **n**etwork (SWN). The DNN, with deeper layers yet smaller number of channels per layer, aims at extracting features of thick branches. Four pooling operations are used for the model to be aware of the overall context of thoracic cavity. While for the SWN, shallower layers with two pooling operations are adopted to prevent thin bronchi from vanishing. The feature channels of SWN are widened to increase representation power. The second step is to predict connectivity using two-stage CNNs. In the first stage, we respectively train our DNN and SWN to learn effective features of large and small bronchi. In the second stage, features from both DNN and SWN are concatenated as the fusion of context knowledge from two scales. Such fused features are utilised for all airway connectivity prediction. Our contributions are summarised as follows: 1) The AirwayNet-SE proposed a solution to the conflict caused by the difference between large and small airways. With connectivity modeling, it leveraged the fusion of context knowledge from two scales to predict whether a voxel is airway and connects to its neighbors. Our method achieved the state-of-the-art Dice coefficient of 93.0% on 10 public scans and 88.7% on 10 private scans, respectively. 2) We released the manual annotations of 60 public CT scans to promote airway segmentation study that requires supervised learning. To the best of our knowledge, this is the largest publicly available dataset of airway annotations.

2. DATA

The experiment dataset contains 70 clinical thoracic CT scans in total, with 60 public CT scans and 10 privately collected CT scans. The acquisition and investigation of data were conformed to the principles outlined in the declaration of Helsinki [13]. We used 20 scans from the training set of EXACT'09 [4] and 40 scans from LIDC-IDRI [14]. The EXACT'09 only provides raw CT images without airway annotation. The LIDC-IDRI (under Creative Commons Attribution 3.0 Unported License) includes 1018 scans with pulmonary nodule annotations. In view of image quality, 40 scans whose slice thickness ≤ 0.625 mm are randomly chosen. The 10 private CT scans were obtained from patients with severe lung diseases such as emphysema and pneumonia. The axial slices of all scans have the same size of 512×512, with their spatial resolution in the range of 0.5–0.781 mm. Their slice thickness varies from 0.45 to 1.0 mm.

For each CT scan, the ground-truth annotation of airway lumen was obtained by: 1) using an interactive segmentation method to generate a rough airway tree via ITK-SNAP [15]; 2) manual correction and delineation by well-trained experts. The annotations of 60 public CT scans are released to promote further study of airway extraction using supervised learning methods[1]. However, the 10 private CT scans and annotations will not be made available online at the moment.

3. METHOD

The first step is to model each airway voxel's connectivity. The second step is to predict connectivity using two stages as follows: 1) Feature learning with large-scale and small-scale contexts, 2) Connectivity prediction based on fused contexts.

3.1. CT pre-processing

To avoid feature learning from irrelevant marginal area, a lung mask extraction technique is employed to exclude invalid regions. We first smooth each CT scan with a Gaussian filter and binarise it with OTSU method. The connected component analysis is performed to keep the largest two components and remove holes. The convex hull of the components is then computed to avoid under-segmentation. Finally, the voxel intensity is truncated inside the window of [-1000, 600] (HU) and normalised to [0, 1]. To reduce GPU cost, the bounding box of lung field is extracted and cropped into overlapping cubes using a sliding window technique. The size of each cube is $32 \times 240 \times 240$ and the sliding stride is [8, 64, 64].

3.2. Step 1: Connectivity modeling using binary labels

Given a voxel $P = (x, y, z)$ in a 3-D volume of binary airway label, its direct neighbor $Q = (u, v, w)$ can be defined by: $d(P, Q) = max(|(x - u)|, |(y - v)|, |(z - w)|) \leq 1$. Since each voxel has 26 direct neighbors, we index those neighbors from 1 to 26 and denote each pair $(P, Q_i), i \in \{1, 2, ..., 26\}$ as a connectivity orientation (see Fig. 2). We use 26 binary labels to represent the corresponding 26 connectivity pairs. For example, if P and its right neighbor Q_{14} are connected

