Tracking and Segmentation of the Airways in Chest CT Using a Fully Convolutional Network

Qier Meng[1], Holger R. Roth[2], Takayuki Kitasaka[3], Masahiro Oda[2],
Junji Ueno[4], and Kensaku Mori[1,2(✉)]

[1] Graduate School of Information Science, Nagoya University, Nagoya, Japan
`kensaku@is.nagoya-u.ac.jp`
[2] Graduate School of Informatics, Nagoya University, Nagoya, Japan
[3] School of Information Science, Aichi Institute of Technology, Toyota, Japan
[4] Graduate School of Health Sciences, Tokushima University, Tokushima, Japan

Abstract. Airway segmentation plays an important role in analyzing chest computed tomography (CT) volumes such as lung cancer detection, chronic obstructive pulmonary disease (COPD), and surgical navigation. However, due to the complex tree-like structure of the airways, obtaining segmentation results with high accuracy for a complete 3D airway extraction remains a challenging task. In recent years, deep learning based methods, especially fully convolutional networks (FCN), have improved the state-of-the-art in many segmentation tasks. 3D U-Net is an example that optimized for 3D biomedical imaging. It consists of a contracting encoder part to analyze the input volume and a successive decoder part to generate integrated 3D segmentation results. While 3D U-Net can be trained for any 3D segmentation task, its direct application to airway segmentation is challenging due to differently sized airway branches. In this work, we combine 3D deep learning with image-based tracking in order to automatically extract the airways. Our method is driven by adaptive cuboidal volume of interest (VOI) analysis using a 3D U-Net model. We track the airways along their centerlines and set VOIs according to the diameter and running direction of each airway. After setting a VOI, the 3D U-Net is utilized to extract the airway region inside the VOI. All extracted candidate airway regions are unified to form an integrated airway tree. We trained on 30 cases and tested our method on an additional 20 cases. Compared with other state-of-the-art airway tracking and segmentation methods, our method can increase the detection rate by 5.6 while decreasing the false positives (FP) by 0.7 percentage points.

Keywords: Airway segmentation · Fully convolutional network · Volume of interest

1 Introduction

Lung cancer is one of the leading causes of cancer-related death. Early detection and diagnosis are essential for decreasing lung cancer related deaths. With the

improvement of high-resolution 3D computed tomography (CT) images, more precise treatment of the disease is feasible. To obtain an accurate diagnosis and treatment, extraction of the airway tree in the 3D CT volume is an important step for analyzing various diseases. However, the segmentation of airways from CT volumes is very difficult due to the limited image intensity contrast and increased noise.

Many studies have focused on automated bronchus recognition from 3D chest CT volumes. Lo et al. [1] summarized airway segmentation methods within the EXACT09 challenge. Although most of the conventional approaches can extract the main bronchus and principle bronchi successfully, they fail when it comes to the extraction of the peripheral bronchi. To achieve more branch extraction in peripheral areas, we introduce the concept of "volume of interest" (VOI) that divides a given CT image into subimages containing a bronchial branch region [2]. Every bronchus branch is segmented inside the subimage, and is unified to form an integrated airway tree.

In recent years, fully convolutional networks (FCN) have outperformed the state-of-the-art in many segmentation tasks [3]. U-Net is one example that consists of a contracting encoder part to analyze the whole image and a successive decoder part to generate an integrated segmentation result [4]. 3D U-Net is the 3D extension of U-Net that replaces all 2D operations with their 3D counterparts [5].

In this paper, we build upon a method to segment the airway by using FCN within the VOI tracking framework [2]. The main contribution of our proposed work includes the implementation of 3D U-Net in the selected VOI tracking along the centerlines of the airway. At the same time, the segmentation results can be used to predict airway extraction in the next branching level. Compared to other state-of-the-art airway tracking and segmentation methods, our proposed method can increase the detection rate and reduce the false positive (FP) rate.

