



**HAL**  
open science

# Learning Bronchiole-Sensitive Airway Segmentation CNNs by Feature Recalibration and Attention Distillation

Yulei Qin, Hao Zheng, Yun Gu, Xiaolin Huang, Jie Yang, Lihui Wang,  
Yue-Min Zhu

► **To cite this version:**

Yulei Qin, Hao Zheng, Yun Gu, Xiaolin Huang, Jie Yang, et al.. Learning Bronchiole-Sensitive Airway Segmentation CNNs by Feature Recalibration and Attention Distillation. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, Oct 2020, Lima, Peru. pp.221-231, 10.1007/978-3-030-59710-8\_22 . hal-03435078

**HAL Id: hal-03435078**

**<https://hal.science/hal-03435078v1>**

Submitted on 18 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Learning Bronchiole-Sensitive Airway Segmentation CNNs by Feature Recalibration and Attention Distillation\*

Yulei Qin<sup>1,2,4</sup>, Hao Zheng<sup>1,2</sup>, Yun Gu<sup>1,2</sup>, Xiaolin Huang<sup>1,2</sup>, Jie Yang<sup>1,2</sup>, Lihui Wang<sup>3</sup>, and Yue-Min Zhu<sup>4</sup>

<sup>1</sup> Institute of Image Processing and Pattern Recognition,  
Shanghai Jiao Tong University, Shanghai, China  
{jieyang, geron762}@sjtu.edu.cn

<sup>2</sup> Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China

<sup>3</sup> Key Laboratory of Intelligent Medical Image Analysis and Precise Diagnosis of Guizhou Province, College of Computer Science and Technology, Guizhou University, Guiyang, China

<sup>4</sup> UdL, INSA Lyon, CREATIS, CNRS UMR 5220, INSERM U1206, Lyon, France

**Abstract.** Training deep convolutional neural networks (CNNs) for airway segmentation is challenging due to the sparse supervisory signals caused by severe class imbalance between long, thin airways and background. In view of the intricate pattern of tree-like airways, the segmentation model should pay extra attention to the morphology and distribution characteristics of airways. We propose a CNNs-based airway segmentation method that enjoys superior sensitivity to tenuous peripheral bronchioles. We first present a feature recalibration module to make the best use of learned features. Spatial information of features is properly integrated to retain relative priority of activated regions, which benefits the subsequent channel-wise recalibration. Then, attention distillation module is introduced to reinforce the airway-specific representation learning. High-resolution attention maps with fine airway details are passing down from late layers to previous layers iteratively to enrich context knowledge. Extensive experiments demonstrate considerable performance gain brought by the two proposed modules. Compared with state-of-the-art methods, our method extracted much more branches while maintaining competitive overall segmentation performance.

**Keywords:** Airway segmentation · Recalibration · Distillation

---

\* Corresponding author: Jie Yang and Yun Gu. This research was partly supported by National Key R&D Program of China (No. 2019YFB1311503), Committee of Science and Technology, Shanghai, China (No. 19510711200), National Natural Science Foundation of China (No. 61977046), International Research Project METISLAB, and Program PHC-Cai Yuanpei 2018 (No. 41400TC).

## 1 Introduction

Extraction of airways from computed tomography (CT) is critical for quantifying the morphological changes of chest (e.g., bronchial stenosis) and thereby indicating the presence and stage of related diseases. The intrinsic tree-like structure of airways demands great efforts for manual delineation and therefore several automatic airway segmentation methods have been proposed [11, 12, 19, 21]. Most of them employed techniques like thresholding, region growing, and tubular structure enhancing. Since the intensity contrast between airway wall and lumen weakens as airways split into thinner branches, these methods often failed to extract peripheral bronchioles and produced leakage outside wall. Recent progress of convolutional neural networks (CNNs) [4, 5, 14, 17] has spawned research on airway segmentation using CNNs [3, 7–9, 13, 15, 18, 20, 22, 25]. 2-D and 2.5-D CNNs [3, 22] were adopted to refine the initial coarsely segmented bronchi. 3-D CNNs were developed for direct airway extraction on CT volume in either an optimized tracking way [13, 25] or a sliding window way [8]. Wang et al. [20] proposed recurrent convolution and radial distance loss to capture airways’ topology. Qin et al. [15] transformed the airway segmentation task into connectivity prediction task for inherent structure comprehension. Graph neural networks [9, 18] were also explored to incorporate nodes’ neighborhood knowledge.