---

[1] Annotations will be available at http://www.pami.sjtu.edu.cn/

airway voxels, then the position P on the 14-th connectivity label is marked as "1". Otherwise, we mark the position P as "0" on the i-th connectivity label if P is disconnected to the neighbor Q_i. After connectivity modeling, the original binary airway label is transformed into 26 connectivity labels of the same size. We concatenate them along the channel axis into a 26-channel 3-D label. The task of binary airway segmentation is therefore converted to the prediction of airway connectivity. The model trained using such a 26-channel label is believed to perform better in grasping the intrinsic tree-like structure of airway than using the original binary label.
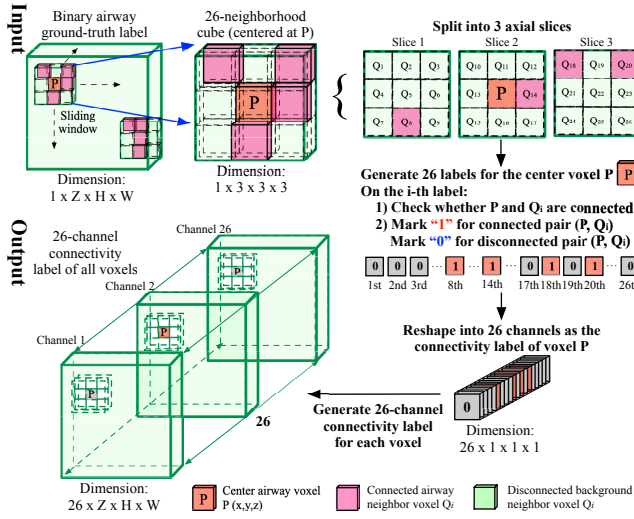

Fig. 2. Illustration of the 26-connectivity modeling.

3.3. Step 2: Connectivity prediction using 3-D CNNs with fused context scale

The illustration of the AirwayNet-SE for connectivity prediction is shown in Fig. 3. It consists of two stages as follows.

Stage 1: Feature learning with large-scale and small-scale contexts. In this stage, we respectively employ the DNN and SWN to learn features of airway connectivity with different context scales. The 3-D U-Net [11], containing a contracting path and an expansive path, is used in both DNN and SWN as backbone. To enlarge the receptive field of DNN, four pooling layers are used and accordingly ten convolution layers (Conv) with batch normalization (BN) and rectified linear unit (ReLU) are set on the contracting path. For SWN, only two pooling layers are kept to preserve the details of "delicate" and thin bronchi. The number of feature channels in DNN and SWN are separately designed to fit for such architecture difference. The trade-off between feature extraction ability and GPU memory limit is considered as well.

Stage 2: Connectivity prediction based on fused context knowledge. In this stage, feature representations from
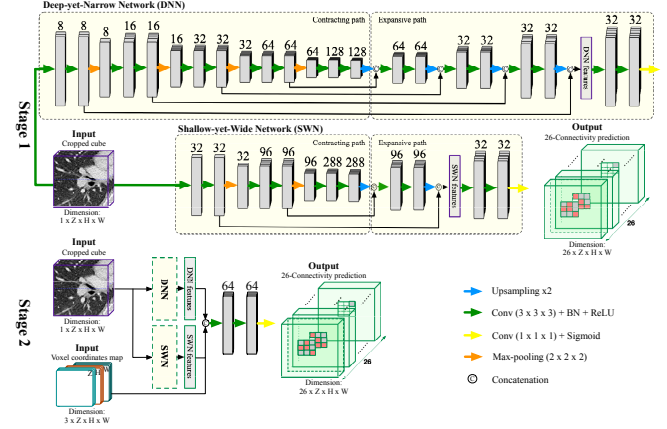

Fig. 3. Illustration of the AirwayNet-SE. The channel number is denoted on each feature map. The first stage is to extract features of two context scales via DNN and SWN. The second stage is to classify the connectivity of airways using fused features with both large-scale and small-scale contexts.

DNN and SWN are concatenated as context knowledge fusion. The voxel coordinates are also included as inputs. They are considered beneficial for the model to comprehend the anatomical structure of airways because the position of airways within the thoracic cavity is not randomly distributed. We build a simple three-layer CNN to learn the mapping from the fused features to the 26-channel connectivity prediction.

Optimization: To train the proposed AirwayNet-SE, the Dice coefficient loss is used in both the two stages because it generally performs better than binary cross-entropy loss in the segmentation task. Given a voxel x in the cropped cube X, its i-th channel of the connectivity label $y_i(x)$, $i \in \{1, 2, ..., 26\}$ and the corresponding prediction output $p_i(x)$, the averaged Dice loss is defined as:

$$\mathcal{L} = 1 - \frac{1}{26} \sum_{i=1}^{26} \frac{2 \sum_{x \in X} p_i(x) y_i(x)}{\sum_{x \in X} (p_i(x) + y_i(x)) + \epsilon}, \quad (1)$$

where ϵ is used to avoid division by zero.

