CoTTNet: Contextual Transformer-based Two-stage Network for Airway Segmentation*

Yanan Wu[1,2], Shuiqing Zhao[1], and Shouliang Qi[1,2(✉)]

[1] College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China
[2] Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Northeastern University, Shenyang, China
qisl@bmie.neu.edu.cn

Abstract. Airway segmentation using computed tomography (CT) scans is essential for diagnosis of lung diseases, i.e., chronic obstructive pulmonary disease. In this paper, we propose a two-stage network which combines 3D convolutional neural network and 3D contextual transformer for airway segmentation. The proposed method achieved favorable performance.

Keywords: Contextual Transformer, Convolutional Neural Network, Airway Segmentation.

1 Introduction

Airway segmentation is a crucial step for the analysis of pulmonary diseases including asthma, bronchiectasis, and emphysema. The accurate segmentation based on X-Ray computed tomography (CT) enables the quantitative measurements of airway dimensions and wall thickness, which can reveal the abnormality of patients with chronic obstructive pulmonary disease (COPD). Besides, the extraction of patient-specific airway models from CT images is required for navigation in bronchoscopic-assisted surgery. Due to the fine-grained pulmonary airway structure, manual annotation is however time-consuming, error-prone, and highly relies on the expertise of clinicians.

* This work was partly supported by the National Natural Science Foundation of China under Grant (Nos. 82072008, 81671773, and 61672146), Key R&D Program Guidance Projects in Liaoning Province (2019JH8/10300051), and the Fundamental Research Funds for the Central Universities (N2124006-3).

2 Method

The proposed method is described in Figure 1. The method consists of two stages. In stage 1, the total airway mask and 3D computed tomography (CT) scans are fed to the proposed network, and the intrapulmonary airway mask and 3D CT scans are fed in stage 2. Then the results of two stages are merged as the final prediction.

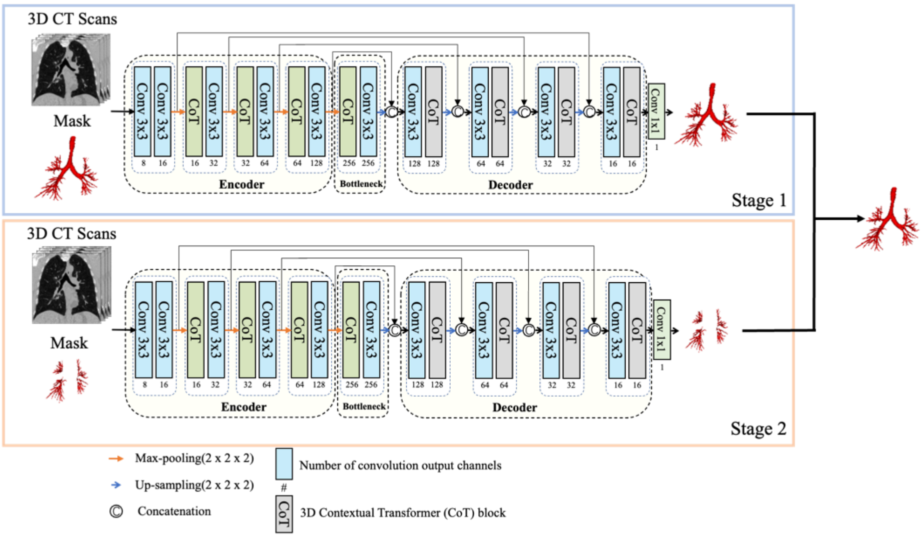

Fig. 1. The proposed two stage network for airway segmentation.

2.1 Model Design

In each stage, the method utilizes the same model. The model is derived from 3D U-Net [1] and modified by replacing the 3x3x3 convolution layer with the contextual transformer (CoT) module. The proposed model has five resolution scales in encoder and decoder paths. At each scale, the encoder and decoder utilize 3D convolutional layer with kernel size $3 \times 3 \times 3$ and CoT module, followed by instance normalization and ReLU.

2.2 Data Processing

The voxel intensity of all scans was truncated within the Hounsfield Unit (HU) window of [−1000, 600] and normalized to [0, 1]. According to the extraction of lung area, we crop the image irrelevant region.

2.3 Overall Process

We divide the airway tree segmentation into two stage, which serve as the main trunk segmentation and the small airway segmentation. CT scans were respectively cropped into sub-volume cubes of the size $64\times192\times192$ and $64\times128\times128$ for the main trunk segmentation and the small airway segmentation.

2.4 Parameter Setting

The Adam optimizer was used with an initial learning rate of 0.01. The Exponential (learning rate) LR was used for declining learning rate, learning rate decays to 0.9 percent of the original after every epoch. We use thresholding (th = 0.5) on the probability outputs.

3 Results

The segmentation results obtained from the two networks are superimposed, and finally a continuous result is obtained using the maximum connected domain. An example of results are provided in Figure 2.

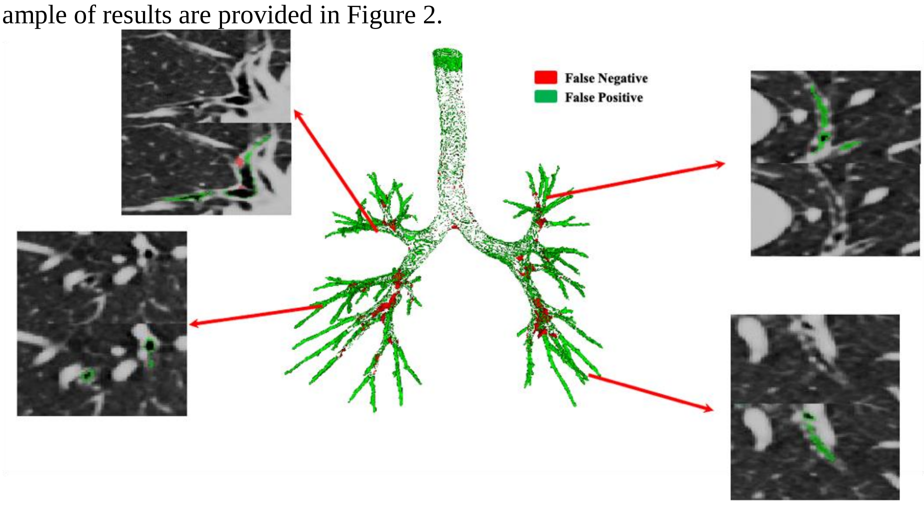

Fig. 2. An example of prediction of our proposed method.

References

1. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, Springer (2015) 234–241