Despite the improved performance by deep learning, there still remain limitations to be overcome. First, severe class imbalance between airways and background poses a threat to the training of 3-D CNNs. Most of the current CNNs heavily rely on airway ground-truth as supervisory signals. Unlike bulky or spheroid-like organs (e.g., liver and kidney), tree-like airways are thin, tenuous, and divergent. It is difficult to train deep models using such sparse and scattered target labels. Although weighted cross-entropy loss was proposed to focus on positive samples, single source of supervisory signals from deficient airway labels still makes optimization ineffective. Second, the characteristics of airways require the model to utilize both global-scale and local-scale context knowledge to perceive the main body (trachea and main bronchus) and limbs (peripheral bronchi). Previous models used 2 or 3 pooling layers and the coarsest resolution features provide limited long-range context. If more layers are simply piled up, the increased parameters may cause over-fitting due to inadequate training data.

To address these two concerns, we present a CNNs-based airway segmentation method with high sensitivity to peripheral bronchioles. Our contributions are threefold: 1) The recalibration module is proposed to maximally utilize the learned features. On one hand, to avoid over-fitting of deep CNNs, the number of feature channels is reduced to limit model complexity. On the other hand, we do not expect the model’s learning capacity to diminish because of such reduction. Under this circumstance, feature recalibration seems to be a reasonable solution. We hypothesize that spatial information of features is indispensable for channel-wise recalibration and should be treated differently from position to position, layer to layer. The average pooling used in [16, 26] for spatial knowledge compression may not well capture the location of various airways in different resolution scales. In contrast, we aim at prioritizing information at key

positions with learnable weights, which provides appropriate spatial hints to model inter-channel dependency and thereafter improves recalibration. 2) The attention distillation module is introduced to reinforce representation learning of airway structure. Attention maps of different features enable us to potentially reveal the morphology and distribution pattern of airways. Inspired by knowledge distillation [6, 24], we refine the attention maps of lower resolution by mimicking those of higher resolution. Finer attention maps (teacher’s role) with richer context can cram coarser ones (student’s role) with details about airways and lung contours. The model’s ability to recognize delicate bronchioles is ameliorated after iteratively focusing on the anatomy of airways. In addition, the distillation itself acts as auxiliary supervision that provides extra supervisory signals to assist training. 3) With extensive experiments on 90 clinical 3-D CT scans, our method achieved superior sensitivity to extraction of thin airways and maintained competitive overall segmentation performance.

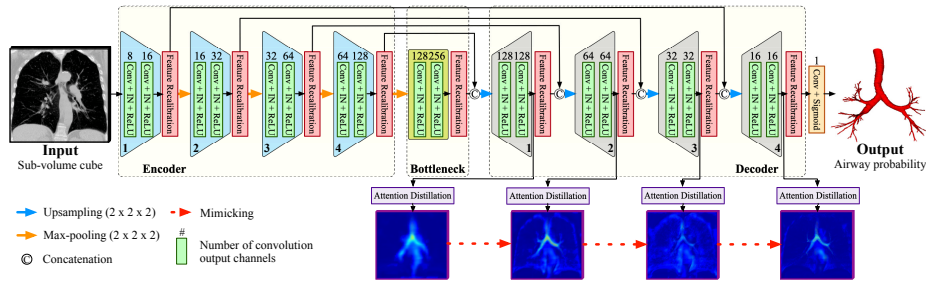


Fig. 1. Overview of the proposed method for airway segmentation.

