

A CT-Based Airway Segmentation Using U²-net Trained by the Dice Loss Function

Kunpeng Wang, Yuexi Dong, Yunpu Zeng, Zhichun Ye, and Yangzhe Wang

Sichuan University - Pittsburgh Institute, Chengdu, China
kunpeng.wang@scupi.cn, chelseadong@stu.scu.edu.cn,
zengyunpu@stu.scu.edu.cn, yezhichun@stu.scu.edu.cn,
wangyangzhe@stu.scu.edu.cn

Abstract. Airway segmentation from chest computed tomography scans has played an essential role in the pulmonary disease diagnosis. The computer-assisted airway segmentation based on the U-net architecture is more efficient and accurate compared to the manual segmentation. In this paper we employ the U²-net trained by the Dice loss function to model the airway tree from the multi-site CT scans based on 299 training CT scans provided by the ATM'22. The derived saliency probability map from the training is applied to the validation data to extract the corresponding airway trees. The observation shows that the majority of the segmented airway trees behave well from the perspective of accuracy and connectivity. Refinements such as non-airway regions labeling and removing are applied to certain obtained airway tree models to display the largest component of the binary results.

Keywords: U²-net · Dice loss function · Airway segmentation · Multi-sites CT scans · Saliency probability map · Non-airway region labeling

1 Introduction

The airway tree is one of the crucial structures of the human respiratory system. The diagnosis and assessment of many pulmonary diseases, such as asthma, bronchiectasis, and emphysema, depend on the careful examination on the structure of the airway tree. The airway tree segmentation on CT scans belongs to the fine-grained image segmentation. Manual segmentation is usually time-consuming, inaccurate, and depends on the clinical experience of physicians, due to which, computer aids are mostly used in clinical practice to help complete the pulmonary airway segmentation.

The development of the convolutional neural networks permits the effective automatic segmentation of airway trees. The U-net architecture as well as different variants based on the original U-net have become the most up-to-date and widely applied convolutional neural network in the field of airway segmentation [4, 5, 8].

The main challenge of the segmentation of the airway tree is the celebrated input imbalance problem [3, 5] since the airway tree normally occupies less

pixel/voxel values in a CT scan. In order to further improve the segmentation performance in terms of the extraction of not only the most obvious features at the main trachea but also the small features at the bronchi, we employ the U²-net, which is a two-level nested U-architecture proposed by Qin *et al.* [1], to segment the complex structure of the airway tree. Note in the original architecture the U²-net is proposed to supervise by the binary cross entropy function. However, due to the imbalance problem in the extraction, we train the U²-net with the Dice loss function [9,10], which has proven to outperform at the imbalance problems. With the 299 annotated CT scans provided by ATM'22 Grand Challenge [2,3,7], we employ the U²-net supervised by the Dice loss function to obtain the most accurate possible airway tree modeling based on the optimization of the Dice Similarity Coefficient (DSC).

2 The Methodology

In this section, we first introduce the preprocessing for the provided training data. Next, we discuss the details about the employed U²-net and the corresponding loss function. The final part of the section revolves around the generation of the binary masks for the validation data with the corresponding labeling and post-treatment. The airway segmentation treatment is illustrated in Fig. 1.

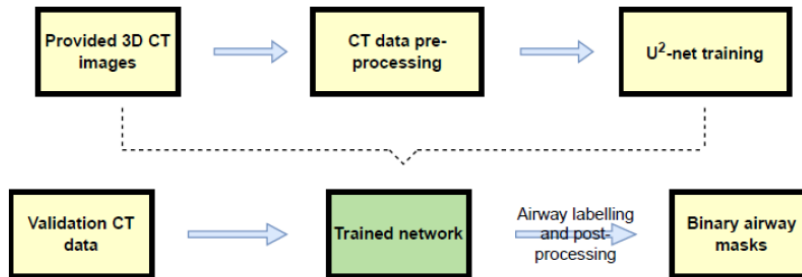


Fig. 1. Flowchart of the employed U²-net segmentation.

2.1 Data Preprocessing

The ATM'22 provides 499 three dimensional CT scans from multi-sites, in which 299 are given for training, 50 for validation and the rest 150 will be used by the challenge organizer for testing the submitted algorithms. The axial slices of the

provided CT scans are of the size 512×512 in length and width but with varying depth sizes.

