Data-efficient multi-label CT segmentation based on shared context learning in multi-view inputs

Yu Jin Seol[1,2] and Jae Sung Lee[1,2,3]
Interdisciplinary Program in Bioengineering, Seoul National University Graduate School, Seoul, Korea[1]
Department of Nuclear Medicine, Seoul National University College of Medicine, Korea[2]
Brightonix Imaging Inc., Seoul, Korea[3]

***Abstract*— A computer-aided diagnostic (CAD) systems for diagnosing diseases and detecting lesions from CT images have been developed. However, deep neural networks with 3D image input as CT dataset require sufficient training data, which is often difficult to obtain and must be label the regions of interest by radiologists. Therefore, we propose a deep neural network based on two parallel U-nets with shared layers to exploit the context of multi-view images in CT segmentation learning. It can examine the images of multi-direction at one time to extract and share weights from different view with a limited number of image dataset and labels. Applying this method, we verified the performance improvement by segmenting multi-labels of lung and trachea in chest CT. It can suggest more efficient diagnosis process and surgical planning.**

***Index Terms*— convolutional neural network, multi-label segmentation, computed tomography**

I. INTRODUCTION

Chest computed tomography (CT) provides a variety of anatomical information required for disease diagnosis and surgical planning. For this reason, computer-aided diagnostic (CAD) systems for diagnosing various diseases and detecting lesions from CT images have been developed. In addition, studies on deep neural networks for segmenting various organs from CT images based on the context of the human body are active [1]. However, deep neural networks with 3D image input require sufficient training data, which is often difficult to obtain in the medical field. Labeling anatomical regions of interest is laborious and time-consuming, placing a great burden on radiologists. Therefore, interest in medical image segmentation using a small number of training images is increasing. In this case, it is desirable to design a network with as few weights as possible in order to avoid overfitting while achieving significant network performance.

In this study, we present a deep neural network based on two parallel U-nets with shared layers to exploit the context of multi-view images in CT segmentation learning. By applying this method, the network learned the multi-labels of lung and trachea by simultaneously examining two directional planes (among axial, sagittal and coronal planes), as in real clinical image readings. The performance of lung and trachea segmentation was compared with that of the conventional U-net using only a single-view input.

II. MATERIALS AND METHODS

A. Dataset

The 40 axial chest CT datasets were randomly selected from the public dataset provided by the OSIC Pulmonary Fibrosis Progression Compositum. The dataset contains labeled masks for lungs and trachea.

The intensities of the CT images were cropped in [-1350, 150] Hounsfield units and re-scaled to [0, 1] to focus only on features most relevant to lung and trachea segmentation. In addition, the axial CT images and ground truth labels were reformatted into coronal and sagittal images to configure multi-directional dataset. The dimension of the coronal and sagittal images were set to 512×512 by padding zero voxels at the top and bottom of the images. For network training and testing, 40 datasets were divided into 32 training data and 8 test data. Seven of the 32 training data were used for validation.

B. Network Architecture

The proposed network architecture is based on the combination of two parallel U-nets [3] with DenseNet-121 [2] subblocks (**Fig. 1**). Each layer of U-net was linked with its previous layer for feature reuse. We also added a sharing layer to share local features and weights extracted from different directions in the last layer of the down-sampling path. By introducing the sharing layer, we can increase the utilization of context for lung and trachea segmentation, improving segmentation performance with limited labels.

C. Training setup

The proposed network was trained for 200 epochs with a batch size of 4 and early stopping. Data augmentation was performed by applying random brightness and contrast adjustment (-30% to 30%). We adopted Dice loss and Adam optimizer with a learning rate of 10^{-4} for network training. Dice similarity coefficient (DSC) and Jaccard index were used for performance evaluation.

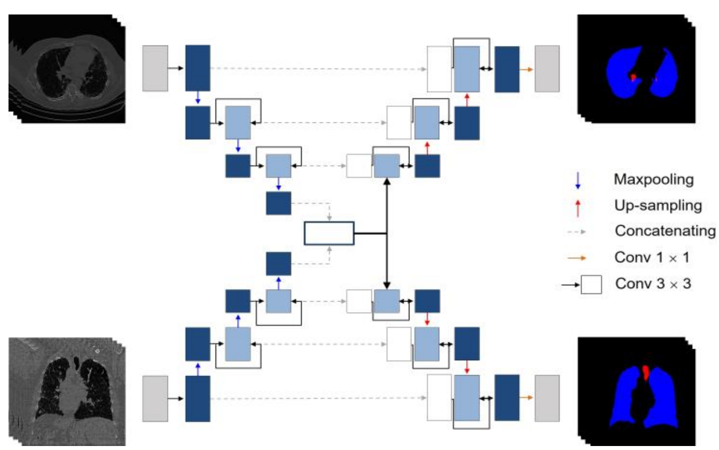

Fig 1. The architecture of proposed network with a sharing layer to exploit the context of multi-directional images in CT segmentation learning