## 2 Method

In this section, we present the design of the above-mentioned two modules: feature recalibration and attention distillation. The overview of the proposed method is illustrated in Fig. 1. For airway segmentation, the CNNs-based prediction can be formulated as  $P = \mathcal{F}(X)$ , where  $P$  denotes the predicted probability of airway at each voxel  $x$  of the input volume  $X$ . Given the airway label  $Y$ , the objective is to learn an end-to-end mapping  $\mathcal{F}$  that minimizes the difference between  $P$  and  $Y$  by CNNs. Assuming the segmentation model has  $M$  convolution layers in total, we denote the activation output of the  $m$ -th convolution feature as  $A_m \in \mathbb{R}^{C_m \times D_m \times H_m \times W_m}$ ,  $1 \leq m \leq M$ . The number of its channel, depth, height, and width are respectively denoted as  $C_m$ ,  $D_m$ ,  $H_m$ , and  $W_m$ .

**Feature Recalibration:** We propose the mapping block  $\mathcal{Z}(\cdot)$  that generates a channel descriptor  $U_m$  to recalibrate the learned feature  $A_m$ . The recalibration by multiplying  $U_m$  with  $A_m$  unearths and strengthens basis channels that affect most the output decision. Unlike previous recalibration methods [16, 26],

we integrate spatial information using weighted combination of features along each spatial dimension. Our hypothesis is that different positions may hold different degree of importance both within  $A_m$  and across resolution scales (e.g.,  $A_{m-1}, A_{m+1}$ ). The operation like adaptive or global pooling is not spatially discriminating for the finest features (containing thin bronchioles that are easily “erased” by averaging) and the coarsest features (containing mostly thick bronchi). Therefore, we introduce the following spatial integration method  $\mathcal{Z}_{spatial}(\cdot)$  that preserves relatively important regions. It can be formulated as:

$$\mathcal{Z}_{spatial}(A_m) = \mathcal{B}(\mathcal{Z}_{spatial}^{Depth}(A_m)) + \mathcal{B}(\mathcal{Z}_{spatial}^{Height}(A_m)) + \mathcal{B}(\mathcal{Z}_{spatial}^{Width}(A_m)), \quad (1)$$

$$\mathcal{Z}_{spatial}^{Depth}(A_m) = \sum_{j=1}^{H_m} h_j \sum_{k=1}^{W_m} w_k A_m[:, :, j, k], \mathcal{Z}_{spatial}^{Depth}(A_m) \in R^{C_m \times D_m \times 1 \times 1}, \quad (2)$$

$$\mathcal{Z}_{spatial}^{Height}(A_m) = \sum_{i=1}^{D_m} d_i \sum_{k=1}^{W_m} w_k A_m[:, i, :, k], \mathcal{Z}_{spatial}^{Height}(A_m) \in R^{C_m \times 1 \times H_m \times 1}, \quad (3)$$

$$\mathcal{Z}_{spatial}^{Width}(A_m) = \sum_{i=1}^{D_m} d_i \sum_{j=1}^{H_m} h_j A_m[:, i, j, :], \mathcal{Z}_{spatial}^{Width}(A_m) \in R^{C_m \times 1 \times 1 \times W_m}, \quad (4)$$

where indexed slicing (using Python notation) and broadcasting  $\mathcal{B}(\cdot)$  are performed. The learnable parameters  $d_i, h_j, w_k$  denote the weight for each feature slice in depth, height, and weight dimension, respectively. Crucial airway regions are gradually preferred with higher weights during learning. To model the inter-channel dependency, we adopt the excitation technique [16] on the compressed spatial knowledge. Specifically, the channel descriptor is obtained by:

$$U_m = \mathcal{Z}(A_m) = f_2(K_2 * f_1(K_1 * \mathcal{Z}_{spatial}(A_m))), \quad (5)$$