Given a chest training CT image, we first normalize the voxel values to prepare for faster computations in the following network training. The original distribution of the CT voxel intensity is around $[-1024, 3071]$. We employ a CT-dependent normalization, that is, the intensity value is normalized according to the mean and standard deviation values of each slice. The normalized voxel intensity is rescaled to around $[0, 2.5]$. Note that the irrelevant dark parts with low intensities in a CT image, which normally enjoy the constant standard deviation value, are further reset to a different value, for example, -9 to distinguish from the rescaled array value 0 .

In our training, we adopt the original size 512×512 of each slice with the GPU memory permitting. However, the depth sizes are doubled, that is, each input slice image is of $512 \times 512 \times 2$ to improve the performance accuracy of the employed U²-net in the feature extraction. Although the treatment may demand more GPU memory, we have shown, with the comparison to the input images in the original depth size, that the double size treatment indeed significantly increases the training accuracy.

2.2 Network Architecture

In the training process we employ the U^{*n*}-net architecture [1] to obtain the saliency probability function for the airway tree. The U^{*n*}-net is a *n*-level nested U-structure, where the exponent *n* denotes the level of the nested U-structures. We adopt $n = 2$ in the application from the practical perspective by following [1]. The U²-net is built by two levels of U-structures. In details, the top level U-structure consists of a six-stage encoder, a five-stage decoder, and a saliency map fusion module and each stage among the encoders and decoders contains the newly proposed Residual U-block (abbreviated as RSU), which is the nested bottom level downsampling upsampling encoder-decoder U-net. Each level of the downsampling/upsampling paths in the corresponding RSU consists of a 3×3 convolutional layer, a batch normalization layer, a rectified linear unit (ReLU) activation function, and a 2×2 max-pooling/upsampling layer. The nested RSU allows each block to use the multiscale features as residuals instead of the original features, which makes the whole architecture maintain high resolution feature with a reasonable requirement of the GPU memory and computing ability.

While the input data are processed by each RSU, the corresponding encoder and decoder with a 3×3 convolution layer, followed by a sigmoid function in each layer of the upper level U-net will generate six side output saliency probability maps $S_{side}^{(6)}, S_{side}^{(5)}, S_{side}^{(4)}, S_{side}^{(3)}, S_{side}^{(2)}, S_{side}^{(1)}$, by applying a 1×1 convolution layer and a sigmoid function to the concatenation of which leads to a final saliency probability map S_{fuse} . In addition to the economical cost of GPU memory and computing capacity, the training of the U²-net does not rely on any backbones due to the aforementioned nested design.

As mentioned in Subsection 2.1, the input is $512 \times 512 \times 2$ duplicated slices, hence, we change the channel number of the side probability output at each level

to 3 to improve the performance of the architecture in extracting information and features at different levels and facilitate the recognition of the trachea as well as the small peripheral bronchi.

2.3 The Loss Function

The training of the U²-net follows an end-to-end supervised manner, whose training loss is defined as the weighted sum of the loss of the side and final output saliency probability maps

$$L = \sum_{n=1}^N w_{side}^{(n)} l_{side}^{(n)} + w_{fuse} l_{fuse}, \quad (1)$$

where $l_{side}^{(n)}$ ($n = 6$, since there are 6 side output in total) is the loss of the side output probability map, l_{fuse} is the loss of the final output, and w_{side} as well as w_{fuse} are the associated weights of each loss. More specifically, we utilize the Dice loss function for each l term (including both l_{side} and l_{fuse}),

$$l = 1 - \frac{2 \sum_{i=1}^M P_{G(i)} P_{S(i)} + \epsilon}{\sum_{i=1}^M P_{G(i)}^2 + \sum_{i=1}^M P_{S(i)}^2 + \epsilon}, \quad (2)$$

where M denotes the total voxel index in a ground truth, $P_{G(i)}$ is the numerical array value of the ground truth and $P_{S(i)}$ represents the corresponding predicted value of the saliency probability map, ϵ is a coefficient used to avoid division by zero and we take $\epsilon = 1$. The network is then trained via optimizing the overall loss L by evaluating the network output with the given ground truth.

