3D Convolutional Neural Networks with Graph Refinement for Airway Segmentation Using Incomplete Data Labels

Dakai Jin, Ziyue Xu(✉), Adam P. Harrison, Kevin George, and Daniel J. Mollura

National Institutes of Health, Bethesda, MD, USA
ziyue.xu@nih.gov

Abstract. Intrathoracic airway segmentation from computed tomography images is a frequent prerequisite for further quantitative lung analyses. Due to low contrast and noise, especially at peripheral branches, it is often challenging for automatic methods to strike a balance between extracting deeper airway branches and avoiding leakage to the surrounding parenchyma. Meanwhile, manual annotations are extremely time consuming for the airway tree, which inhibits automated methods requiring training data. To address this, we introduce a 3D deep learning-based workflow able to produce high-quality airway segmentation from incompletely labeled training data generated without manual intervention. We first train a 3D fully convolutional network (FCN) based on the fact that 3D spatial information is crucial for small highly anisotropic tubular structures such as airways. For training the 3D FCN, we develop a domain-specific sampling scheme that strategically uses incomplete labels from a previous highly specific segmentation method, aiming to retain similar specificity while boosting sensitivity. Finally, to address local discontinuities of the coarse 3D FCN output, we apply a graph-based refinement incorporating fuzzy connectedness segmentation and robust curve skeletonization. Evaluations on the EXACT'09 and LTRC datasets demonstrate considerable improvements in airway extraction while maintaining reasonable leakage compared with a state-of-art method and the dataset reference standard.

Keywords: Airway tree segmentation · CT · Convolutional neural network · Graph method · Incomplete label

1 Introduction

The intrathoracic airway tree is one of the major organs in the respiratory system. Its morphological changes, such as lumen shape or wall thickness, often

Z. Xu—This work is supported by the Intramural Research Program of the National Institutes of Health, Clinical Center and the National Institute of Allergy and Infectious Diseases. We also thank Nvidia for the donation of a Tesla K40 GPU.

Q. Wang et al. (Eds.): MLMI 2017, LNCS 10541, pp. 141–149, 2017.
DOI: 10.1007/978-3-319-67389-9_17

relate to the progression of several pulmonary diseases, e.g. asthma and chronic obstructive pulmonary disease (COPD) [1]. The development of *in vivo* computed tomography (CT)-based quantitative airway tree analysis provides valuable information for disease detection, etiologies investigation, and treatment-effects evaluation. Airway tree segmentation is also often a prerequisite for quantitative morphological analysis in image-based approaches.

Due to the structural complexity, this 3D segmentation task is extremely tedious and time consuming using manual or semi-automatic methods, taking more than 7 h or up to 2.5 h [2], respectively. In addition, annotations may be error prone. On the other hand, many automatic methods have been reported, including threshold-based [3], morphology-based [4] and 2D learning-based [5, 6]. Among these, different variations of region growing (RG) are often used. However, imaging resolution limits, airway wall thinness, and other artifacts can all harm CT contrasts, causing RG methods to struggle with leakage into the adjacent lung parenchyma. In contrast, 2D learning-based methods [5,6] potentially add more robustness; however, the extracted airway branches are limited due to the failure of exploiting 3D spatial information, which can be important for 3D airway tree delineation.

In the past decade, deep learning approaches have gained significant success within the natural image domain, due to their unparalleled performance on challenging data [7,8]. Using this powerful technique, several deep learning based-approaches have already been applied to medical image classification and segmentation tasks [6,9–11], most of which work in 2D. For airway segmentation, Charbonnier et al. [6] proposed to use 2D convolutional neural networks (CNNs) for detecting leakages, greatly increasing the extracted airway length compared to previous results. However, their framework does not directly provide a probability map for airway candidates and 3D continuity and tree structural information are not fully utilized under 2D settings. Finally, a limiting factor uniting all these learning-based methods [5,6] is the extreme cost in obtaining annotated airway segmentation as training data. Avoiding the need for fully-annotated data is a crucial quality toward wider adoption within clinical settings.

