Medical Image Analysis 51 (2019) 13–20

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net☆

Jihye Yun[a,#], Jinkon Park[b,#], Donghoon Yu[b], Jaeyoun Yi[b], Minho Lee[a], Hee Jun Park[a], June-Goo Lee[c], Joon Beom Seo[d], Namkug Kim[a,d,*]

[a] Department of Convergence Medicine, Biomedical Engineering Research Center, University of Ulsan College of Medicine, Asan Medical Center, 88 Olympic-Ro 43-Gil Songpa-Gu, Seoul 05505, South Korea
[b] Coreline Soft, Co., Ltd., Sung-Myung B/D 5F, World Cup buk-ro 6-gil 49, Mapo-Gu, Seoul 03991, South Korea
[c] Biomedical Engineering Research Center, University of Ulsan College of Medicine, Asan Medical Center, 88 Olympic-Ro 43-Gil Songpa-Gu, Seoul 05505, South Korea
[d] Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 88 Olympic-Ro 43-Gil Songpa-Gu, Seoul 05505, South Korea

ARTICLE INFO

Article history:
Received 25 September 2017
Revised 8 October 2018
Accepted 18 October 2018
Available online 19 October 2018

Keywords:
Airway
Artificial intelligence
Convolutional neural net
Deep learning
Machine learning
Segmentation

ABSTRACT

We propose a novel airway segmentation method in volumetric chest computed tomography (CT) and evaluate its performance on multiple datasets. The segmentation is performed voxel-by-voxel by a 2.5D convolutional neural net (2.5D CNN) trained in a supervised manner. To enhance the accuracy of the segmented airway tree, we simultaneously took three adjacent slices in each of the orthogonal directions including axial, sagittal, and coronal and fine-tuned the parameters that influence the tree length and the number of leakage. The gold standard of airway segmentation was generated by a semi-automated method using AVIEW^{TM}. The 2.5D CNN was trained and evaluated on a subset of inspiratory thoracic CT scans taken from the Korean obstructive lung disease study, which includes normal subjects and chronic obstructive pulmonary disease patients. The reliability and further practicality of our proposed method was demonstrated in multiple datasets. In eight test datasets collected by the same imaging protocol, the percentage detected tree length, false positive rate, and Dice similarity coefficient of our method were 92.16%, 7.74%, and 0.8997 ± 0.0892, respectively. In 20 test datasets of the EXACT'09 challenge, the percentage detected tree length was 60.1% and the false positive rate was 4.56%. Our fully automated (end-to-end) segmentation method could be applied in radiologic practice.

1. Introduction

Multi-detector computerized tomography (MDCT) offers volumetric images of the airway tree geometry at the sublobal level with submillimeter resolution. Quantifying the peripheral geometry from MDCT images is important for diagnosis and treatment planning of pulmonary diseases involving airway pathology, such as chronic obstructive pulmonary disease (COPD), cystic fibrosis, and interstitial lung diseases (Barnes and Hansel, 2004; Pu et al., 2012). Airway tree segmentation plays an especially important role in pulmonary disease analysis because it quantifies the anatomical features, including the airway wall thickening, wall area, lumen area, wall–lumen area ratio, wall–lumen diameter ratio, and changes in lumen diameter.

Recent review papers on automated airway analysis (Pu et al., 2012; van Rikxoort and van Ginneken, 2013) have overviewed the various kinds of airway segmentation methods. A significant number of these methods perform basic operations on the voxel intensities, such as thresholding or region growing (e.g., as in Fabijasńka (2009); Graham et al. (2010); and van Rikxoort et al. (2009)), which depend on the attenuation differences between the lumen, bronchial wall, and surrounding lung parenchyma. However, in smaller bronchi, the attenuation differences are tempered by the limited resolution and partial vol-

☆ **Conflict of Interest**: No author associated with this paper has disclosed any potential conflict of interest.

* Corresponding author.
E-mail address: namkugkim@gmail.com (N. Kim).
# These authors contributed equally to this work.

https://doi.org/10.1016/j.media.2018.10.006
1361-8415/

ume effects. Consequently, the segmentation often leaks into the lung parenchyma and/or misses small airway branches. Other algorithms are based on morphological or geometrical considerations (e.g., Fabijańska (2009); Fetita et al. (2004); and Graham et al. (2010)), machine learning (e.g., Kitasaka et al. (2010); Lo et al. (2010); and Breitenreicher et al. (2013)), and template matching (e.g., van Rikxoort et al. (2009)). To restrict the segmentation process, these methods usually rely on the appearance of airways in the computed tomography (CT) scans (i.e., dark ellipses in 2D and dark tubes in 3D) or on predefined anatomical rules (e.g., Fan et al. (2000) and Kitasaka et al. (2002)) including (1) adjacency to vessels, (2) low airway intensity, (3) the degree of airway wall existence, (4) absence of a closed loop among the airway branches, (5) no abrupt change in branching angle, and (6) a progressive decrease in diameter. However, these rules and assumptions are easily violated by noise, artifacts, and variations among the anatomy and diseases.

In the 2009 airway segmentation challenge (EXACT'09) (Lo et al., 2012), 15 algorithms were evaluated on a heterogeneous set of 20 CT scans. All participating methods were expected to optimize the trade-off between increasing the airway tree length and reducing the number of leakages. In most methods, this trade-off was determined by optimizing the parameters that influence the two competing factors. When all participating algorithms were combined, they yielded significantly longer airway trees than each algorithm alone, without increasing the leakage. This indicates that different algorithms provide complementary information. In many airway segmentation algorithms, parameter selection is essential for building a decent tree length without leaks (Lo et al., 2012). However, fine-tuning the parameters that influence the tree length and the number of leakages is often a difficult and tedious task and may also depend on the quality of the CT scan. The trade-off usually favors limiting the number of leakages at the cost of detecting fewer smaller bronchi.