2.4 Post-processing

We implement a threshold of 0.5 on the output map S_{fuse} to obtain the predicted mask for the 50 CT for validation. However, the predicted mask may contain spurious non-airway regions. Notably, the airway and non-airway regions can be generally topologically viewed in a Hausdorff space (see Fig. 2 for example). We then refine the segmented results by removing the distinguishable large non-airway parts. Due to the separability, we assign different labels to each separate region, for example, 1 for the airway region and 2 for the non-airway region as shown in Fig. 3. Similar post-processing methods can be found in [6]. Extracting the mask numerically labelled with 1 and applying to the validation CT scans, we finally obtained the refined airway tree segmentation results.

3 Experiments and Results

The U²-net is implemented in the Pytorch framework and trained on NVIDIA GeForce GTX 1080TI with 11GB memory as well as RTX 3090 with 24GB memory. The ADAM (Adaptive Moment Estimation) and SGD (Stochastic Gradient



Fig. 2. Unrefined airway tree segmentation for the validation case ATM_292_0000.

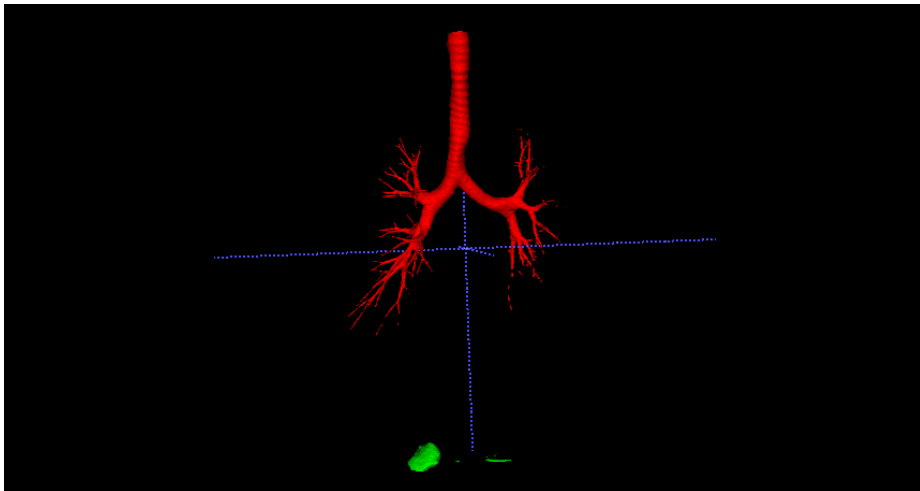


Fig. 3. Labeled airway tree segmentation for the validation case ATM_292_0000.

Descent) are chosen as the respective global and local optimizers. We employ the learning rate of 0.001 with momentum of 0.9 and set the batch size to 2. Also, we also choose 25% of the input data of 299 CT scans as the test during the training of the network. We examine the training loss and validation loss over epochs. The training of one epoch takes about 10 hours. The training loss stabilizes at 0.0525 while the validation loss converges to 0.0555 both after around 12.5 epochs as shown in Fig. 4. We terminate the training of the network after 20 epochs and employ the thus-obtained parameters.

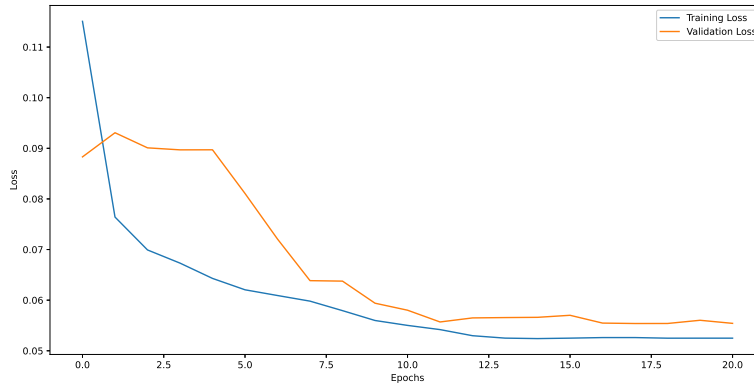


Fig. 4. Training loss and validation loss over epochs.

Fig. 2 shows the unrefined result of the predicted mask applying to the validation case ATM_292_0000. Following the post-processing method in Subsection 2.4, the refined airway segmentation of the case ATM_292_0000 is presented in Fig. 5.