To address these gaps, we introduce a new 3D deep fully convolutional network (FCN) workflow designed for training on weakly annotated tree structures. Our method is based on the following two modules. (1) A 3D FCN is trained to perform efficient end-to-end learning and inference suitable for 3D airway trees. Crucially, instead of training on manually annotated airways, our workflow trains on incomplete and weakly labeled annotations produced by a previous, highly-specific and moderately-sensitive automated method [4]. We introduce a new domain-specific sampling scheme tailored for tree-like structures, and train our 3D FCN to learn the real airway patterns achieving a similarly high specificity, but with markedly increased sensitivity. (2) Graph-based methods are applied to refine the FCN output, by first using fuzzy-connectedness RG, which combines the 3D FCN probability map and the original CT intensity; and followed by a skeletonization guided leakages removal. We evaluate our method on the Extraction of Airways from CT 2009 (EXACT09) and Lung Tissue Research Consortium (LTRC) datasets, demonstrating high specificity while extracting

on average 30 and 70 more tree branches, respectively, compared to the highly-specific baseline method [4].

2 Method

Our method's overall operation is illustrated in Fig. 1. In the following, we first provide details on our proposed 3D FCN architecture, followed by an explanation of our principled and domain-specific data sampling method that allows training with incomplete labels. Finally, we discuss our graph-based refinement methods.

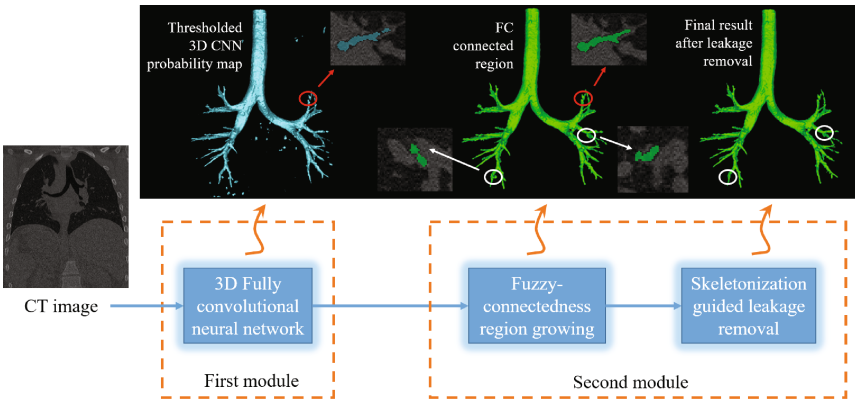

Fig. 1. Schematic illustration of the workflow of the new airway segmentation method.

2.1 3D FCN Architecture

A high quality airway probability map is crucial for a successful segmentation. Although 2D CNN-based airway segmentation may achieve improved performance compared with previous learning based methods [5], we postulate that 3D spatial information, especially the 3D branch-level continuity and junction-bifurcating patterns, is important for segmenting airway structures as well as avoiding local leakages to the neighboring lung parenchyma.

Due to high memory and computational demands, training on one entire 3D chest or whole body CT image is not possible with standard equipment. As a result, images are often broken up into relatively small 3D regions of interest (ROIs) [10,11]. One major problem with this is that the size of large organs or anatomies can exceed the ROI of the training data, therefore hampering the learning ability and robustness. In contrast, the airway is a relatively small-scale structure with tree-like topology. Thus, a relatively small 3D ROI contains sufficient spatial information to characterize the airway branch and junction-level patterns. Moreover, since conservative GR can reliably extract large airway branches, e.g. trachea and left-right main bronchi, we focus our FCN on learning

the middle and small size airway branches. Thus, airways are an application highly amenable for 3D FCNs, even with current technological limitations.

We adapt and modify the 3D U-Net architecture [11]. To help preserve very small airways at peripheral sites, we trim the previously deeper 3D U-Net to contain only two pooling layers, resulting in three different spatial resolutions. Downsizing the depth of our network allowed us to increase the width of the convolutional layers at the middle and higher levels, potentially helping to learn the structure variations in 3D space. Additional upsampling path with short-cuts from the corresponding pooling layers to the upsampling layers is also incorporated aiming to better capture finer scale airways. Kernels used in all convolutional layers are $3 \times 3 \times 3$ except the last layer using a $1 \times 1 \times 1$ kernel. Apart from these modifications, we use the same architectural choices as 3D U-Net [11]. See Fig. 2 for the details of our network architecture. Cross-entropy loss is used in training. In our end-to-end training, the loss function is computed over all voxels in the ROI. Due to the sparsity of airway structures in the training data, the distribution of airway and background voxels is highly biased. Therefore, a global class-balancing weight is applied to the loss function, which is shown below:

$$loss = -\beta \sum_{j \in Y_+} \log \hat{y}_j - (1-\beta) \sum_{j \in Y_-} \log (1-\hat{y}_j), \tag{1}$$

where $\hat{y}_j$ is the computed value after the final convolutional layer, Y_+ and Y_- represent the set of the foreground and background airway labels, and $\beta = mean\,(|Y_-|/|Y|)$ is global weight precomputed over the entire training data.