Unlike traditional hand-crafted feature-based approaches, deep-learning-based approaches self-learn the informative features directly from a database. Such a data-driven approach might provide an efficient and robust method for solving the above trade-off problem. In particular, convolutional neural net (ConvNets) LeCun et al. (1998), LeCun et al. (2015), Schmidhuber (2015), which uses trainable filter banks with an extensive weight-sharing scheme, can quickly outperform state-of-the-art approaches in many image classification tasks (e.g., Krizhevsky et al. (2012), Sermanet et al. (2014), Simonyan and Zisserman (2014), and Szegedy et al. (2014)). Furthermore, these ConvNets have been employed a powerful framework that is highly suited to airway segmentation. Charbonnier et al. (2017) proposed a multi-view ConvNet for classifying candidate branches as true airways or leakages. Jin et al. (2017) refined the coarse output of 3D ConvNet by incorporating fuzzy connectedness segmentation and the curve skeletonization. Juarez et al. (2018) applied 3D ConvNet in the local volume of interest with data augmentation of elastic deformation.

In this paper, we propose a novel airway segmentation method based on a 2.5 dimensional convolutional neural net (2.5D CNN). To increase the airway tree length and reduce the number of leakages, we took three adjacent slices along each of the axial, sagittal, and coronal directions simultaneously, as detailed in Section 2. The method is evaluated on our own dataset and on the 20 test datasets of the EXACT'09 challenge (Lo et al., 2012), which is an internationally recognized open contest for airway segmentation. This test set comprises a completely independent heterogeneous set of inspiration and expiration scans with different reconstruction kernels and various lung diseases of different severities.

2. Materials and methods

2.1. Patient datasets

For algorithm development, we selected multi-centered thoracic CT scans from the Korean obstructive lung disease (KOLD) cohort, which includes stable patients with obstructive lung disease who were prospectively obtained from the pulmonary clinics of 11 hospitals in South Korea between June 2005 and September 2009. The inclusion criteria of the KOLD cohort are described elsewhere (Lee et al., 2008; Chae et al., 2010; Lee et al., 2011). COPD was diagnosed by airflow limitations that are not fully reversible (post-bronchodilator FEV1/forced vital capacity $<$70% and a smoking history [$>$10 pack-year]). Among the 183 patients who met our diagnostic COPD criteria, we excluded 22 patients with physician-diagnosed asthma. From the remainder, we randomly selected 69 COPD patients who had undergone volumetric CT with the same type of scanner at seven selected hospitals. This study was approved by the institutional review boards of all hospitals, and each patient provided written informed consent. The CT images were represented by a 512×512 matrix, and pixel dimensions were ranged depending on the participant's body size.

The method was validated on two types of clinical datasets; 8 test sets separately selected from our own KOLD data and the EXACT'09 public dataset of 20 test cases. The EXACT'09 datasets were originally designed for comparing the performances of different segmentation algorithms by the same evaluation measures.

2.2. CT protocol

Volumetric CT examinations without broncho-dilatation were acquired within a day of pulmonary function testing. All CT examinations were obtained by a 16-MDCT scanner (Somatom Sensation; Siemens Medical Systems, Erlangen, Germany). The scan parameters were 100 eff. mAs, 140 kVp, 16×0.75 mm collimation, and unity pitch. For this scanner, the scale of the attenuation coefficients ranges from -1024 to 3072 HU. The patients, who had been pretrained to reach full inspiration and expiration, were instructed to hold their breath at full inspiration and expiration just before the scan. The acquired data were reconstructed using a standard kernel at 0.75-mm thickness with 0.75-mm increments. All CT scanners were calibrated monthly after maintenance using water as the standard phantom and calibrated daily using air. All scans were obtained within 24 hours after air calibration. The imaging data were stored in the Digital Imaging and Communications in Medicine format.

2.3. Gold standard of airway segmentation

Except for the reference dataset (the 20 test datasets of the EXACT'09 challenge), all images in the development and clinical validation sets were semiautomatically segmented. We segmented the airways in the CT data of 77 patients (59 patients (76.6%) for training, 10 patients (13%) for validation, and 8 patients (10.4%) for testing) using the airway segmentation module of AVIEW^{TM} (Coreline Soft Inc., Seoul, South Korea) (see Fig. 1), which segments the airway regions by Hough transform, a gray-scale morphological operation, and tubular structure detection. After segmentation, the isolated and/or disconnected airway candidates were specified by an experienced research assistant, which usually required 1–2 hours. Finally, the segmentation result was confirmed by a thoracic radiologist with 15 years' experiences. Generating the gold standard dataset of airway segmentation is extremely difficult, tedious, and time-consuming, thus we collected the gold standard segmentation mask with the aim of a minimum amount of dataset that can be

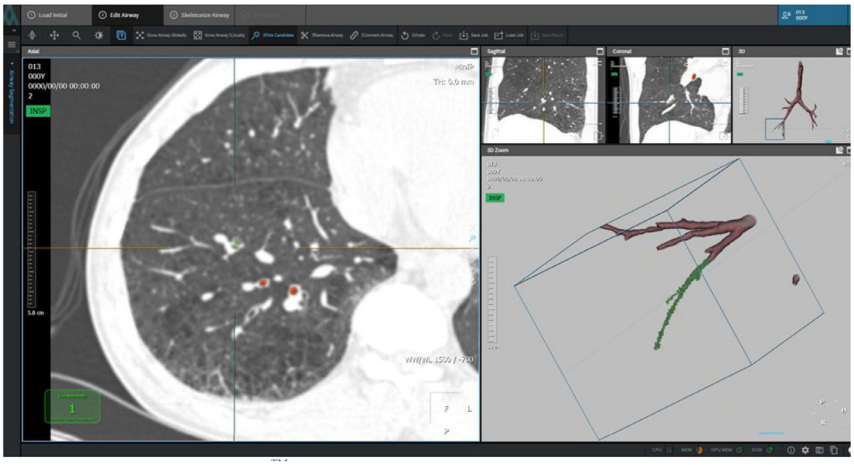

Fig. 1. Airway segmentation module of AVIEW^{TM} for generating the gold standard segmentation mask. After segmentation using this module, the isolated and disconnected airway candidates must be identified by semiautomatic selection, which takes 1–2 hours by an expert. The segmentation result is confirmed by a thoracic radiologist with more than 10 years' experiences.

