Lung segmentation on HRCT and volumetric CT of diffuse interstitial lung disease using U-Net

Beomhee Park[1] **Heejun Park**[1] **Joon Beom Seo**[1,2] **Sang Min Lee**[1,2] **Namkug Kim**[1,3*]
[1]University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea
[2]Department of Radiology
[3]Department of Convergence Medicine
{beomheep, heejun.park1129, joonbeom.seo,
sangmin.lee.md, namkugkim}@gmail.com

Abstract

Building the gold standard for training a deep-learning-based model is a time-consuming and very expensive task, especially in medicine. In the case of lung segmentation, blurred boundaries from various types of pulmonary diseases make the problem more difficult. In general, the accurate and robustness model needs a large size of dataset, cost and time effective labeling process is vital. Therefore, we proposed a method to generate the gold standard efficiently based on initial segmentation using a conventional method for lung segmentation with various types of diffuse interstitial lung disease (DILD). The accuracy and robustness of the deep-learning-based segmentation method trained with these data were evaluated and compared with the conventional method using different protocols, including high-resolution computed tomography (HRCT) and volumetric CT.

1 Introduction

Computed tomography (CT) has become an important tool for diagnosing pulmonary diseases. In particular, high-resolution computed tomography (HRCT) and volumetric CT are recognized as effective diagnostic tools for diffuse interstitial lung disease (DILD) as they help in detecting chronic changes in lung parenchyma [1]. Many studies have therefore developed computerized image analysis and computer-aided diagnosis techniques that utilize HRCT scans to more effectively diagnose pulmonary diseases, including DILD.

When analyzing images of diseased organs, segmenting the organ is a critical step that usually precedes the main image analysis. If segmentation errors cause the organ's borders to be set incorrectly, this is very likely to affect the subsequent analysis. Particularly in lung segmentation, it can be hard to identify the lung borders of patients with pulmonary diseases as such diseases further reduce the distinction between their lung tissue and the surrounding structures.

Although a deep-learning based model is considered as one of the segmentation methods to solve these difficulties, it requires a large size of dataset to train. Therefore, we proposed an efficient labeling method and evaluated a convolutional neural net (CNN)-based segmentation model that trained the generated data using two clinical protocols including HRCT and volumetric chest CT of DILD which suffer from blurred boundaries for various types of lung parenchyma disorders.

2 Material

In this study, we created a dataset consisting of HRCT images of 619 patients, including 211 with cryptogenic organizing pneumonia (COP), 210 with usual interstitial pneumonia (UIP), and 196 with

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

nonspecific interstitial pneumonia (NSIP), together with independent volumetric CT images of 30 patients with UIP. The HRCT images used to train the model were scanned by GE and Siemens with 1–2 mm thick and inter-slice intervals of 5–10mm, and the volumetric CT images scanned by Siemens with submilimeter thickness without intervals were used to provide additional extra-validation.

3 Methods

The overall procedure consists of three phases. Since identifying and segmenting lung regions with DILD from scratch are a time-consuming and expensive task, we instead took a two-step approach. First, lung segmentation in HRCT was performed using a conventional image processing method. Although this provided generally reasonable segmentation results for normal lungs, it was not appropriate for severe DILD cases. A thoracic radiologist therefore manually corrected the lung regions in axial images to create gold standards. Then, the deep-learning model was trained using these gold standards.

3.1 Building Gold Standard

In general, the CT values (HU) for normal lung tissues are relatively low compared with those for other tissues due to the air in the lungs. A threshold-based method was therefore employed for the initial segmentation as it was able to extract the lung region easily and effectively. The Insight Segmentation and Registration Toolkit (ITK) was used to apply rolling-ball and hole-filling operations to smooth and fill the lung region, as well as other lung tissues such as major vessels and airways in the lung's interior [2]. Based on the initial segmentation, a thoracic radiologist more easily corrected the lung region in the axial direction, for creating gold standards.

3.2 Generalized Segmentation using U-Net

U-Net [3] is one of the most widely used CNN architectures for image segmentation. Its network can be divided into two main parts including the left side reduces the dimensionality and the right side extends the original dimensionality. The most important characteristic of U-Net is a connection (concatenation function) between the left and right sides, which allows to give more accurate segmentation results by preventing it from losing information. We used this network for our generalized lung segmentation model and modified its structural hyper-parameters to accept 512 x 512 inputs.