3.4. Airway candidate generation

The predicted airway connectivity needs to be transformed back into conventional binary airway candidates. First, we apply a thresholding $t = 0.5$ on the probability output. The pairwise adjacent voxels in each $3 \times 3 \times 3$ neighborhood should agree with each other in connectivity. Therefore, we only keep voxels that connect each other in mutual connectivity channels. Then, we perform channel-wise summation to obtain the 1-channel airway candidate output and multiply it with the lung field mask to generate valid airway regions. No post-processing is employed for refinement.

4. EXPERIMENTS AND RESULTS

In view of the dataset under investigation, we randomly chose 50 public CT scans for training and hyper-parameter fine-tuning. The remaining 10 public and 10 private scans were used for evaluation. Performance comparison and ablation studies are conducted to confirm the validity of our method.

4.1. Implementation details

To improve the model's generalizability on diverse CT scans, data augmentation is performed on-the-fly during training via random horizontal flipping and Gaussian smoothing. In the first stage, we densely sampled cubes near trachea and main bronchus regions for training DNN. The cubes containing thin peripheral bronchiole are mostly chosen for SWN to learn their complicated branching patterns. In the second stage, all cubes near airways are used for network training, resulting in around 6000 samples. We implemented our method in Keras and adopted Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$) with learning rate set as 10^{-4}. The training converged after 30 epochs for networks in stages 1 and 2.

4.2. Evaluation metrics and results

The performance of the proposed AirwayNet-SE was evaluated in terms of three metrics: 1) Dice coefficient (DSC), 2) True positive rate (TPR), and 3) False positive rate (FPR). As shown in Table 1, the proposed AirwayNet-SE achieved the highest DSC of 93.0% and 88.7% on both the public and private testing sets, respectively. Compared to DNN and SWN that only rely on context of single scale for airway extraction, the AirwayNet-SE increased the DSC over 2% on average, respectively, demonstrating the effectiveness of feature fusion from two context scales. The comparison with state-of-the-art methods [7, 8, 9] was conducted. These methods were implemented by ourselves and fine-tuned on the current dataset. For the public testing set, Juarez et al. [8] segmented more conservatively and they had a lower FPR and a competitive DSC. But our TPR is 3% higher than theirs with smaller standard deviation, meaning that more airway voxels can be detected by the AirwayNet-SE. Qin et al. [9] achieved a higher TPR than ours and higher false positives are observed on both two testing sets. Since DSC is a comprehensive metric that considers both sensitivity and specificity, the highest DSC of our method with comparable TPR and FPR verified its robustness and superiority. Qualitative comparison of the segmentation results is visualized in Fig. 4. Compared to SWN and DNN, more peripheral branches were successfully detected by AirwayNet-SE and there exists less discontinuity in the predicted airway regions. Furthermore, SWN performed worse on thick bronchi due to its relatively small scale of context. In contrast, the prediction of DNN was prone to missing thin bronchiole and produce ruptures. With the fused context knowledge of both small and large scales, our AirwayNet-SE

Table 1. Results (%) of the proposed AirwayNet-SE in comparison with state-of-the-art methods and ablation study (mean±standard deviation). The DNN, SWN, and AN-SE stand for deep-yet-narrow network, shallow-yet-wide network, and AirwayNet-SE (AN-SE), respectively.

Public testing set			
Method	DSC	TPR	FPR
Jin et al. [7]	90.5±4.0	94.7±2.6	0.044±0.029
Juarez et al. [8]	92.8±3.5	89.2±6.5	**0.008±0.004**
Qin et al. [9]	90.9±4.3	92.7±3.5	0.033±0.027
DNN	90.1±3.8	93.3±3.3	0.041±0.024
SWN	89.5±5.4	**95.4±2.4**	0.055±0.043
AN-SE	**93.0±3.1**	92.4±4.0	0.018±0.012
Private testing set			
Method	DSC	TPR	FPR
Jin et al. [7]	86.3±5.4	81.7±8.9	0.027±0.022
Juarez et al. [8]	87.2±5.3	81.0±8.8	**0.016±0.021**
Qin et al. [9]	88.5±4.0	**86.5±6.0**	0.033±0.029
DNN	86.9±6.8	82.0±11.2	0.022±0.021
SWN	87.4±4.7	83.4±7.6	0.027±0.023
AN-SE	**88.7±5.3**	84.6±8.7	0.020±0.017

avoided the aforementioned limit of single-scale context and achieved more accurate segmentation results.
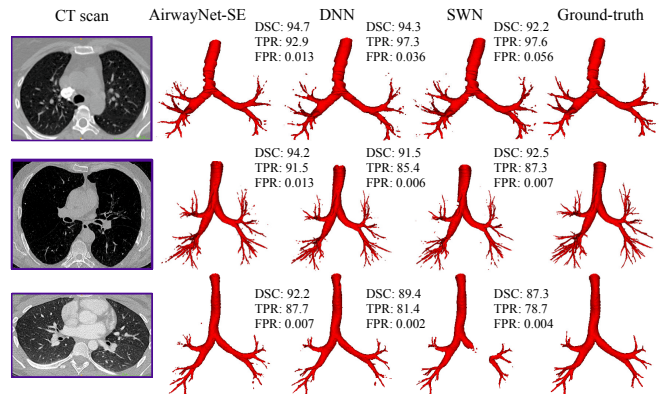
Fig. 4. Comparison of airway segmentation results between the AirwayNet-SE, DNN, SWN, and ground truth.