2   Method

Figure 1 shows the workflow of our method. We segment the trachea by region growing and set the initial VOI based on the segmented result. 3D U-Net is then used to segment the airway region in each VOI sequentially. To implement the
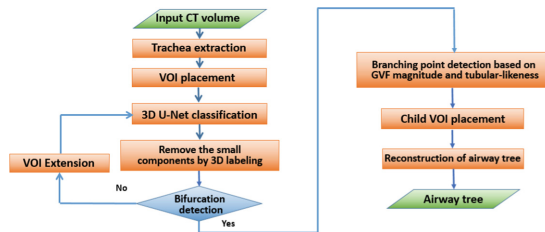

Fig. 1. Flowchart of the proposed method

3D U-Net to obtain the bronchus regions, three different kinds of VOI of different sizes are set. The size of the input volume of the network is $132 \times 132 \times 116$ voxels. The size of output volume of the network is $44 \times 44 \times 28$ voxels. The size of the VOIs used in airway tracking algorithm is dynamically adjusted in the airway tracking algorithm depending on the diameter of the bronchi.

2.1 3D U-Net Based Airway Tracking Algorithm

The employed network architecture is the 3D extension of U-Net proposed by Ronneberger et al. [4], which was initially utilized for biomedical image analysis. While U-Net is an entirely 2D architecture, the network utilized in this paper applied all the operations in their 3D version, such as 3D convolutions, 3D max pooling, and 3D up-convolutions [5]. We use the open source[1] implementation of 3D U-Net developed in Caffe [6]. The 3D U-Net architecture consists of an analysis and a synthesis path with four resolution levels each. Each resolution level in the analysis path contains two $3 \times 3 \times 3$ convolutional layers, each followed by rectified linear units (ReLu), and then a $2 \times 2 \times 2$ max pooling with strides of two in each dimension. In the synthesis path, the convolutional layers are replaced by up-convolutions of $2 \times 2 \times 2$ with strides of two in each dimension. These are followed by two $3 \times 3 \times 3$ convolutions each with a ReLu. Furthermore, 3D U-Net employs shortcut (or skip) connections from layers of equal resolution in the analysis path in order to provide higher-resolution features to the synthesis path [5]. The last layer contains a $1 \times 1 \times 1$ convolution that reduces the number of output channels to the number of class labels which is $K = 2$ in our specialized airway extraction case. The model can be trained to minimize a newly introduced weighted voxel-wise cross-entropy loss:

$$\mathcal{L} = \frac{-1}{N} \left\{ \lambda \times \left(\sum_{x \in N_0} \log \left(\hat{p}_k(x) \right) \right) + (1 - \lambda) \times \left(\sum_{x \in N_1} \log \left(\hat{p}_k(x) \right) \right) \right\}, \qquad (1)$$

where $\hat{p}$ are the softmax output class probabilities.

λ is a weight factor, N are the total number of voxels x, and $k \in [0, 1]$ indicates the lung tissue label of the airway region and the ground truth of the airway regions. The input to this loss function are real valued predictions $x \in [-\infty, +\infty]$ from the last convolutional layer. We apply weight λ to the loss function (Eq. 1) in order to balance the lung tissue as background with respect to the airway regions as foreground. We choose λ with $\lambda = \frac{N_1}{N}$, where N is the number of voxels for the lung region, and N_1 corresponds to the number of airway (foreground) voxels. Figure 2 shows examples of airway region extraction by 3D U-Net.

2.2 VOI Placement

Due to the requirement of 3D U-Net, three kinds of VOIs are defined. Figure 3 shows the three kinds of VOI. One is the V_{TRACK} which is set to contain a