The saliency function obtained from the U²-net generally performs well on the validation CT scans. The observation shows that though the end part of the bronchial reflects certain weakness in connectivity, the main segmented airway tree including the tracheal and main sub-branches behaves reasonably well in terms of connectivity and accuracy.

4 Conclusion

We have obtained the airway tree segmentation of the given validation CT scans using the U²-net. The experiment has indeed demonstrated that the U²-net can generate a rather goodness of fit saliency probability function without any *a priori* backbones requiring a reasonable GPU memory and computational cost. Our refined airway tree modeling shows good connectivity and accuracy by pure

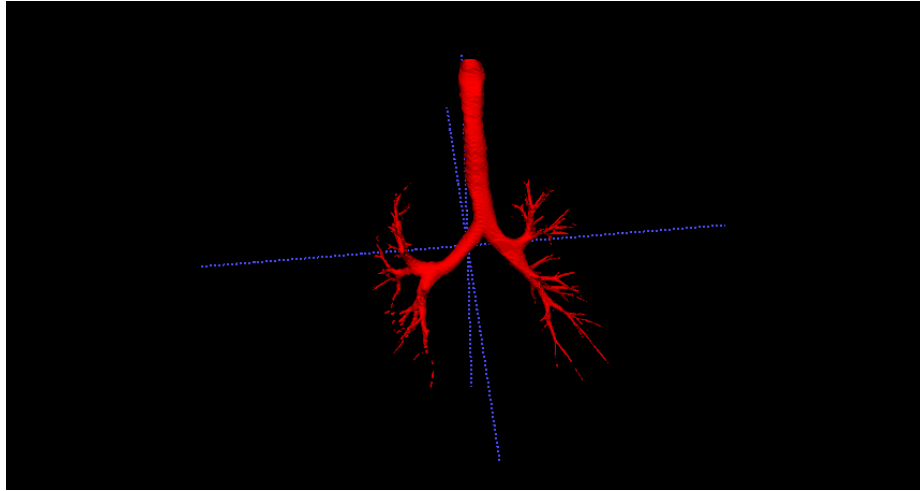


Fig. 5. Final airway tree segmentation for the validation case ATM_292_0000.

observation. Yet, we still need the ground truth to numerically examine the derived airway tree model. Also, the post-processing treatment can also improve, for example, using the connectivity analysis given the ground truth of the validation data.

References

1. Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O. R.,: U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognition* **106**, 107404 (2020)
2. Lo, P., Van Ginneken, B., Reinhardt, J.M., et al.: Extraction of airways from CT (EXACT'09). *IEEE Transactions on Medical Imaging* **31**(11), 2093–2107 (2012)
3. Zheng, H., Qin, Y., Gu, Y., et al.: Alleviating class-wise gradient imbalance for pulmonary airway segmentation. *IEEE Transactions on Medical Imaging* **40**(9), 2452–2462 (2021)
4. Zhang, M., Yu, X., Zhang, H., et al.: FDA: Feature Decomposition and Aggregation for Robust Airway Segmentation//Domain Adaption and Representation Transfer, and Affordable Healthcare and AI for Resource Diverse Global Health, pp. 25–34. Springer, Cham (2021)
5. Garcia-Uceda, A., Selvan, R., Saghir, Z. et al.: Automatic airway segmentation from computed tomography using robust and efficient 3-D convolutional neural networks. *Sci Rep* **11**, 16001 (2021)
6. Cheng, G., Wu, X., Xiang, W., et al.: Segmentation of the Airway Tree from Chest CT Using Tiny Atrous Concolutional Network. *IEEE Access*.**9**, 33583–33594 (2021)
7. Yu, W., Zheng, H., Zhang, M., et al.: BREAK: Bronchi Reconstruction by gEodesic transformation And sSkeleton embedding. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), pp. 1–5. IEEE (2022)

8. Qin, Y., Chen, M., Zheng H., et al.: Airwaynet: A voxel-connectivity aware approach for accurate airway segmentation using convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 212–220. Springer, Cham (2019)
9. Milletari, F., Navab, N., Ahmadi, S.: V-net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: 2016 Fourth International Conference on 3D Vision (SDV), pp. 565–571. IEEE Computer Society, Los Alamitos (2016). <https://doi.org/10.1109/3DV.2016.79>
10. Jadon, S.: A survey of loss functions for semantic segmentation. In: 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). IEEE, (2020)