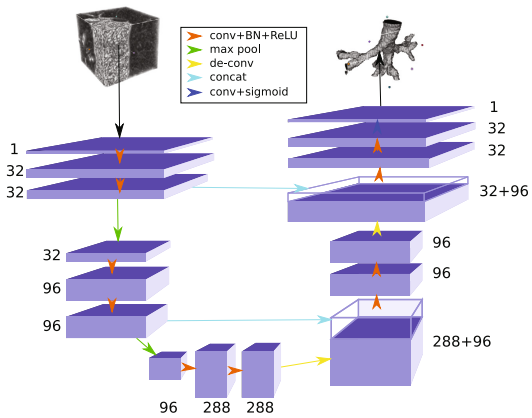

Fig. 2. 3D FCN architecture designed for small airway tubular structures. The number of channels is denoted at the side of each feature map.

2.2 Training with Incomplete Labeling

Obtaining fully-annotated medical image segmentation is always a bottleneck, but it is particularly costly, tedious, and error-prone for airways. For this reason, we investigate whether incomplete, but automatically generated, labels can serve as effective training data for our 3D FCN method. Specifically, we use labels generated from a previous conservative segmentation method that has only a small number of airway leakages, but is unable to capture all branches. In effect, this means that baseline labels are highly specific but not sufficiently sensitive. Without loss of generality, for this work we use Xu *et al.*'s method [4] as the baseline segmentation. Other methods with high specificity and moderate sensitivity are also suitable to serve as the baseline segmentation. Our aim is to retain similar specificity of the baseline segmentation, while increasing the sensitivity.

Instead of training using a sliding box across the entire volume, we use a more principled approach tailored for tree-like structures. (1) To ensure high-quality ROIs with positive examples, we extract sub-volumes along the baseline segmentation. Specifically, we generate the centerline representation of the baseline airway tree, and along each centerline we randomly sample 15% of the voxels and crop a $80 \times 80 \times 72$ voxel-size ROI centered at this centerline voxel. To augment the data, we also randomly shift within 15 voxels in each dimension. In this way, there are roughly 500 ROIs containing positive examples for one CT image. This ensures that there are sufficient examples capturing the local airway branching patterns, e.g., various locations, sizes and branch orientations, as well as the lung background parenchyma. (2) To obtain sufficient negative examples, 500 $80 \times 80 \times 72$ ROIs are also randomly sampled from each CT image.

One potential drawback of this incomplete labeling training strategy is that the baseline segmentation will contain false negatives. Fortunately, however, the use of the global weight, β, helps address this issue. Recall that with the paucity of positive examples, the loss weighting reduces the impact of each background voxel on the training process. For this reason, the detrimental effects of any false negatives will also be reduced. In contrast, foreground voxels are more highly weighted, meaning the impact of true positives will be magnified. Thus, when the labels are noisy, the weighting scheme is suited for high-specificity, moderate-sensitivity settings. This exactly describes the baseline segmentation we use.

2.3 Graph-Based Refinement

In the second module of our workflow, we propose to use graph-based refinement to address the local discontinuity, variations, and coarse spatial resolution of the probability map from the 3D FCN. First, we apply fuzzy connectedness segmentation to grow the airway region by incorporating the 3D FCN airway probability information and the CT intensity into the *fuzzy affinity* function in FC method. In this way, information from the high resolution of the original CT attenuations are used to enhance the accurate, but sometimes, coarse FCN output. The initial seeds for the FC are generated by setting a high threshold

(> 0.8) on the FCN probability map. The general form for the fuzzy affinity function incorporates adjacency, μ_α, feature homogeneity, μ_ψ, and expected object features, μ_ϕ. Let p and q denote any two voxels in 3D digital space. The adjacency component is defined as: $\mu_\alpha(p,q) = 1$, if p and q are 26-adjacent, and '0' otherwise. $\mu_\psi(p,q)$ captures the homogeneity between p and q, with a smaller value for more similar pair, whereas $\mu_\phi(p,q)$ calculates how well the pair of p and q match the expected feature distribution. We refer the reader to [4] for more details on specific forms of these affinity functions. In our context, two features are used at a given voxel p: intensity $I(p)$ and the 3D FCN probability map $FCN(p)$. Both $\mu_\psi(p,q)$ and $\mu_\phi(p,q)$ incorporate these two features into their affinity calculation. The combined affinities are given as follows:

$$\mu_k(p,q) = \gamma\mu_k(FCN(p), FCN(q)) + (1-\gamma)\mu_k(I(p), I(q)), \quad k \in \{\psi, \phi\} \qquad (2)$$

where γ is the factor to control the influence of intensity as compared with the 3D FCN features. In this paper, $\gamma = 0.75$ is used since FCN output features higher contrast in avoiding leakage.