---

[1] http://lmb.informatik.uni-freiburg.de/resources/opensource/unet.en.html.
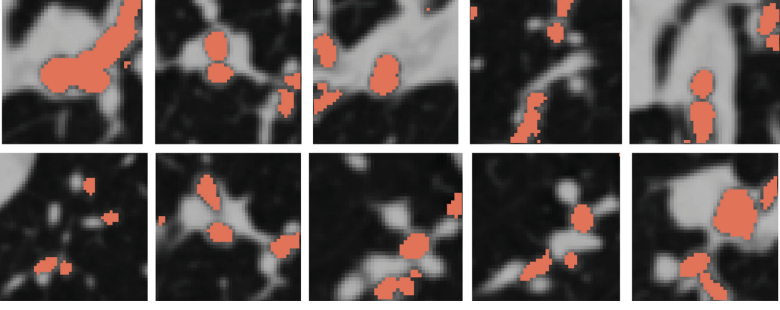
Fig. 2. Extraction results by 3D U-Net.

bronchial branch in the airway tracking algorithm. Since the 3D U-Net requires a constant size of images as input, we prepare two wider VOIs, V_{SEG} and V_{3DU}. The V_{SEG} of 3D U-Net with a size of $44 \times 44 \times 28$ voxels is used to save the segmentation output of 3D U-Net. The V_{3DU} is used as the input of 3D U-Net for training and testing with a size of $132 \times 132 \times 116$ voxels. The height of the V_{TRACK} is 28 voxels, and the length and width is twice the average diameter of the airway regions. We resample the segmentation result from V_{SEG}, save it in V_{TRACK}, and reconstruct all of the candidate airway regions sequentially to form a complete airway tree.
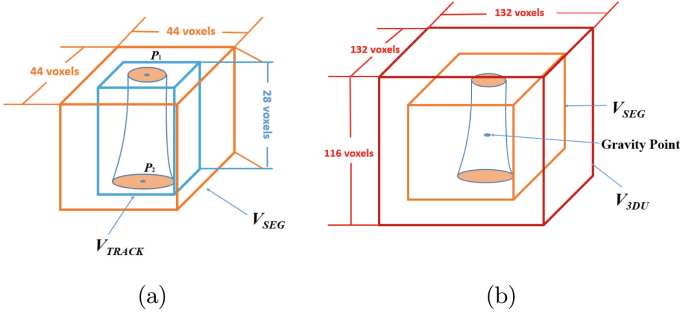

(a)   (b)

Fig. 3. Illustration of three different kinds of VOI. (a) Shows that V_{SEG} and V_{TRACK} have the same original point. The orange VOI is V_{SEG}, and the blue VOI is the V_{TRACK}. (b) Shows that V_{3DU} corresponds to V_{SEG}. The red VOI is V_{3DU}.

2.3 Airway Tracking Algorithm

We introduce the 3D U-Net in VOI-based airway tracking algorithm (3D U-VOI) in this section. The airway tracking algorithm starts to place a VOI at the trachea and the airway trees are segmented to place child VOIs from the root along the centerline. In each VOI, an airway region is extracted by 3D U-VOI.

The size of each VOI is determined according to the diameter of airway regions. The VOI direction is determined by the centerline extracted in the last VOI. The explanation of the airway tracking algorithm is outlined as follows:

1. Trachea extraction: A trachea region is roughly segmented by a method similar to [7]. The threshold value used in the region growing method is manually determined to extract the trachea region.

2. VOI placement: A VOI is defined as a cuboid with eight vertices and two points: original point $\mathbf{P}_1$, which determines the position of the VOI, and $\mathbf{P}_2$, which are the centers of the two faces shown in Fig. 3(a).

3. 3D U-Net-based airway region extraction in VOI: In order to extract the bronchial branches after trachea extraction, the 3D U-Net is used to extract the bronchial regions in the VOI.

4. Leakage removal: After obtaining the airway regions in each VOI, the 3D labeling process is used for removing the small components detected by 3D U-Net.

5. Furcation detection: The furcation regions are detected by analyzing the number of the connected components on the VOI image surface. The number of connected components indicates the number of furcations. The following conditions will be checked iteratively in each VOI:

– $\mathbf{N_c = 0}$. In this VOI, there is no bronchus region to be detected, and the tracing will be terminated.

– $\mathbf{N_c = 1}$. In this VOI, furcation hasn't been detected, but the bronchial region is continued to be traced.

– $\mathbf{N_c = 2}$ or $\mathbf{3}$. In this VOI, the bifurcation and trifurcation has been detected.