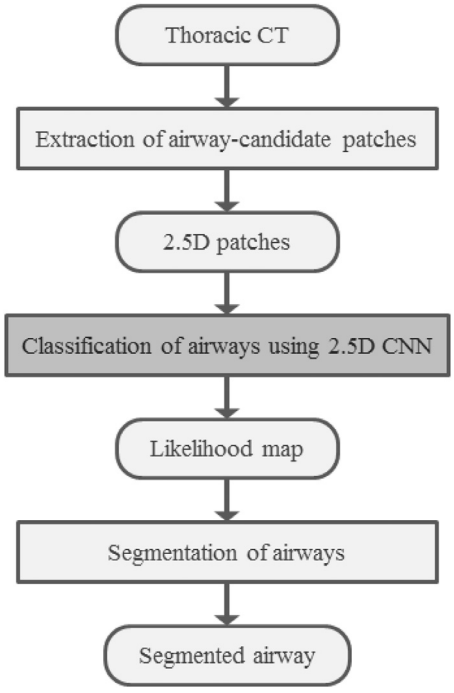

Fig. 2. Process of deep-airway segmentation using 2.5D convolutional neural net (2.5D CNN).

trained and tested effectively (total 77 CT scans as mentioned before).

2.4. Deep learning with convolutional neural net

Our system proceeds through three main steps (see Fig. 2). To enhance the accuracy of the airway segmentation, we first selected the airway-candidate patches generated in 2.5D. From the selected 2.5D airway-candidate patches, we then trained the 2.5D CNN to classify the center pixel of each patch as airway or non-airway, thus producing a likelihood map. Finally, the segmented airways were obtained from the likelihood map by connected component labeling with thresholding.

2.4.1. Extraction of 2.5D airway-candidate patches

Our method segments the airways by voxel-by-voxel classification based on the patches around each voxel. To train the designed segmentation network, we selected sample points, and extracted patches from them. For the extraction of these airway-candidate patches, the considerations about airway and non-airway regions were determined by two thoracic radiologists with more than 20 years' experiences. Since increasing the true tree length and reducing the number of leakages are crucial factors for the airway segmentation, they decided to extract the airway-candidate patches from airway and non-airway regions. Furthermore, to distinguish airway and non-airway regions, the surrounding regions of the airway were divided into five kinds of sub-regions according to the distance from the airway regions. Using the patches extracted with this strategy, our proposed network could learn the overall structures of the airway regions and differentiate between airway regions and non-airway regions including airway wall, parenchyma, pulmonary vessels, etc.

We selected 2,250 sample points per one thoracic CT scan (1,000 and 1,250 sample points from the airway and non-airway regions, respectively). The airway regions of the training datasets were based on the gold standard. The non-airway regions were divided into five kinds of sub-regions at different distances by measuring the 3D Euclidean distance from the airway regions (1 voxel, 2–4 voxels, 5–7 voxels, 8–10 voxels, and 11–30 voxels), and then we selected 250 sample points from each non-airway subregion, obtaining 1,250 sample points in total. From selected sample points, we extracted airway-candidate patches P_i^j which are from three adjacent slices in each of the orthogonal directions (axial, sagittal, and coronal), which is intending to increase the length of the segmented airway tree while reducing the number of leakages. Since these airway-candidate patches are two-dimensional images but contain constrained three-dimensional information, we called them as 2.5D patches. The patch size was fixed at 32×32 pixels, ensuring complete capture of the surrounding information. Consequently, nine patches of 32×32 pixels were generated at each sample point; three patches in each of the axial, sagittal, and coronal directions include the adjacent slices located at -1, 0, and $+1$ voxels from the sample point. To train the 2.5D CNN, we were extracted total number of 1,397,250 patches (training: 1,194,750 patches from 59 patients, validation: 202,500 patches from 10 patients), which were evenly distributed from 69 patients. The patches around each sample point were resampled using a lung windowing setting with level -700 HU and width 700 HU and were directly input to the subsequent classification step of the 2.5D CNN.