3.3 Experimental Results

The HRCT of 79 patients and volumetric CT images of 30 patients were used for testing. To accurately evaluate performance of our lung segmentation method, we compared the conventional (CM) and deep-learning-based methods (DL) with the gold standard using four metrics including the Dice similarity coefficient (DSC), Jaccard similarity coefficient (JSC), mean surface distance (MSD), and Hausdorff surface distance (HSD). Table 1 compares the results for both methods (CM and DL) for

Table 1: DILD segmentation results, compared with the manual correction baseline
((a): HRCT, (b): volumetric CT, CM: conventional method, DL: deep-learning method)

			DSC (%)	JSC (%)	MSD (mm)	HSD (mm)
(a)	COP	CM	97.28 ± 3.14 *	94.88 ± 5.72 *	0.99 ± 0.99 *	37.75 ± 18.61
		DL	98.79 ± 0.70	97.62 ± 1.36	0.34 ± 0.24	31.39 ± 18.16
	UIP	CM	98.21 ± 2.35	96.58 ± 4.28	0.63 ± 0.71 *	33.54 ± 13.81 *
		DL	98.84 ± 0.45	97.70 ± 0.87	0.24 ± 0.13	23.45 ± 9.15
	NSIP	CM	98.28 ± 1.44	96.66 ± 2.76	0.67 ± 0.62 *	32.55 ± 15.48 *
		DL	98.92 ± 0.42	97.87 ± 0.83	0.21 ± 0.09	19.93 ± 6.08
	Total	CM	97.88 ± 2.53 **	95.97 ± 4.64 **	0.77 ± 0.82 **	34.84 ± 16.34 **
		DL	98.84 ± 0.55	97.71 ± 1.07	0.27 ± 0.18	25.47 ± 13.63
(b)	UIP	DL	98.29 ± 0.56	96.65 ± 0.11	0.80 ± 0.21	26.57 ± 11.26

paired t-test (*p-value < 0.01, **p-value < 0.001)

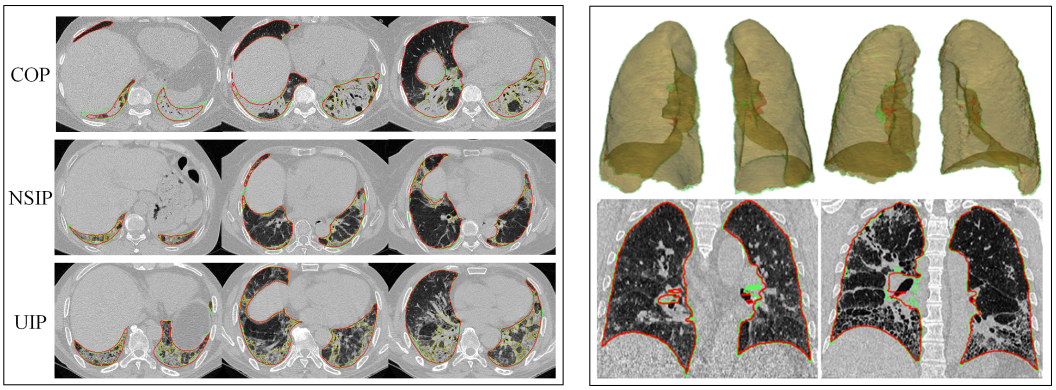

Figure 1: Segmentation results of deep-learning-based method (DL) with U-net at the lowest DSC values for HRCT (left) and volumetric CT (right)
(green: glod standard; yellow: CM; red: DL)

all four metrics on the test dataset. Since the gold standards were based on the CM results, its results would be correlated accordingly. However, DL showed better results for all disease classes than CM, with smaller standard deviations in every case, which means good generalization. The DSC results showed that the method was most accurate for NSIP and least accurate for COP. Although the overall results were significantly different for all four metrics ($p < 0.001$), only the DSC, JSC, and MSD were significantly different ($p < 0.01$) for COP, and the MSD and HSD were significantly different for UIP and NSIP.

4 Discussion and Conclusion

Segmentation of lungs with DILD is difficult because of the large lung parenchyma variability among patients, including irregular tissue patterns on the lung images. We have efficiently created gold standards from conventional image processing results and used them to train the U-net-based segmentation model. Our method demonstrated performance improvements in terms of four different accuracy metrics, for patients scanned with two different clinical protocols including HRCT and volumetric CT. In addition, even in case of the lowest DSC scores due to severe parenchymal disorders, there were no critical errors and the lung region was well segmented, as shown in Figure 1. The main differences occurred in the hilum area of a lung that was not clearly defined even in the gold standard image. This model still needs to be applied and validated to a wide range of chest CT scans for additional diseases, different vendors and different protocols from multi-centers.

Acknowledgments

This work was supported by the Industrial Strategic technology development program (10072064) funded by the Ministry of Trade Industry and Energy (MI, Korea) and by Kakao and Kakao Brain corporations.

References

[1] Teri J Franks, Jeffrey R Galvin, and Aletta Ann Frazier. The use and impact of hrct in diffuse lung disease. *Current Diagnostic Pathology*, 10, 2004.

[2] Shiying Hu, Eric A. Hoffman, and Joseph M. Reinhardt. Automatic lung segmentation for accurate quantitation of volumetric x-ray ct images. *IEEE Transaction on Medical Imaging*, 20, 2001.

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.