6. VOI extension: We extend the VOI to the bronchus running direction by one voxel. After that the Step (3) is processed again.

7. Branching point detection: We apply the gradient vector flow (GVF) magnitude and tubular-likeness function based on GVF to extract the centerline [9]. The branching point is detected based on this centerline.

8. Child VOI placement: After the branching point detection, the child VOIs V_{TRACK} are placed based on the branching point detected and the center points of the components on the VOI surface. Except for V_{TRACK}, the V_{3DU} and V_{SEG} are placed for the 3D U-Net shown in Fig. 3(b). After the child VOI placement, the extraction procedure is repeated [8].

9. Airway tree reconstruction: Finally, we reconstruct all of the candidate bronchus region to form an integrated airway tree.

3   Experiment and Results

To evaluate our proposed method, we used 50 clinical CT scans acquired with standard dose. The size of each slice scan was 512×512 pixels with a pixel size in the range of 0.625–0.976 mm. The number of slices in each dataset ranged from 239 to 962 slices with varying thicknesses from 0.625 to 2.00 mm. The data

set was randomly divided into two groups: the training data set which contains 30 cases and the testing data set containing 20 cases. We performed two comparisons. A comparison between the results of airway tracking algorithm with 3D U-Net and the results of only using 3D U-Net in a sliding-window approach without the tracking informations. Sliding-window U-Net method means randomly generating the sliding window around the candidate regions obtained by dilating the groundtruth data by ten voxels. Two indices are used for evaluation: (a) Dice Similarity Coefficients (DSCs), (b) false positive rate (FPR), the number of extracted voxels which were not bronchial voxels according to ground truth. The results are given in Fig. 4 and Table 1.
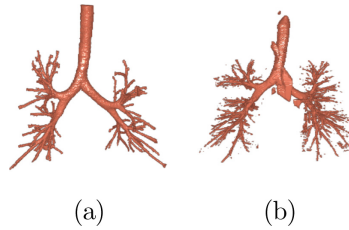

(a)   (b)

Fig. 4. Comparison between the result of airway tracking algorithm with 3D U-Net and the result of only using 3D U-Net.

Table 1. Comparison of segmentation results for airway tree in 20 chest CT examination volumes between the sliding-window-based U-Net and tracking-based U-Net in a sliding-window fashion.

DSCs	SW U-Net	Tracking U-Net
Mean	68.7%	86.6%
Std	20.6%	5.90%
Min	22.4%	75.9%
Max	83.6%	98.6%
FPR	SW U-Net	Tracking U-Net
Mean	0.209%	0.017%
Std	0.075%	0.013%
Min	0.077%	0.001%
Max	0.403%	0.053%

We also compared our proposed method with two methods, one is Kitasaka's method [2] based on the VOI and region growing, and the other is Meng's method based on local intensity analysis and machine learning technique [10]. The segmentation results, yielded by applying our proposed method to each of the 20 chest CT examinations, are provided in Table 2.

Table 2. Comparison of segmentation results for airway tree in 20 chest CT examination volumes. Here, TPR and FPR denotes the true positive and false positive rate of detected branches, respectively. DSCs is the Dice similarity score. (Note: The ground truth data was generated manually and BD indicates branch detection.)