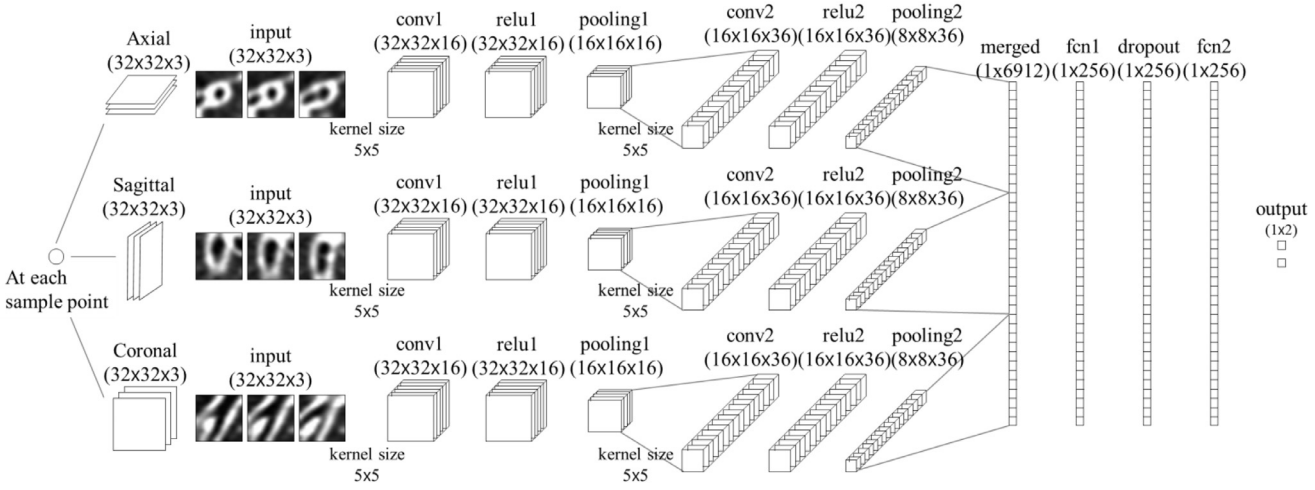

Fig. 3. Architecture of 2.5D CNN for airway segmentation. A patch group in each direction is passed through one of three separate identical stacks of convolution and max-pooling layers. Three parallel stacks are combined into the first fully connected layer (256 units). To prevent overfitting, a connection dropout probability of 0.5 was added prior to the second fully connected layer. Finally, the probabilities of being part of the airway are calculated by the softmax function, which is applied to the output of the last fully connected layer.

2.4.2. Classification of airways by the 2.5D CNN

Our 2.5D CNN was designed and trained to classify an airway-candidate point into an airway or a non-airway region. A single sample point was represented by nine 32×32 pixel patches (three patches in each of the axial, sagittal, and coronal directions). Similarly to Setio et al. (2016), the 2.5D CNN was designed by optimizing several parameters that define the architecture of the network, i.e., the number of convolution layers, number and size of the filters in each layer, and the dropout and learning rates. These parameters were optimized by training multiple 2.5D CNNs with different sets of parameters and evaluating the performance by the area under curve of the receiver operating characteristic of the exclusive validation set. For tuning the hyper parameters and an unbiased evaluation of 2.5D CNN, we selected 10 random patients (13%) from the datasets. The final architecture of the 2.5D CNN is outlined in Fig. 3.

To classify an airway-candidate sample point, the patch groups along the axial, sagittal, and coronal direction were processed in parallel. The patch group in each direction was passed through one of three separate identical stacks of convolution and max-pooling layers. The convolution layers comprised a first convolution layer with 16 learnable filters and a second convolution layer with 36 learnable filters. All filters in both layers were sized at 5×5 pixels. Each convolution layer is followed by a rectified linear unit activation and a max-pooling operation. Three parallel stacks of the orthogonal directions were concatenated using a merged layer (6,912 units), followed by two fully connected layers. To prevent overfitting, a connection dropout probability of 0.5 was added prior to the second fully connected layer (Srivastava et al., 2014). Finally, the probabilities of airway regions were calculated by the softmax function applied to the output of the last fully connected layer and stored in a likelihood map. The designed 2.5D CNN was implemented in publicly available Tensorflow v0.10.0 (Google Inc., Mountain View, CA, USA).

Given the training datasets, the 2.5D CNN was learned by minimizing the training error namely, the cross-entropy error between the inferred probability and the gold standard label by stochastic gradient descent (SGD). The cross-entropy cost function in the airway segmentation is given by the following:

$$L(y,\ f) = -y \log f - (1-y) \log (1-f), \tag{1}$$

where f and y denote the inferred probability and the corresponding desired output, respectively. The performance of SGD was improved by applying the RMSProp heuristic (Hinton, 2012), and the weights were updated using mini-batches of 512 samples. The training error was minimized by running the back-propagation algorithm over 353 training epochs. The network configuration giving the highest accuracy on the validation set was selected as the optimal configuration.

2.4.3. Segmentation of airways using connected component labeling with thresholding

From the airway-candidate patches classified by the proposed 2.5D CNN, we constructed a likelihood map containing the necessary information for the airway segmentation. The likelihood map was calculated by the softmax function applied to the output of the last fully connected layer. For the binary classification, we considered voxels with higher likelihood values (≥ 0.5) as discriminative airway voxels. Since the softmax function is a logistic function that squashes a k-dimensional vector z of arbitrary real values to a k-dimensional vector $\sigma(z)$ values in the range [0, 1] that add up to 1, the threshold value 0.5 is a reasonable value to discriminative between airways and non-airways. After applying connected component labeling to these selected voxels and selecting the largest component, we obtained the final segmented airways.

2.4.4. Optimization of airway classification

An unseen thoracic CT image can be straightforwardly segmented by evaluating the trained 2.5D CNN on every patch in the net. This technique applies a sliding window to a patch around the pixel. The drawbacks of this approach are twofold: 1) all pixels must be invoked and 2) there are many redundant computations due to the overlapping of different patches. Both of these problems were solved by the following strategy.

Our segmentation process begins with the fully automated segmentation method for initial airway mask by using the region growing method and simple thresholding by −950 HU which is referred value from Reinhardt et al., 1997, because main bronchus could be easily segmented and we want to train our deep learning method to focus on the small airway branches which are very hard to be segmented. In addition, this initial segmentation could reduce many redundant computations due to patch overlapping, improving the performance. As the initial airways by using the region growing method have short tree length (see Fig. 5(b)), we extracted airway-candidate patches around their ends and extended them with high probability. Therefore, we extracted airway-candidate

Table 1
Performance measures of EXACT'09 (Lo et al., 2012).

