Airway tree segmentation with 3DUnet

Van Khoa LE

Median Technologies, 1800 Rte des Crêtes Batiment B, 06560 Valbonne, France.

Abstract. Airway tree segmentation in chest computed tomography (CT) is an important subject in lung disease analysis [2]. This paper presents an approach for this task using 3DUnet model with enhanced data augmentation and test time augmentation pipeline. Our method achieves a high Dice score of 0.93 in the validation set for airway segmentation.

Keywords: 3DUnet · segmentation. · CT scan

1 Introduction

Our model is a unified framework for 3D segmentation based on 3D convolutional neural network. The model receives 3D chest CT scan as input and generates an output 3D mask with only one class of the airway. The solution can be divided into three modules, pre-processing, segmentation, and post-processing.

2 Solution

2.1 Pre-processing

Pre-processing steps contain two stages: resampling and normalization. Since the voxel spacing of each CT scan is heterogeneous, we need to resample all the volumes and masks to the same spacing. The chosen spacing is the median spacing among all the scans in the training set. Third-order spline interpolation is used for images and the nearest neighbor algorithm is applied for masks. Images are then standardized, where the mean and standard deviation of the foreground voxels are used to scale the HU value of each image.

2.2 Architecture

Our model is 3DUnet which is implemented in the toolbox nnunet [1]. Instance normalization is used instead of batch normalization due to the small batch size. The number of feature maps is 320. We use 3d convolutional kernels with size [3,3,3]. The loss used for training is the sum of dice loss and cross-entropy loss. We also applied deep supervision to better train the feature map of each layer of our model. In deep supervision, feature maps in each layer are also trained additionally by a branch, using the ground truth mask as the label. The final loss is combined between the loss from each layer and the final loss at the output of the decoder.

2.3 Training

In a 3D deep learning model, the volume data needed to be fed into the model is big, and normally we can not fit the whole volume into GPU. A patch cropped from the image and its corresponding mask are used as the input and output of the model respectively. The median shape of all images in the training set after resampling was used as patch size to train at each iteration. A small batch size of 2 is used, which facilitates a larger patch size to fit into GPU memory. For segmentation tasks where the mask's area is big such as lung airway, a large patch helps the model extract contextual information of the foreground class at once and could give much better results. At least one sample is forced to have a foreground class in the mask in a batch.

Several augmentations were applied to avoid overfitting such as rotations, scaling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma correction, and mirroring.

2.4 Inference

At the inference, the whole volume is predicted with a sliding window, the window size equal to the patch size in training, and adjacent windows overlap each other by half of their size. Test time augmentation by flipping volume along all axes is applied.

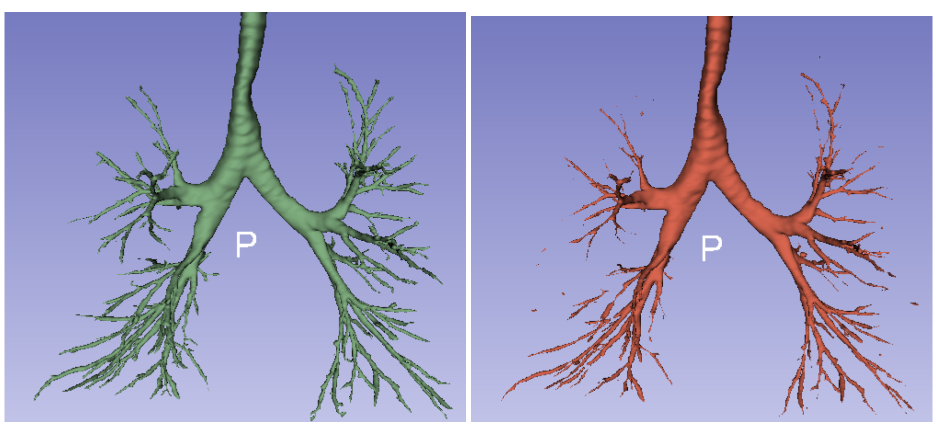

Fig. 1. An example of airway segmentation groundtruth (left) and our prediction mask (right).

3 Experiment

The training dataset contains 300 images and masks. The dataset is split into 5 folds and trained for 1000 epochs with 250 iterations over 250 mini-batches. The batch size is 2 and train with an initial learning rate of 0.01. The learning rate is decay with the policy $(1 - epoch/epoch_{max})^{0.9}$. A patch size of (128,160,112) voxels and spacing (0.78,0.5,0.78) were used, which is the median spacing among images in the dataset. We used a server with A100 GPUs to train our models. The result obtained on 5 folds of the training set is in table 1. An example of the ground truth and our prediction before post processing is in figure 1.

Table 1. Table captions should be placed above the tables.

Fold	Dice	Jaccard	Precision	Recall
1	0.9	0.83	0.9	0.9
2	0.92	0.85	0.92	0.92
3	0.92	0.85	0.93	0.92
4	0.92	0.86	0.93	0.91
5	0.92	0.85	0.92	0.92

References

1. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., Maier-Hein, K. H. (2020). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 1-9.
2. Lo, Pechin, et al. "Extraction of airways from CT (EXACT'09)." IEEE Transactions on Medical Imaging 31.11 (2012): 2093-2107.