Case	Proposed method				Kitasaka et al. [2]				Meng et al. [10]			
	BD	TPR	FPR	DSCs	BD	TPR	FPR	DSCs	BD	TPR	FPR	DSCs
1	114	85.1%	0.3%	86.7%	110	82.1%	1.8%	85.9%	116	86.6%	0.2%	58.9%
2	71	80.7%	0.2%	92.2%	58	65.9%	0.6%	79.7%	68	77.3%	0%	81.0%
3	142	71.7%	0.1%	89.3%	137	72.9%	0.7%	89.0%	140	70.7%	0.2%	78.4%
4	106	88.3%	0.2%	97.0%	97	80.8%	1.3%	85.1%	103	85.8%	0.2%	85.1%
5	82	96.5%	0.2%	87.2%	77	90.6%	0.7%	82.5%	80	94.1%	0.2%	82.6%
6	98	89.1%	0.1%	85.1%	85	77.3%	0.8%	85.1%	92	83.6%	0.1%	95.5%
7	95	72.5%	0.2%	80.2%	88	67.2%	0.9%	69.5%	91	69.5%	0.5%	95.2%
8	36	73.5%	0.2%	91.4%	30	61.2%	0.4%	80.2%	28	57.1%	0%	84.8%
9	58	77.3%	0.2%	86.8%	55	73.3%	0.6%	86.8%	54	70.1%	0.2%	86.9%
10	153	89.5%	0.1%	76.8%	149	87.1%	0.6%	80.7%	151	88.3%	0.1%	85.1%
11	101	84.2%	0%	75.9%	98	81.7%	0.4%	75.9%	102	85.0%	0.6%	70.1%
12	92	77.3%	0%	90.9%	95	79.8%	0.3%	90.9%	93	78.2%	0.2%	96.2%
13	89	68.9%	0.2%	87.5%	93	72.1%	1.6%	89.6%	91	70.5%	0%	80.2%
14	111	87.4%	0%	83.2%	108	85.0%	0.5%	85.1%	106	83.5%	0.2%	95.4%
15	54	90.0%	0.1%	80.3%	48	80.0%	1.1%	72.0%	52	86.7%	0%	85.1%
16	37	82.2%	0.2%	82.4%	31	68.9%	0.5%	70.3%	33	67.3%	0.2%	82.6%
17	33	67.3%	0%	98.6%	28	57.1%	1.7%	98.9%	30	61.2%	0.1%	96.7%
18	80	66.7%	0.1%	85.5%	76	63.3%	1.1%	85.5%	75	62.5%	0.1%	91.4%
19	81	61.8%	0.1%	88.3%	75	57.3%	0.3%	88.3%	82	62.6%	0.1%	98.9%
20	57	82.6%	0%	86.3%	52	75.4%	1.4%	86.3%	55	79.7%	0.5%	85.5%
AVG	84.5	79.6%	0.1%	86.6%	79.5	74.0%	0.8%	83.3%	82.1	76.0%	0.2%	86.0%
p-value	−	−	−	−	0.0027	0.0126	0.0015	0.011	0.0014	0.0013	0.0375	0.045

We measured four indices for evaluation: (a) number of branches extracted, (b) true positive rate (TPR), the ratio of extracted branches and total number of branches and (c) computing the false positive rate (FPR), the false branches which don't belong to ground truth, (d) Dice Similarity Coefficients (DSCs).

Training: In the training phase, we input both of the 3D CT volumes selected from the lung area and the corresponding ground truth data sets. We dilated the ground truth data by ten voxels as the candidate regions, and randomly selected the subvolume images inside the candidate regions. Data augmentation with random 3D rotations at 360° was done on-the-fly, which results in 35k images simulating tracking iterations. Training on 30 cases took 2 days for 35k

iterations on an NVIDIA GeForce GTX TITAN X with 12 GB memory, and one iteration generates one subvolume image.

Testing: We obtained the V_{3DU} in the airway tracking algorithm. After fixing the size and direction of the V_{3DU}, we input V_{3DU} into the trained network to obtain the candidate airway areas and save them into V_{SEG}. After that, we re-sampled the V_{TRACK} used in the airway tracking algorithm by cubic spline interpolation.