Measurement	Description
Branch count	The number of branches that are detected correctly
Branch detected	The fraction of branches that are detected, with respect to the total number of branches present in the reference
Tree length	The sum of the length of the centerlines of all correctly detected branches
Tree length detected	The fraction of tree length that is detected correctly relative to the reference
Leakage count	The number of disconnected sources where leakage occurs (26 connectivity)
Leakage volume	The volume of regions that are not marked as "correct" in the reference
False positive rate	The fraction of total segmented volume that is not marked as "correct" relative to the reference

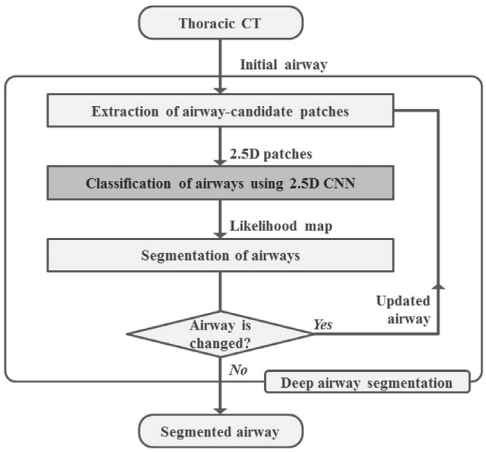

Fig. 4. Process of optimized deep-airway segmentation.

patches from the sample points located six voxels (heuristically selected by thoracic radiologists) from the end of the initial airways and classified them as airway or non-airway by the trained 2.5D CNN. The airway mask was updated by using only the voxels with high likelihood values (≥ 0.5) connected to the initial airway mask. In addition, a voxel with a high likelihood value (≥ 0.95) was updated as an airway even when not connected to the initial airway mask. The updated airway mask was input to the extraction process of the airway-candidate patches of next iteration step. The proposed deep-airway segmentation iteratively updates the initial airway mask until the mask is unchanged between two successive updates. Fig. 4 shows the processing steps during each update. After this initial segmentation on main bronchus of airway, we could change the segmentation into the classification problem. This optimization by iterative probability-based patching on the end area of previous airway mask dramatically improved the performance from 45–70 min to 2–8 min (2 min in all except two cases).

2.5. Evaluation

The performance of the proposed deep-airway segmentation method was evaluated on two separate independent datasets; our own test data and data from the EXACT'09 challenge. In the case of our own test data, each segmentation result was evaluated and compared with the gold standard. For a quantified comparison, we computed the Dice similarity coefficient (DSC) and the performance measures of EXACT'09 (Lo et al., 2012). The DSC evaluates both the reproducibility performance and the spatial overlap accuracy of the segmentations. The EXACT'09 challenge is no longer maintained and new submissions are not processed. Therefore, to compare our results to those of the other competitors in the challenge, we requested the reference standard from the organizers of the challenge and re-implemented the evaluation algorithm (as described in Lo et al. (2012)). For consistency with the airway segmentation methods participating in the challenge, our method was evaluated by the seven performance measures of EXACT'09 (Lo et al., 2012) described in Table 1.

3. Experiments and results

3.1. Experimental setup

The proposed 2.5D CNN was trained on the inspiratory thoracic CT scans of 69 patients (59 patients for training and 10 patients for validation). During the training, the maximum validation accuracy was reached after 353 epochs which took about 2.5 days on an NVIDIA GeForce GTX 1080 with 8 GB of memory. Our method was validated on our own dataset and the EXACT'09 challenge data.

3.2. Experiment 1: our own test data

In the first experiment, the proposed deep-airway segmentation method was validated on our own dataset. Table 2 summarizes the emphysema indices and the computed performance measures in the eight datasets. For a quantitative evaluation of the segmented airways, we computed the DSC and performance measures of EXACT' 09 (Lo et al., 2012), which imply the tree length and leakage of the segmented airway, and compared them with the manual segmentation results using AVIEW^{TM}. The gold standard was established by the same process to that of the training datasets.

From the DSC, the average spatial overlap was determined as 0.8997 ± 0.0892, and the value was lower for severe COPD patients with large low-attenuation areas. As the emphysema index of the patient increased, the other measures were lower. Given that the DSC is more sensitive to the thin tubular structures of airways than other measures, the good reproducibility performance of our method was confirmed. Although this clinical validation dataset was substantially different from the training dataset, our method achieved a high sensitivity (detected tree length = 92.16%) at a low false positive rate (2.63%). Even if excluding the trachea and the main bronchus, we could get a low false positive rate (7.74%). Fig. 6 shows the results of all eight cases. For a single thoracic CT, our deep-airway segmentation required 1–2 min on a standard PC with a GPU GeForce GTX1080.

3.3. Experiment 2: EXACT'09 challenge data

In the second experiment, the performance of our method was compared with the performances of the 15 algorithms in the EXACT'09 challenge (Lo et al., 2012). The airway segmentation algorithms in EXACT'09 were evaluated on a heterogeneous set of 20 CT scans, including inspiration and expiration scans with different reconstruction kernels and interstitial lung diseases with different severities. The participating methods were evaluated by seven performance measures, as mentioned in subSection 2.5. For a fair comparison, we evaluated our method by the same measures.