Because of the high quality of the 3D FCN map, the segmented results after FC based refinement do not typically contain large leakages. Mostly, they are small boundary protrusions or branch like structures adjacent to airway branches as shown in Fig. 1. Therefore, post-pruning methods are applied to remove some leakages. Here, we use a robust curve skeletonization approach [12] to remove the leakages appeared as the boundary protrusions. The method generates the curve skeleton of airway tree by iteratively growing new skeletal branches computed as a minimum-cost path. The meaningfulness of a skeletal branch is defined by its global context and scales, therefore, enhancing its robustness in stopping false branches that comes from the leaked segmentation region. Based on the skeleton representation of the airway tree, we also compute the scales of terminal branches and prune those whose maximal local radius exceeds its minimal radius by three times, i.e., the extreme scale difference indicates leakage.

3 Experiments and Results

We evaluated our airway segmentation method on CT scans from the EXACT09 challenge [13] and the LTRC [14] with slice resolution 0.5 to 1.25 mm. Images from EXACT09 were acquired using different scanning protocols and reconstruction parameters. The dataset ranges from clinical dose to ultra low-dose scans, from healthy subjects to diseased patients, and from full inspiration to full expiration. In contrast, images from LTRC are those of patients with COPD and interstitial lung disease using more clinical standardized imaging protocols.

We trained our 3D FCN from scratch using 12 randomly chosen training images from the EXACT09 challenge because of its imaging diversity. This results in around 12000 training samples. For testing, we used the entire 20 volume EXACT09 test set and another randomly selected 20 volumes from LTRC. Optimization was performed using stochastic gradient descent, with initial learning rates of 1e-6, a momentum of 0.99, and a weight decay of 0.05. Training converged after 3∼4 epochs. When testing, we applied overlapped sliding windows

with sub-volumes sized to $80 \times 80 \times 72$ and strides of $48 \times 48 \times 40$. Afterwards, the probability maps of sub-volumes are aggregated to get the whole volume prediction. In general, it takes roughly 10 min to process one CT volume.

Here, we compared our segmentation results to the baseline results of [4]. On the other hand, because LTRC provides the airway reference segmentations, we compare our LTRC results with both baseline results of [4] and the dataset reference standard [14]. Note that the LTRC reference standard is prone to leakages. Three evaluation metrics are used: segmented branch count, tree length, and leakage count. Two experts manually examined the airway segmentations, judging if a segmented branch is a true branch or belongs to a leakage.

Table 1. Quantitative results (mean ± standard deviation) of new method compared against the baseline method Xu *et al.* [4] and LTRC reference standard [14].

	Additional branch count	Additional tree length (cm)	Additional tree length %	Additional leakage count
v.s. Xu *et al.* (EXACT09)	32.2 ± 32.1	66.8 ± 89.9	46.0 ± 49.4	5.2 ± 5.1
v.s. Xu *et al.* (LTRC)	78.8 ± 28.9	158.7 ± 59.7	82.9 ± 53.2	4.7 ± 3.2
v.s. reference (LTRC)	93.6 ± 32.6	181.4 ± 68.2	94.7 ± 60.8	1.3 ± 4.2

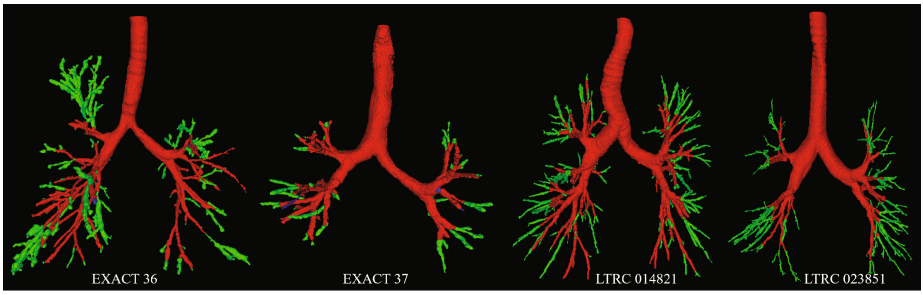

Fig. 3. 3D rendering of example airway segmentations of our method compared against Xu *et al.* [4] and the dataset reference standard [14] for EXACT09 and LTRC, respectively. Overlap regions are colored in red, and green indicates additional branches extracted by our method. Branches detected by the method of Xu *et al.* [4] or the reference standard [14], but missed by ours, are colored in blue (Color figure online).

