Exploring nnUNet with Centerline Dice for Pulmonary Airway Segmentation

Ziyan Huang, Yuncheng Yang, Haoyu Wang, Jin Ye, Can Tu, Jingqi Niu, Shiyi Du, Zhongying Deng and Junjun He

1 Shanghai AI Lab, Shanghai, China

Abstract. nnUNet is arguably the most important baseline for medical image segmentation challenges in recent years. Here, we apply nnUNet to the MICCAI Airway Tree Modeling Challenge 2022 and try to beat the default setting generated automatically by nnUNet. We adjust the model size and input patch size to acquire an accurate model. We also use centerline dice loss as loss function to get segmentation results with high connectivity .

Keywords: Segmentation · Pulmonary airway · Medical image

1 Introduction

Pulmonary airway segmentation is a key prerequisite for the analysis of pulmonary diseases. However, manual annotation is time-consuming and laborious. So, there is a great need to segment the airway automatically. In recent years, deep learning based methods are widely used in medical image segmentation. One of the most general and successful segmentation framework is nnUNet [1] and many top solutions of challenges in recent years are built based on it. We argue that it can also provide a solid baseline for pulmonary airway segmentation.

The Airway Tree Modeling challenge 2022 (ATM22) [3] is a competition that aims at extracting the airway tree structure with high topological completeness and accuracy for clinical use. Though nnUNet can achieve promising results on the DSC metric, it ignore the topological completeness of airway and may fail in the Tree length detected rate (TD) and the branches detected rate (BD).

In this paper, we explore nnUNet for pulmoary airway segmentation. Due to the topology specificity of pulmoary airway, we try to modify the default setting of nnUNet and make it more suitable for this task. Specifically, we try to use different model size, patch size, data augmentation and loss function. Validation results on the training dataset suggest model size have barely impact but the input patch size may be a non-negligible influence. We also notice that default nnUNet only focus on Dice coefficient as performance metric but ignore the completeness of prediction result which is important for tubule segmentation. To handle this problem, we use the summation of centerline dice loss [2] and default nnUNet loss (dice loss and cross entropy loss) as our loss function.

2 Method

We try to modify the default setting of nnUNet and make it more suitable for pulmoary airway segmentation.

2.1 Bigger Model

We modify the model size of nnUNet and argue bigger model can improve the model performance. Specifically, we change the convolution number per resolution stage and the channel number of nnUNet. The default nnUNet has two convulution block for each stage, the channel number in the first resolution stage is 32 and double it when the resolution is halved. We try to increase the convolution number to 3 and increase the channel number in first stage to 64.

2.2 Bigger Input Patch Size

As nnUNet makes a compromise on the size of input patch to make the network trainable within 10GB GPU memory. We argue bigger patch size can benefit the model to have wider receptive field.

2.3 Centerline Dice Loss in Tubular Structure Segmentation

Centerline is a representation of the topological structure of the 3D pulmonary airway [2]. The connectivity of the tubular structure can be seen as the the connectivity of the centerline. Therefore we use the skimage library to get the skeleton of the input pulmonary airway. The skeleton image serve as the weight map of the final loss, which is the centerline dice loss (clDice loss).

3 Experiments

3.1 Dataset and evaluation measures

ATM22 collected 500 CT scans from multi-sites. The airway tree structures are carefully labeled by three radiologists with more than five years of professional experience. The intra-class imbalance among the trachea, main bronchi, lobar bronchi, and distal segmental bronchi affects the segmentation performance of peripheral bronchi.

3.2 Implementation details

All experiments are conducted with CUDA 11.3, Pytorch 1.12 on NVIDIA Tesla A100 with 80GB VRAM. We train our big model and small model generally following the default training procedure of nnUNet. In all our experimental settings, the batch size is fixed to 2, we train model for 1000 epochs and each epoch contains 250 iterations. We use SGD optimizer with Nesterov momentum

0.99, weight decay 1e-3. The learning rate starts at 0.01 and is decayed following the poly learning rate policy: $(1 - epoch/1000)^{0.9}$. We adopt data augmentation of additive brightness, gamma, rotation, scaling, elastic deformation on the fly during training.

4 Results and discussion

We split the provided 300 training data into training and validation sets in an 80%-20% ratio respectively. We conduct experiments on only 80% training data and report the DSC on the 20% validation data. We conduct ablation study on the model size and input patch size.

4.1 Quantitative Results of Different Model Size

We conduct experiments with different depth and width of nnUNet. As shown in Table 1, we do not observe significant difference between different model size.

Table 1. DSC of different model size on our split validation set. The second row is the default setting of nnUNet.

Conv Num Per Stage	Width in First Stage	DSC
2	16	0.9057
2	32	0.9040
2	64	0.9067
3	32	0.9079
3	64	0.9067

4.2 Quantitative Results of Different Input Patch Size

We conduct experiments with different input patch size of nnUNet. As shown in Table 2, the default setting of nnUNet that use anisotropic input patch size is not satisfactory.

4.3 Effectiveness of Centerline Dice Loss

We use the model trained with clDice and without clDice, then we do inference on the official validation image and evaluate the results on the online evaluation platform. We report the results of leaderboard in Table 3

Table 2. DSC of different input crop size on our split validation set. The first row is the default setting of nnUNet.

Input Patch Size	DSC
default	0.9040
(96,96,96)	0.9145
(128,128,128)	0.9220
(160,160,160)	0.9185

Table 3. The official evaluation metric value (TD, BD, DSC, Precision) comparison of model trained with and without clDice loss

Loss Function	TD	BD	DSC	Precision
$Dice + CE$	0.894157 ± 0.077628	0.832090	0.945648	0.954437
$Dice + CE + 0.1 \times clDice$	0.909825 ± 0.069923	0.861022	0.946302	0.951879

5 Conclusion

In this paper, we explore nnUNet for pulmonary airway segmentation. We find this task does not require a big model but the input patch size may have a non-negligible impact. For tubule segmentation, clDice loss can help nnUNet generate segmentation results with high connectivity.

Acknowledgements The authors of this paper declare that the segmentation method they implemented for participation in the ATM 2022 challenge is fully automatic without any manual intervention.

References

1. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021) 1
2. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H.: cldice-a novel topology-preserving loss function for tubular structure segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16560–16569 (2021) 1, 2
3. Zhang, M., Yu, X., Zhang, H., Zheng, H., Yu, W., Pan, H., Cai, X., Gu, Y.: Fda: Feature decomposition and aggregation for robust airway segmentation. In: Albarqouni, S., Cardoso, M.J., Dou, Q., Kamnitsas, K., Khanal, B., Rekik, I., Rieke, N., Sheet, D., Tsaftaris, S., Xu, D., Xu, Z. (eds.) Domain Adaptation and Representation Transfer, and Affordable Healthcare and AI for Resource Diverse Global Health. pp. 25–34. Springer International Publishing, Cham (2021) 1