Table 3 summarizes the computed performance measures in the 20 test cases of EXACT'09. Fig. 7 compares the performances of the proposed deep-airway segmentation method and the 15 algorithms of EXACT'09. The detected tree length and false positive rate of

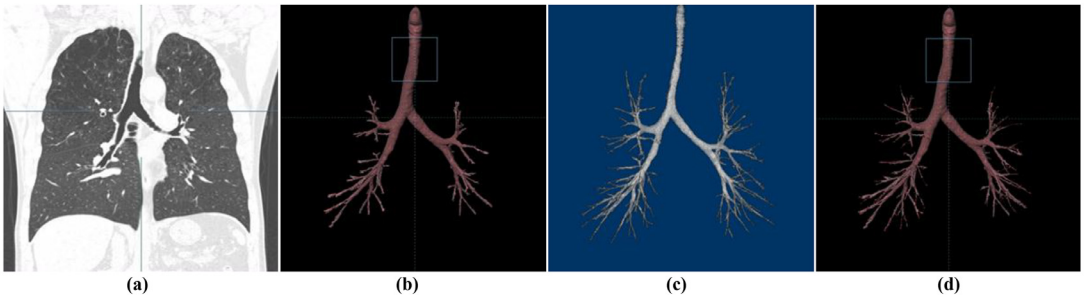

Fig. 5. Optimized deep-airway segmentation. (a) A thoracic CT scan. (b) Initial airway. (c) Result of 2.5D CNN. (d) Result of manual segmentation.

Table 2
Performance of the proposed deep-airway segmentation evaluated on our own test data.

Data set	Emphysem a index	DSC	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage volume (mm^3)	False positive rate (%)	
								With trachea and main bronchus	Without trachea and main bronchus
1	43.4	0.7921	179	83.33	290.1	86.97	9520.5	1.74	6.78
2	18.3	0.8856	337	84.97	425.0	89.76	3597.8	3.50	9.37
3	9.9	0.9750	225	90.42	346.9	93.94	2178.8	2.57	5.98
4	10.6	0.7813	318	88.13	444.8	90.24	13,736.9	2.19	7.25
5	19.8	0.9789	209	96.45	313.7	97.32	3132.8	3.51	7.80
6	47.5	0.8268	244	90.26	333.1	93.03	20,850.7	2.4	13.83
7	5.2	0.9852	179	93.64	248.7	93.34	15,312.0	1.74	4.40
8	8.1	0.9727	247	91.28	322.0	92.65	2004.9	2.82	6.49
		0.8997 ± 0.0892		**89.81** ± 4.31		**92.16** ± 3.13		**2.63**	**7.74**

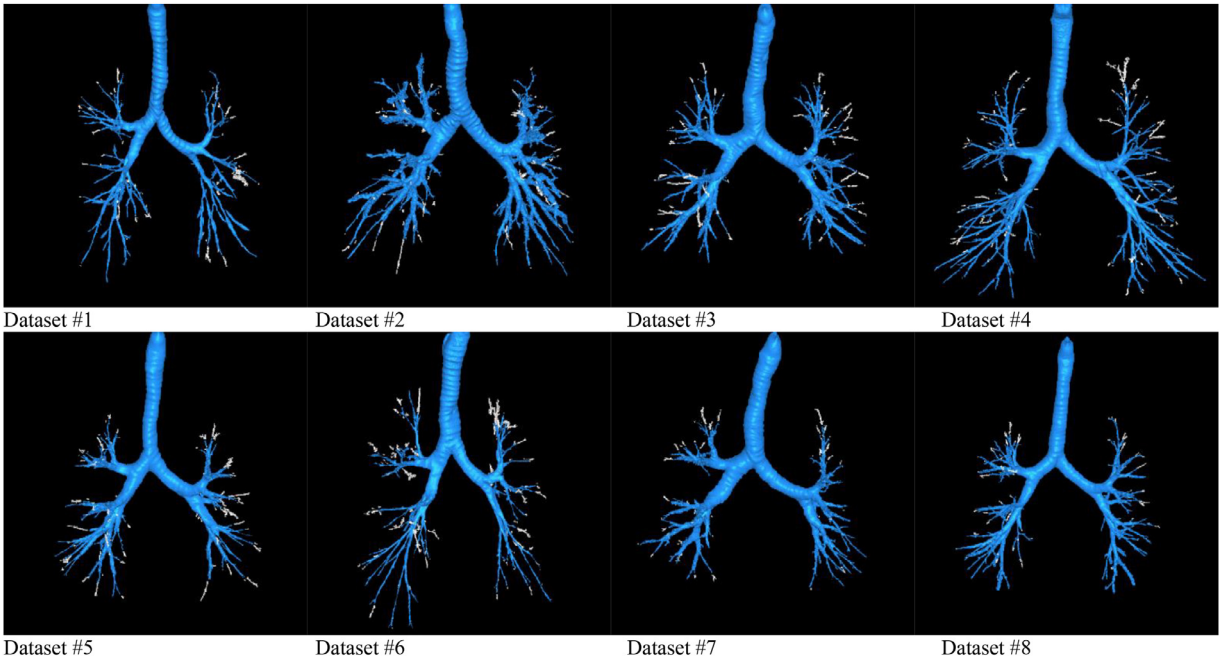

Fig. 6. Results of the first experiment showing eight evaluations on our own test data. Blue regions were manually segmented by a thoracic radiologist with more than 10 years of clinical experiences. White parts represent potential leakage regions that were falsely segmented by our method. Each result is rendered in anterior–posterior view. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

Table 3
Performance of the proposed deep-airway segmentation on EXACT'09 challenge data.