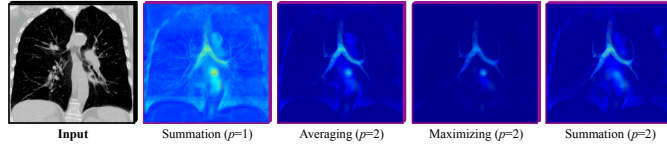
where  $K_1, K_2$  are 3-D kernels of size  $1 \times 1 \times 1$  and “\*” denotes convolution. Convolution with  $K_1$  decreases the channel numbers to  $C_m/r$  and that with  $K_2$  recovers back to  $C_m$ . The ratio  $r$  is the compression factor.  $f_1(\cdot)$  and  $f_2(\cdot)$  are non-linear activation functions. We choose Rectified Linear Unit (ReLU) as  $f_1(\cdot)$  and Sigmoid as  $f_2(\cdot)$  in the present study. Multiple channels are recombined through such channel reduction and increment, with informative ones emphasized and redundant ones suppressed. The final recalibrated feature  $\hat{A}_m = U_m \odot A_m$  is calculated as element-wise multiplication between  $U_m$  and  $A_m$ .

**Attention Distillation:** Recent studies [6, 24] on knowledge distillation showed that attention maps serve as valuable knowledge and can be transferred layer-by-layer from teacher networks to student networks. The motivation of our proposed attention distillation is that activation-based attention maps, which guide where to look at, can be distilled and exploited during backward transfer. Without

separately setting two different models, later layers play the role of teacher and “impart” such attention to earlier layers in the same model. Besides, to tackle the challenge of insufficient supervisory signals caused by severe class imbalance, the distillation can be viewed as another source of supervision. It produces additional gradient that helps to train deep CNNs for airway segmentation. Specifically, the attention distillation is performed between two consecutive features  $A_m$  and  $A_{m+1}$ . Firstly, the activation-based attention map is generated by  $G_m = \mathcal{G}(A_m), G_m \in R^{1 \times D_m \times H_m \times W_m}$ . Each voxel’s absolute value in  $G_m$  reflects the contribution of its correspondence in  $A_m$  to the entire segmentation model. One way of constructing the mapping function  $\mathcal{G}(\cdot)$  is to compute the statistics of activation values  $A_m$  across channel:  $G_m = \sum_{c=1}^{C_m} |A_m[c, :, :, :]|^p$ . The element-wise operation  $|\cdot|^p$  denotes the absolute value raised to the  $p$ -th power. More attention is addressed to highly activated regions if  $p > 1$ . Here, we adopt channel-wise summation  $\sum_{c=1}^{C_m} (\cdot)$  instead of maximizing  $\max_c(\cdot)$  or averaging  $\frac{1}{C_m} \sum_{c=1}^{C_m} (\cdot)$  because it is comparatively less biased. The summation operation retains all implied salient activation information without ignoring non-maximum elements or weakening discriminative elements. Visual comparison in Fig. 2 exhibits that summation with  $p > 1$  intensifies most the sensitized airway regions. Secondly, trilinear interpolation  $\mathcal{I}(\cdot)$  is performed to ensure that the processed 3-D attention maps share the same dimension. Then, spatial-wise Softmax  $\mathcal{S}(\cdot)$  is applied to normalize all elements. Finally, we drive the distilled attention  $\hat{G}_m$  closer to  $\hat{G}_{m+1}$  by minimizing the following loss function:

$$\mathcal{L}_{distill} = \sum_{m=1}^{M-1} \|\hat{G}_m - \hat{G}_{m+1}\|_F^2, \hat{G}_m = \mathcal{S}(\mathcal{I}(G_m)), \quad (6)$$

where  $\|\cdot\|_F^2$  is the squared Frobenius norm. With the current  $\hat{G}_m$  iteratively mimicking its successor  $\hat{G}_{m+1}$ , visual attention guidance is transmitted from the deepest to the shallowest layer. Note that such distillation process does not require extra annotation labor and can work with arbitrary CNNs freely. To ensure that the direction of knowledge distillation is from back to front, we detach the  $\hat{G}_{m+1}$  from the computation graph for each  $m$  in Eq. 6. Hence, the backward propagating gradients will not change the value of  $\hat{G}_{m+1}$ .