The quantitative results are summarized in Table 1. Our method significantly increases the number of detected branches as well as the tree length in both

EXACT09 and LTRC datasets. In the EXACT09 dataset, an average of over 32 more branches and 46% tree length is obtained as compared to the baseline method [4]. In the LTRC dataset, further improvements are observed, i.e., 78 more tree branches and roughly 83% more tree length, as compared to the baseline method of [4]. Further improvements are observed when compared against the reference standard [14]. Meanwhile, the additional leakage counts are limited (< 5) in both datasets. As a result, we demonstrate markedly increased sensitivity, while retaining good specificity. Selected examples from both datasets are displayed in Fig. 3, visualizing the impact of these improvements. As shown, leakages in the new method are typically small, which will minimize the volume-based leakage metrics used in the EXACT09 challenge [13].

4 Conclusion

In this paper, we introduce a robust and accurate 3D deep learning-based workflow for airway tree segmentation, which is able to produce high-quality segmentations from incompletely labeled training data. Using a 3D fully convolutional neural network approach, we develop a training and domain-specific sampling scheme tailored for tree structure that uses incomplete labels from a previous highly specific segmentation method. Followed by the 3D FCN prediction, a graph-based refinement step is applied to address local discontinuities of the coarse 3D FCN output by incorporating fuzzy connectedness segmentation and the curve skeletonization. The airway segmentation results on the EXACT09 and LTRC dataset demonstrate considerable improvements in expanding the airway branch numbers and the tree length while maintaining limited leakage compared to a state-of-art method or the dataset reference standard. Importantly, these results are obtained without the need for highly cost manual airway annotations, increasing our method's usability and applicability going forward. Next step of our work involves more comprehensive evaluations, and full EXACT09 challenge metric will be performed to further demonstrate the effectiveness of our method.

References

1. Kuwano, K., Bosken, C.H., Paré, P.D., Bai, T.R., Wiggs, B.R., Hogg, J.C.: Small airways dimensions in asthma and in chronic obstructive pulmonary disease. Am. Rev. Respir. Dis. **148**(5), 1220–1225 (1993)
2. Tschirren, J., Yavarna, T., Reinhardt, J.M.: Airway segmentation framework for clinical environments. In: Second International Workshop on Pulmonary Image Analysis, London, UK, pp. 227–238 (2009)
3. 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: Second International Workshop on Pulmonary Image Analysis, pp. 341–349 (2009)
4. 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. Med. Imag. Anal. **24**(1), 1–17 (2015)

5. Lo, P., Sporring, J., Ashraf, H., Pedersen, J.J., de Bruijne, M.: Vessel-guided airway tree segmentation: a voxel classification approach. Med. Imag. Anal. **14**(4), 527–538 (2010)
6. 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. Med. Imag. Anal. **36**, 52–60 (2017)
7. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556 (2014)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, June 2016
9. 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). doi:10.1007/978-3-319-24574-4_28
10. Merkow, J., Marsden, A., Kriegman, D., Tu, Z.: Dense volume-to-volume vascular boundary detection. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9902, pp. 371–379. Springer, Cham (2016). doi:10.1007/978-3-319-46726-9_43
11. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). doi:10.1007/978-3-319-46723-8_49
12. Jin, D., Iyer, K.S., Chen, C., Hoffman, E.A., Saha, P.K.: A robust and efficient curve skeletonization algorithm for tree-like objects using minimum cost paths. Pattern Recogn. Lett. **76**, 32–40 (2016)
13. Lo, P., Van Ginneken, B., Reinhardt, J.M., et al.: Extraction of airways from CT (EXACT'09). IEEE Trans. Med. Imaging **31**(11), 2093–2107 (2012)
14. Karwoski, R.A., Bartholmai, B., Zavaletta, V.A., Holmes, D., Robb, R.A.: Processing of CT images for analysis of diffuse lung disease in the lung tissue research consortium. In: Proceedings of SPIE 6916, Medical Imaging 2008: Physiology, Function, and Structure from Medical Images (2008)