	Branch count	Branch detected (%)	Tree length (cm)	Tree length detected (%)	Leakage count	Leakage volume (mm^3)	False positive rate (%)
Mean	163.4	65.7	129.3	60.1	94.1	726.4	4.56
SD	79.4	13.1	66.0	11.9	61.9	779.0	3.73
Min	28	35.0	21.5	32.7	6	13.0	0.39
1st quartile	112	59.4	83.5	52.3	44	132.4	1.75
Median	153	67.4	118.8	58.5	88	376.1	3.24
3rd quartile	263	79.9	210.6	75.7	159	1434.3	8.28
Max	310	87.2	232.5	77.7	209	2662.3	12.56

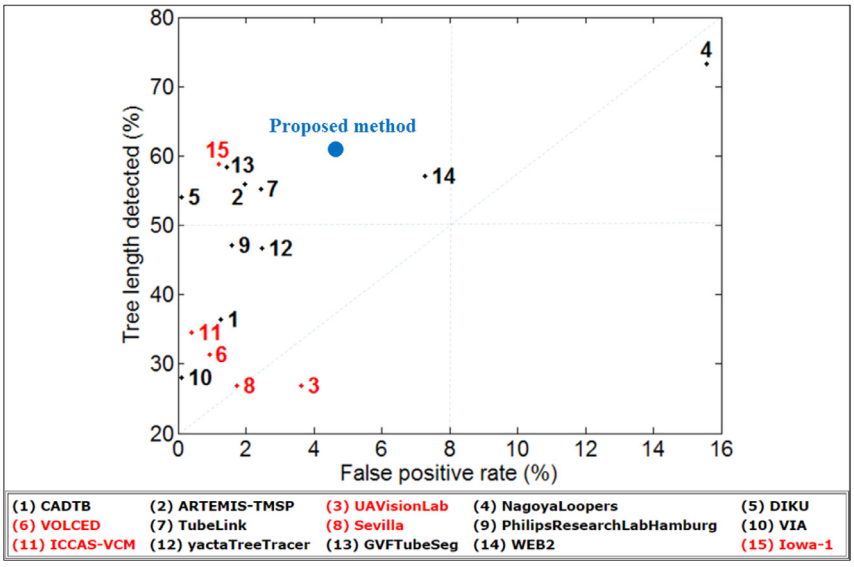

Fig. 7. Comparison of accuracy in our method and in 15 algorithms of EXACT'09. Average false positive rate versus detected tree length (red denotes a semiautomated method). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

our method were 60.1% and 4.56%, respectively. Since the reference datasets were not publically available in case of the 20 test cases of the EXACT'09 challenge, we could not redefine the ground-truth airway regions and recalculate the accuracy. However, most of the airways resulting from our method were confirmed as true airways by a thoracic radiologist with more than 10 years of clinical experiences, and thus our false positive rate could be lower than that from EXACT'09 evaluation. Our method was outperformed by the leakage reduction algorithm of Charbonnier et al. (2017), with a detected tree length of 65.4% and a false positive rate of 1.68%. However, our fully automated (end-to-end) segmentation method lowers the execution time to 2–8 min (2 min in all except 2 cases), without any additional processing or user interactions.

4. Discussion and conclusion

Airway segmentation in volumetric CT is among the most challenging tasks in medical image processing. The segmentation process must overcome the limited spatial resolution of CT, complexity of the airway tree, partial volume effect of tapering vessels, and discontinuities caused by bodily processes such as heartbeat. In addition, the segmentation results are usually widely different from a radiologist's assessment. This discrepancy might arise from the robustness of the human eye and brain, and incompleteness of medical image processing. However, deep learning is believed to mimic the human visual cortex in an artificial way (Kheradpisheh et al., 2016). Therefore, deep learning is expected to improve the diagnostic accuracy of radiologists by delivering a quantitative analysis of suspicious lesions. It also provides automatic report generation and voice recognition, both of which will benefit the clinical workflow and shorten the reading time. In this study, we explored the applicability of this deep learning technique to airway segmentation.

To ensure the accuracy and generalizability of our deep-airway segmentation method, we validated the method on multiple datasets. The training dataset consisted of multi-centered datasets chosen from the KOLD cohort, confirming that the proposed method can handle variations of datasets. In the 20 test cases of the EXACT'09 challenge, our method achieved high sensitivity but also a rather high false positive rate. This was mainly attributed to the sub-ideal quality of the given reference standard. In the evaluation of our own dataset, the false positive rate might arise from the fundamental limitation of human interpretation. In both cases, most of the false positives were confirmed as true airways by the thoracic radiologist. Overall, the multiple-dataset validation demonstrated the reliability and further practicality of the proposed method.

Computerized analyses related to the human airway require accurate identification of the airway tree. Semi-automated detection of a human airway tree is possible in theory but typically infeasible in practice; therefore, fully automated methods are highly desired. By training end-to-end model on CT scans, we predicted the segmentation of the entire volume at once. This fully automated segmentation method provides a robust segmentation that is free from user bias.

However, for severe COPD patients with large low-attenuation areas, the performance of our model degrades and the execution time varies despite the optimization. To improve the robustness of our approach to disease severity, we could adopt the 3D ConvNet (Milletari et al., 2016), which entirely captures the 3D information of airways in its 3D convolutional layers.

Acknowledgment

This work was supported by the Technological Innovation R&D Program (S2464035) funded partially by the Small and Medium Business Administration (SMBA, Korea) and by the Industrial Strategic technology development program (10072064) funded by the Ministry of Trade Industry and Energy (MI, Korea).

References

Barnes, P.J., Hansel, T.T., 2004. Prospects for new drugs for chronic obstructive pulmonary disease. Lancet 364 (9438), 985–996.
Breitenreicher, D., Sofka, M., Britzen, S., Zhou, S.K., 2013. Hierarchical discriminative framework for detecting tubular structures in 3D images. IPMI 23, 328–339.
Chae, E.J., Seo, J.B., Song, J.W., Kim, N., Park, B.W., Lee, Y.K., Oh, Y.M., Lee, S.D., Lim, S.Y., 2010. Slope of emphysema index: an objective descriptor of regional heterogeneity of emphysema and an independent determinant of pulmonary function. Am. J. Roentgenol. 194 (3), W248–W255.
Charbonnier, J.P., van Rikxoort, E.M., Setio, A.A., Schaefer-Prokop, C.M., van Ginneken, B., Ciompi, F., 2017. Improving airway segmentation in computed tomography using leak detection with convolutional networks. Med. Image Anal. 36, 52–60.
Fabijasńka, A., 2009. Two-pass region growing algorithm for segmenting airway tree from MDCT chest scans. Comput. Med. Imaging Graph. 33, 537–546.
Fan, L., Chen, C.W., 2000. Reconstruction of Airway Tree Based On Topology and Morphological Operations. SPIE Medical Imaging, SPIE, San Diego, CA, pp. 46–57.