**Fig. 2.** The difference among mapping functions  $\mathcal{G}(\cdot)$  for the lattermost attention map.

**Model Design:** The proposed method employs U-Net [5] architecture. To enlarge the receptive field of CNNs and facilitate the feature learning of long-range

relationship, four pooling layers are used with five resolution scales. At each scale, both encoder and decoder have two convolution layers (kernel size  $3 \times 3 \times 3$ ) followed by instance normalization (IN) and ReLU. Feature recalibration is applied at the end of each resolution scale. Since high-level features in decoder are also of high-resolution, we perform the decoder-side attention distillation to pass down the fine-grained details that are missing in previous low-resolution attention maps. The encoder-side distillation is not favored because low-level features are more local-scale and not strongly related to airway description. Preliminary experiments confirmed that such design is optimum for our task.

**Training Loss:** To deal with hard samples, we use both the Dice [14] and Focal loss [10] for the segmentation task. Given the binary label  $y(x)$  and prediction  $p(x)$  of each voxel  $x$  in the volume set  $X$ , the combined loss is defined as:

$$\mathcal{L}_{seg} = -\left(\frac{2 \sum_{x \in X} p(x)y(x)}{\sum_{x \in X} (p(x) + y(x)) + \epsilon}\right) - \frac{1}{|X|} \left(\sum_{x \in X} (1 - p_t(x))^2 \log(p_t(x))\right), \quad (7)$$

where  $p_t(x) = p(x)$  if  $y(x) = 1$ . Otherwise,  $p_t(x) = 1 - p(x)$ . Parameter  $\epsilon$  is used to avoid division by zero. Together with attention distillation loss, the total loss function is given by  $\mathcal{L}_{total} = \mathcal{L}_{seg} + \alpha \mathcal{L}_{distill}$ , where  $\alpha$  balances these two terms.

### 3 Experiments and Results

**Datasets:** For evaluation, we conducted airway segmentation experiments on 90 3-D thoracic CT scans collected from two public datasets: 1) 70 scans from LIDC [1]; 2) 20 scans from the training set of EXACT’09 [12]. For the LIDC dataset, 70 CT scans whose slice thickness  $\leq 0.625$  mm were randomly chosen. For the EXACT’09 dataset, 20 training scans were provided without airway annotation. The airway ground-truth of each CT scan was manually corrected based on the interactive segmentation results of ITK-SNAP [23]. The acquisition and investigation of data were conformed to the principles outlined in the declaration of Helsinki [2]. The axial size of all CT scans was  $512 \times 512$  pixels with the spatial resolution of 0.5–0.781 mm. The number of slices in each scan varied from 157 to 764 with the slice thickness of 0.45–1.0 mm. All 90 CT scans were randomly split into training set (63 scans), validation set (9 scans), and testing set (18 scans). In the experiments, model training and hyper-parameter fine-tuning were performed only on the training set. The model with the best validation results was chosen and tested on the testing set for objectivity.

**Setup:** The CT pre-processing included Hounsfield Unit truncation, intensity normalization, and lung field segmentation [15]. In training phase, sub-volume CT cubes of size  $64 \times 224 \times 304$  were densely cropped near airways. Random horizontal flipping, shifting, and voxel intensity jittering were applied as on-the-fly data augmentation. The Adam optimizer ( $\beta_1 = 0.5, \beta_2 = 0.999$ ) with the learning rate  $2 \times 10^{-4}$  was used and the training converged after 50 epochs. In validation and testing phases, we performed the sliding window prediction with axial stride of 48. Results were averaged on overlapping margins and thresholded by

**Table 1.** Results (%) of comparison on the testing set (Mean±Standard deviation).

Method	BD	TD	TPR	FPR	DSC
3-D U-Net [5]	87.2±13.7	73.8±18.7	85.3±10.4	0.021±0.015	91.5±2.9
V-Net [14]	91.0±16.2	81.6±19.5	87.1±13.6	0.024±0.017	92.1±3.6
VoxResNet [4]	88.2±12.6	76.4±13.7	84.3±10.4	0.012±0.009	92.7±3.0
Wang et al. [20]	93.4±8.0	85.6±9.9	88.6±8.8	0.018±0.012	93.5±2.2
Juarez et al. [9]	77.5±20.9	66.0±20.4	77.5±15.5	<b>0.009±0.009</b>	87.5±13.2
Qin et al. [15]	91.6±8.3	82.1±10.9	87.2±8.9	0.014±0.009	<b>93.7±1.9</b>
Juarez et al. [8]	91.9±9.2	80.7±11.3	86.7±9.1	0.014±0.009	93.6±2.2
Jin et al. [7]	93.1±7.9	84.8±9.9	88.1±8.5	0.017±0.010	93.6±2.0
Our proposed	<b>96.2±5.8</b>	<b>90.7±6.9</b>	<b>93.6±5.0</b>	0.035±0.014	92.5±2.0