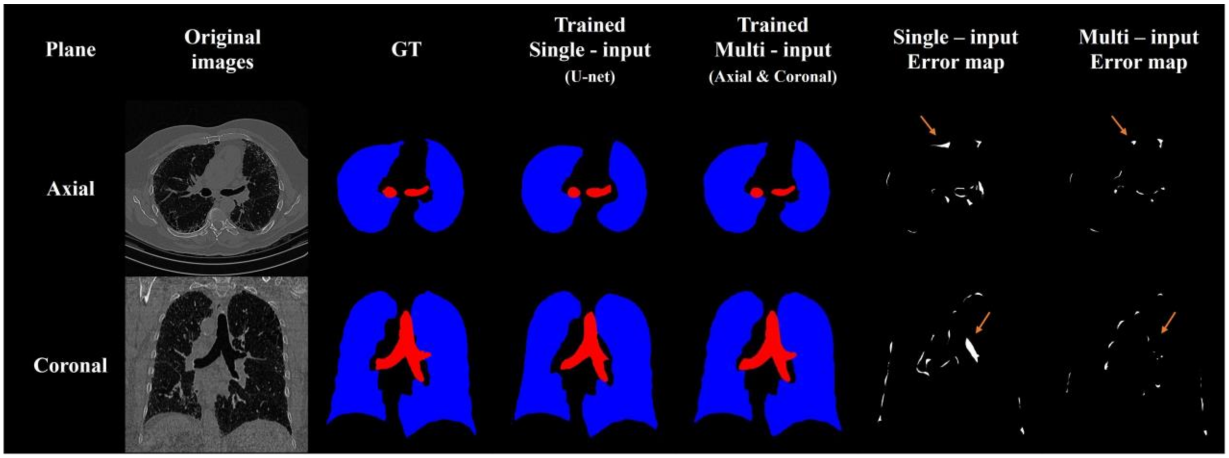

Fig 2. Ground truth (GT) and predicted results from a conventional single-input U-net and the proposed network using multi-view input (axial and coronal planes)

III. Result

The performance of the conventional U-net using each view as a single input and the proposed network with multi-view input segmenting lung and trachea regions are summarized in Tables 1 and 2, respectively, and **Fig. 2**.

Furthermore, we verified the performance of the model trained with combining multi-input of "axial & sagittal" and "sagittal & coronal" plane. The results of each trained model are shown in **Fig 3**.

Table 1. Lung and trachea segmentation performance with conventional U-net with single input

Single- input (U-net)				
Plane	Lung		Trachea	
	DSC	Jaccard	DSC	Jaccard
Axial	0.932	0.901	0.922	0.879
Coronal	0.936	0.909	0.919	0.868
Sagittal	0.917	0.902	0.899	0.845

Table 2. Lung and trachea segmentation performance with proposed multi-input network

Multi- input (Axial & Coronal)				
Plane	Lung		Trachea	
	DSC	Jaccard	DSC	Jaccard
Axial	0.948	0.921	0.936	0.890
Coronal	0.949	0.914	0.942	0.887

IV. Summary and Conclusion

The proposed method, which shares the extracted features and weights from each direction in the down-sampling path of U-nets, allowed robust and accurate lung and trachea segmentation with a limited number of chest CT dataset and labels. In addition, the proposed network outperformed conventional U-net with single directional input. The performance improvement was most significant when axial and coronal planes were used together for network training. Higher performance improvement was achieved in trachea rather than lung, indicating that the proposed multi-view learning is most efficient for segmenting relatively small anatomical structures.

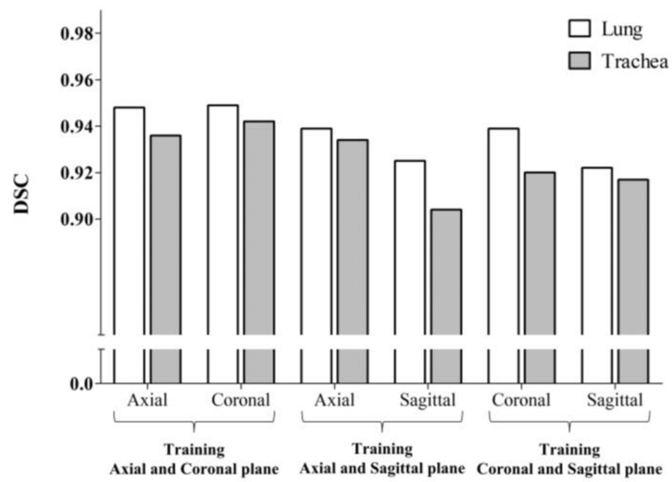

Fig 3. The graph of lung and trachea segmentation performance with all proposed multi-input network

In future studies, we plan to extend the network structure to learn the context from all three directions together and apply the proposed method for segmenting smaller fine structures than trachea.

References

[1] M. Khan *et al.* "Deep Neural Architectures for Medical Image Semantic Segmentation" *IEEE Access,* vol. 9, pp. 83002-83024, Jun. 2021.

[2] S. Cai *et al.*, "Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network," *Quant Imaging Med Surg,* vol. 1 0 , no. 6, pp. 1275-1285, Jun. 2020.

[3] O. Ronneberger *et al.*, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *International Conference on Medical image computing and computer-assisted intervention. Springer, Cham,* vol. 9 3 5 1 , pp. 234-241, Nov. 2015