Figure 5 shows the evaluation results by the proposed method, as well as comparison with two other methods [2, 10].
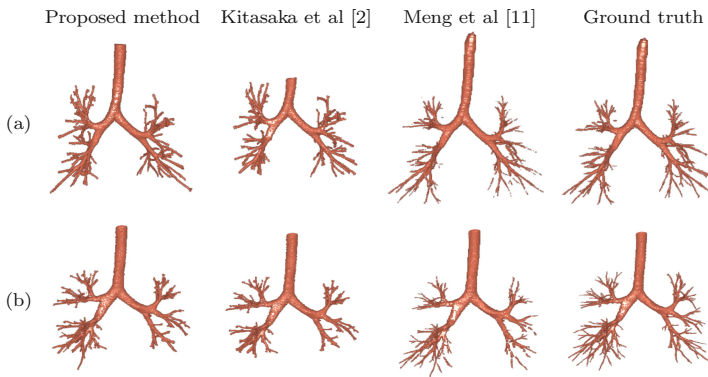

Proposed method   Kitasaka et al [2]   Meng et al [11]   Ground truth

(a)

(b)

Fig. 5. Comparison of airway segmentation results of the proposed, previous methods [2, 10], and ground truth. Upper and lower rows show the results of Case 3 and 8, respectively.

4   Discussion

According to the Fig. 4 and Table 1, it can be seen that the result of 3D U-Net in sliding-window can extract many peripheral bronchi but also generate many false positive (FP) regions as well. It is apparent that the proposed method can extract many peripheral bronchi and avoid the FP regions effectively. The 3D U-Net is effective to detect the peripheral bronchi in the chest CT images, however it produces many FP detections when applied in a straight-forward sliding window approach as proposed in [5]. After combining the 3D U-Net with the airway tracking algorithm, the false positive regions can be decreased significantly. Table 2 and Fig. 5 demonstrate that our proposed method outperforms the other two methods. To further illustrate the application results of our proposed method, we show a set of automatic airway segmentations in Fig. 5, which illustrates that our proposed method can extract more peripheral bronchi than

the method by Kitasaka et al. [2] and can extract more integrated branches than Meng et al. [10]. Our proposed method utilizes 3D U-Net to extract the airway region in each VOI, which effectively extracts the bronchus region and avoids leakage.

5   Conclusion

This paper introduced an effective integration of 3D U-Net into airway tracking and segmentation in combination with VOI-based algorithm. The proposed method was possible to improve segmentation accuracy significantly. Due to the memory restriction of the GPU, the input volume size to the network is fixed. In the future, networks with dynamic input sizes can be explored.

References

1. Lo, P., Sporring, J., Ashraf, H., Pedersen, J.J., de Bruijne, M.: Vessel-guided airway tree segmentation: a voxel classification approach. Med. Image Anal. **14**(11), 527–538 (2010)
2. Kitasaka, T., Mori, K., Hasegawa, J.I., Toriwaki, J.I.: A method for extraction of bronchus regions from 3D branch tracing and image sharpening for airway tree chest X-ray images by analyzing structural features of the bronchus. Forma **17**, 321–338 (2002)
3. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR, pp. 3431–3440 (2015)
4. 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
5. 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
6. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. In: ACM Multimedia 2014 Open Source Software Competition, pp. 675–687 (2014)
7. Mori, K., Hasegawa, J., Toriwaki, J., Anno, H., Katada, K.: Automated extraction and visualization of bronchus from 3D CT images of lung. In: Ayache, N. (ed.) CVRMed 1995. LNCS, vol. 905, pp. 542–548. Springer, Heidelberg (1995). doi:10.1007/978-3-540-49197-2_71
8. Meng, Q., Kitasaka, T., Oda, M., Ueno, J., Mori, K.: Airway segmentation from 3D chest CT volumes based on volume of interest using gradient vector flow. In: International Forum on Medical Imaging in Asia, vol. 11, pp. 192–195 (2017)

9. Bauer, C., Bischof, H.: A novel approach for detection of tubular objects and its application to medical image analysis. In: Rigoll, G. (ed.) DAGM 2008. LNCS, vol. 5096, pp. 163–172. Springer, Heidelberg (2008). doi:10.1007/978-3-540-69321-5_17

10. Meng, Q., Kitasaka, T., Nimura, Y., Oda, M., Ueno, J., Mori, K.: Automatic segmentation of airway tree based on local intensity filter and machine learning technique in 3D chest CT volume. Int. J. Comput. Assist. Radiol. Surg. **12**(2), 245–261 (2017)