**Table 2.** Results (%) of ablation study on the testing set (Mean±Standard deviation).

Method	BD	TD	TPR	FPR	DSC
Baseline	91.6±9.2	81.3±11.5	87.2±8.6	<b>0.014±0.008</b>	<b>93.7±1.7</b>
Baseline + cSE [26]	95.1±6.2	88.5±8.3	92.4±5.5	0.033±0.015	92.3±1.9
Baseline + PE [16]	95.7±5.1	88.4±7.9	92.3±5.9	0.037±0.019	91.8±2.8
Baseline + FR	96.1±5.9	<b>90.8±7.5</b>	92.9±5.9	0.034±0.016	92.3±2.3
Baseline + AD	94.9±6.9	88.3±8.2	91.8±6.2	0.029±0.014	92.8±1.4
Our proposed	<b>96.2±5.8</b>	90.7±6.9	<b>93.6±5.0</b>	0.035±0.014	92.5±2.0

0.5 for binarization. No post-processing was involved. For the hyper-parameter setting, we empirically chose  $\alpha = 0.1$ ,  $\epsilon = 10^{-7}$ ,  $p = 2$ , and  $r = 2$ . Such setting may not be optimum and elaborate tuning may be conducted in future work.

**Metrics:** Considering clinical practice, only the largest connected component of segmented airways was kept and five metrics were used: (a) Branches detected (BD), (b) Tree-length detected (TD), (c) True positive rate (TPR), (d) False positive rate (FPR), and (e) Dice coefficient (DSC). We referred to [12] for airway tree-based metric definition (a)–(d) and the trachea region is excluded. For overall segmentation estimation by DSC, trachea is included in computation.

**Quantitative results:** Since we adopted U-Net as backbone, comparison experiments were performed with other encoder-decoder CNNs: the original 3-D U-Net [5], its variants V-Net [14], and VoxResNet [4]. We also compared our method with five state-of-the-art methods: Wang et al. [20], Juarez et al. [9], Qin et al. [15], Juarez et al. [8], and Jin et al. [7]. All these methods were re-implemented in PyTorch and Keras by ourselves and fine-tuned on the same dataset. Table 1 shows that our method achieved the highest BD, TD, and TPR with a compelling DSC. The first three metrics directly reflect that our model outperformed the others in the detection of airways, especially distal parts of thin branches. Its superior sensitivity also comes with a side effect of slight performance decline in FPR and DSC. This may be ascribed to: 1) Our model successfully detected some true thin airways that were too indistinct to be annotated properly by experts. When calculating the evaluation metrics, these real branches were counted as false positives and therefore causing higher FPR and lower DSC. 2) A little leakage was produced at bifurcations where the contrast between airway lumen and wall was fairly low. In this situation, the segmentation model was inclined to predict voxels as airway due to the reinforced feature learning of tiny branches. However, it is clinically worthwhile that much more



branches were segmented at such a trivial loss. Furthermore, we conducted ablation study to investigate the validity of each component in our method. The model trained without feature recalibration (FR) and attention distillation (AD) was indicated as baseline. Two very recently proposed recalibration methods (cSE [26] and PE [16]) were introduced into our baseline for comparison. As shown in Table 2, FR brings more performance gain than cSE and PE under the same experiment setting, confirming the importance of reasonably integrating spatial knowledge for recalibration. All components (cSE, PE, FR, and AD) increased the baseline model’s sensitivity of detecting airways and also produced more false positives. In view of the metrics BD, TD, and TPR, both modules of FR and AD do contribute to the high sensitivity of our model.