Fetita, C.I., Prêteux, F., Beigelman-Aubry, C., Grenier, P., 2004. Pulmonary airways: 3D reconstruction from multislice CT and clinical investigation. IEEE Trans. Med. Imaging 23, 1353–1364.
Graham, M.W., Gibbs, J.D., Cornish, D.C., Higgins, W.E., 2010. Robust 3D airway tree segmentation for image-guided peripheral bronchoscopy. IEEE Trans. Med. Imaging 29, 982–997.
Hinton, G., 2012. Rmsprop: divide the gradient by a running average of its recent magnitude. Neural Netw. Mach. Learn., Coursera lecture 6e.
Jin, D., Xu, Z., Harrison, A.P., George, K., Mollura, D.J., 2017. *3D convolutional neural networks with graph refinement for airway segmentation using incomplete data labels*. In: International Workshop on Machine Learning in Medical Imaging, pp. 141–149.
Juarez, A.G.U., Tiddens, H.A.W.M., de Bruijne, M., 2018. Automatic airway segmentation in chest CT using convolutional neural networks. In: Image Analysis for Moving Organ, Breast, and Thoracic Images, pp. 238–250.
Kitasaka, T., Yano, H., Feuerstein, M., Mori, K., 2010. Bronchial region extraction from 3D chest CT image by voxel classification based on local intensity structure. In: Proc. Third International Workshop on Pulmonary Image Analysis, pp. 21–29.
Kitasaka, T., Mori, K., Hasegawa, J., Toriwaki, J., 2002. A method for extraction of bronchus regions from 3D chest X-ray CT images by analyzing structural features of the bronchus. FORMA-TOKYO- 17 (4), 321–338.
Kheradpisheh, S.R., Ghodrati, M., Ganjtabesh, M., Masquelier, T., 2016. Deep networks can resemble human feed-forward vision in invariant object recognition. Sci. Rep. 6, 32672.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, 25, pp. 1097–1105.
LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Lee, J.S., Huh, J.W., Chae, E.J., Seo, J.B., Ra, S.W., Lee, J.H., Kim, E.K., Lee, Y.K., Kim, T.H., Kim, W.J., Lee, J.H., Lee, S.M., Lee, S., Lim, S.Y., Shin, T.R., Yoon, H.I., Sheen, S.S., Oh, Y.M., Lee, S.D., 2011. Predictors of pulmonary function response to treatment with salmeterol/fluticasone in patients with chronic obstructive pulmonary disease. J. Korean Med. Sci. 26 (3), 379–385.
Lee, Y.K., Oh, Y.M., Lee, J.H., Kim, E.K., Lee, J.H., Kim, N., Seo, J.B., Lee, S.D., 2008. Quantitative assessment of emphysema, air trapping, and airway thickening on computed tomography. Lung 186 (3), 157–165.
Lo, P., Sporring, J., Ashraf, H., Pedersen, J.J.H., de Bruijne, M., 2010. Vessel-guided airway tree segmentation: a voxel classification approach. Med. Image Anal. 14, 527–538.
Lo, P., van Ginneken, B., Reinhardt, J.M., Tarunashree, Y., de Jong, P.A., Irving, B., Fetita, C., Ortner, M., Pinho, R., Sijbers, J., Feuerstein, M., Fabijanska, A., Bauer, C., Beichel, R., Mendoza, C.S., Wiemker, R., Lee, J., Reeves, A.P., Born, S., Weinheimer, O., van Rikxoort, E.M., Tschirren, J., Mori, K., Odry, B., Naidich, D.P., Hartmann, I.J., Hoffman, E.A., Prokop, M., Pedersen, J.H., de Bruijne, M., 2012. Extraction of airways from CT (EXACT'09). IEEE Trans. Med. Imaging 31, 2093–2107.
Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. Fourth International Conference on 3D Vision (3DV). arXiv 1606.04797.
Pu, J., Gu, S., Liu, S., Zhu, S., Wilson, D., Siegfried, J.M., Gur, D., 2012. CT based computerized identification and analysis of human airways: a review. Med. Phys. 39 (5), 2603–2616.
Reinhardt, JosephM., D'Souza, N., Hoffman, EricA., 1997. Accurate measurement of intrathoracic airways. IEEE Trans. Med. Imaging 16 (6), 820–827.
van Rikxoort, E.M., van Ginneken, B., 2013. Automated segmentation of pulmonary structures in thoracic computed tomography scans: a review. Phys. Med. Biol. 58 (17), 187–220.
van Rikxoort, E., Baggerman, W., van Ginneken, B., 2009. Automatic segmentation of the airway tree from thoracic CT scans using a multi-threshold approach. In: The Second International Workshop on Pulmonary Image Analysis, pp. 341–349.
Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117.
Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., 2014. OverFeat: Integrated recognition, localization and detection using convolutional networks. In: International Conference on Learning Representations (ICLR 2014), 1312. arXiv, p. 6229.
Setio, A.A.A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., van Riel, S., Wille, M.W., Naqibullah, M., Sanchez, C., van Ginneken, B., 2016. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans. Med. Imaging 35, 1160–1169.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv 1409, 1556.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., *Going deeper with convolutions*. 2014. arXiv:1409.4842.