5. CONCLUSION

This paper introduced a two-step AirwayNet-SE for airway segmentation. With connectivity modeling, it fused context of two scales to improve performance. Experimental results proved that our approach was effective at overcoming the distribution difference between large and small airways. The airway annotations were also released to boost research on airway extraction using supervised learning methods.

6. REFERENCES

[1] Eva M Van Rikxoort, Wouter Baggerman, and Bram van Ginneken, "Automatic segmentation of the airway tree from thoracic CT scans using a multi-threshold approach," in *Proc. of Second International Workshop on Pulmonary Image Analysis*, 2009, pp. 341–349.

[2] Ziyue Xu, Ulas Bagci, Brent Foster, Awais Mansoor, Jayaram K Udupa, and Daniel J Mollura, "A hybrid method for airway segmentation and automated measurement of bronchial wall thickness on CT," *MedIA*, vol. 24, no. 1, pp. 1–17, 2015.

[3] Pechin Lo, Jon Sporring, Haseem Ashraf, Jesper JH Pedersen, and Marleen de Bruijne, "Vessel-guided airway tree segmentation: A voxel classification approach," *MedIA*, vol. 14, no. 4, pp. 527–538, 2010.

[4] Pechin Lo, Bram Van Ginneken, Joseph M Reinhardt, Tarunashree Yavarna, Pim A De Jong, Benjamin Irving, Catalin Fetita, Margarete Ortner, Rômulo Pinho, Jan Sijbers, et al., "Extraction of airways from CT (exact'09)," *IEEE TMI*, vol. 31, no. 11, pp. 2093–2107, 2012.

[5] Jean-Paul Charbonnier, Eva M Van Rikxoort, Arnaud AA Setio, Cornelia M Schaefer-Prokop, Bram van Ginneken, and Francesco Ciompi, "Improving airway segmentation in computed tomography using leak detection with convolutional networks," *MedIA*, vol. 36, pp. 52–60, 2017.

[6] Qier Meng, Holger R Roth, Takayuki Kitasaka, Masahiro Oda, Junji Ueno, and Kensaku Mori, "Tracking and segmentation of the airways in chest CT using a fully convolutional network," in *MICCAI*. Springer, 2017, pp. 198–207.

[7] Dakai Jin, Ziyue Xu, Adam P Harrison, Kevin George, and Daniel J Mollura, "3D convolutional neural networks with graph refinement for airway segmentation using incomplete data labels," in *MLMI*. Springer, 2017, pp. 141–149.

[8] A Garcia-Uceda Juarez, HAWM Tiddens, and M de Bruijne, "Automatic airway segmentation in chest CT using convolutional neural networks," in *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pp. 238–250. Springer, 2018.

[9] Yulei Qin, Mingjian Chen, Hao Zheng, Yun Gu, Mali Shen, Jie Yang, Xiaolin Huang, Yue-Min Zhu, and Guang-Zhong Yang, "Airwaynet: A voxel-connectivity aware approach for accurate airway segmentation using convolutional neural networks," *arXiv e-prints*, 2019.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.

[11] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *MICCAI*. Springer, 2016, pp. 424–432.

[12] Raghavendra Selvan, Thomas Kipf, Max Welling, Jesper H. Pedersen, Jens Petersen, and Marleen de Bruijne, "Extraction of airways using graph neural networks," *arXiv e-prints*, 2018.

[13] World Medical Association et al., "World medical association declaration of helsinki. ethical principles for medical research involving human subjects.," *Bulletin of the World Health Organization*, vol. 79, no. 4, pp. 373, 2001.

[14] G Samuel III, "Armato iii, geoffrey mclennan, luc bidaut, michael f mcnitt-gray, charles r meyer, anthony p reeves, binsheng zhao, denise r aberle, claudia i henschke, eric a hoffman, et al. the lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans," *Medical physics*, vol. 38, no. 2, pp. 915–931, 2011.

[15] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.