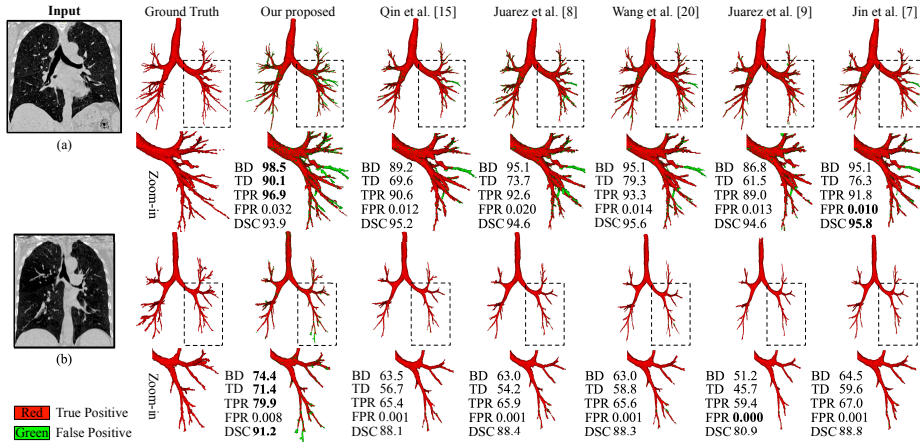
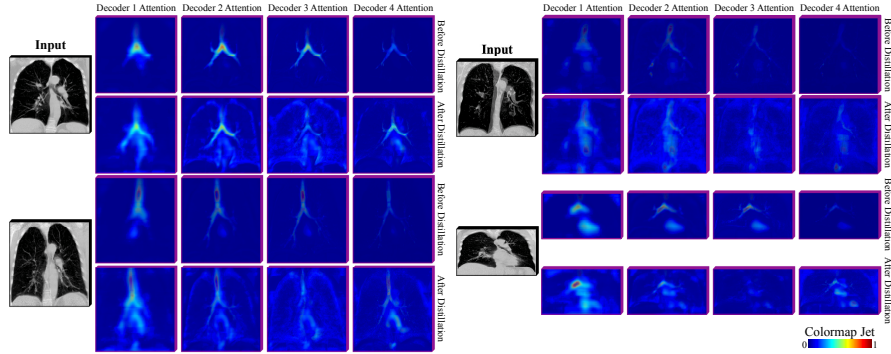


Fig. 3. Rendering of segmentation results of (a) easy and (b) hard testing cases.

**Qualitative results:** Segmentation results are shown in Fig. 3 to demonstrate the robustness of our method on both easy and hard cases. In line with Table 1, all methods performed well on extracting thick bronchi. Compared with state-of-the-art methods, more visible tiny branches were reconstructed by the proposed method with high overall segmentation accuracy maintained. In Fig. 3 (a) and (b), some false positives were identified as true airways after retrospective evaluation of labels. These branches were unintentionally neglected due to annotation difficulty. Moreover, to intuitively assess the effect of attention distillation, attention maps from decoder 1–4 are visualized in Fig. 4. After distillation, the activated regions become more distinct and the task-related objects (e.g., airway, lung) are enhanced accordingly. The improved attention on airways and lung boundary explains that the model comprehended more context and therefore achieved higher sensitivity to intricate bronchioles. Another interesting finding is that although the last attention map is not refined in distil-

lation, it still gets polished up because better representation learned at previous layers in turn affects late-layer features.



**Fig. 4.** Pseudo-color rendering of attention maps (decoder 1–4) before and after distillation process. These maps are min-max scaled and rendered with Jet colormap.

## 4 Conclusion

This paper proposed a highly sensitive method for CNNs-based airway segmentation. With the spatial-aware feature recalibration module and the gradually-reinforced attention distillation module, feature learning of CNNs becomes more effective and relevant to airway perception. Extensive experiments showed that our method detected much more bronchioles while maintaining competitive overall segmentation performance, which corroborates its superior sensitivity over state-of-the-art methods and the validity of the two constituent modules.

## References

1. Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., 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* **38**(2), 915–931 (2011)
2. Association, W.M., et al.: World medical association declaration of helsinki. ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization* **79**(4), 373 (2001)
3. Charbonnier, J.P., Van Rikxoort, E.M., Setio, A.A., Schaefer-Prokop, C.M., van Ginneken, B., Ciompi, F.: Improving airway segmentation in computed tomography using leak detection with convolutional networks. *MedIA* **36**, 52–60 (2017)
4. 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)

5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: MICCAI. pp. 424–432. Springer (2016)
6. Hou, Y., Ma, Z., Liu, C., Loy, C.C.: Learning lightweight lane detection CNNs by self attention distillation. In: ICCV. pp. 1013–1021. IEEE (2019)
7. Jin, D., Xu, Z., Harrison, A.P., George, K., Mollura, D.J.: 3D convolutional neural networks with graph refinement for airway segmentation using incomplete data labels. In: MLMI. pp. 141–149. Springer (2017)
8. Juarez, A.G.U., Tiddens, H., de Bruijne, M.: 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. Juarez, A.G.U., Selvan, R., Saghir, Z., de Bruijne, M.: A joint 3D unet-graph neural network-based method for airway segmentation from chest CTs. In: MLMI. pp. 583–591. Springer (2019)
10. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV. pp. 2980–2988. IEEE (2017)
11. Lo, P., Sporning, J., Ashraf, H., Pedersen, J.J., de Bruijne, M.: Vessel-guided airway tree segmentation: a voxel classification approach. *MedIA* **14**(4), 527–538 (2010)
12. Lo, P., Van Ginneken, B., Reinhardt, J.M., Yavarna, T., De Jong, P.A., Irving, B., Fetita, C., Ortner, M., Pinho, R., Sijbers, J., et al.: Extraction of airways from CT (EXACT’09). *IEEE TMI* **31**(11), 2093–2107 (2012)
13. Meng, Q., Roth, H.R., Kitasaka, T., Oda, M., Ueno, J., Mori, K.: Tracking and segmentation of the airways in chest CT using a fully convolutional network. In: MICCAI. pp. 198–207. Springer (2017)
14. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: IC3DV. pp. 565–571. IEEE (2016)
15. 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: MICCAI. pp. 212–220. Springer (2019)
16. Rickmann, A.M., Roy, A.G., Sarasua, I., Navab, N., Wachinger, C.: ‘Project & Excite’ modules for segmentation of volumetric medical scans. In: MICCAI. pp. 39–47. Springer (2019)
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
18. Selvan, R., Kipf, T., Welling, M., Pedersen, J.H., Petersen, J., de Bruijne, M.: Extraction of airways using graph neural networks. *arXiv e-prints* (2018)
19. Van Rikxoort, E.M., Baggerman, W., van Ginneken, B.: 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. pp. 341–349 (2009)
20. 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: MICCAI. pp. 348–356. Springer (2019)
21. Xu, Z., Bagci, U., Foster, B., Mansoor, A., Udupa, J.K., Mollura, D.J.: A hybrid method for airway segmentation and automated measurement of bronchial wall thickness on CT. *MedIA* **24**(1), 1–17 (2015)
22. Yun, J., Park, J., Yu, D., Yi, J., Lee, M., Park, H.J., Lee, J.G., Seo, J.B., Kim, N.: Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net. *MedIA* **51**, 13–20 (2019)

23. Yushkevich, P.A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J.C., Gerig, G.: User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage* **31**(3), 1116–1128 (2006)
24. Zagoruyko, S., Komodakis, N.: Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. In: ICLR. OpenReview.net (2017)
25. Zhao, T., Yin, Z., Wang, J., Gao, D., Chen, Y., Mao, Y.: Bronchus segmentation and classification by neural networks and linear programming. In: MICCAI. pp. 230–239. Springer (2019)
26. Zhu, W., Huang, Y., Zeng, L., Chen, X., Liu, Y., Qian, Z., Du, N., Fan, W., Xie, X.: AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical Physics* **46**(2), 576–589 (2019)