Automatic Segmentation using 3DResUNet for ATM segmentation challenge

Abdul Qayyum[1], Moona Mazher[2], Imran Razzak[3]

[1] ENIB, UMR CNRS 6285 LabSTICC, Brest, 29238, France
[2] Department of Computer Engineering and Mathematics, University Rovira I Virgili, Spain
[3]UNSW, Sydney, Australia
Team Name: Dolphins

1 Methods

1.1 Proposed Method

Our proposed model consisted of two stages. In the first stage, we have proposed 3DResUNet with a deep supervision technique. The proposed model was trained on the training dataset and validation dataset used to predict the labels. These labels are called pseudo labels. In the second stage, the nnUNet [1] model was trained using pseudo and training datasets. The pseudo labels with validation cases were used in the second stage with the original training dataset to train the nnUNet. The detailed description is shown in Figure.1.

3D-ResUnet with Deep Supervision: A framework of the proposed model is presented as an encoder, a decoder, and a baseline module. The 1x1 convolutional layer with softmax function has been used at the end of the proposed model. The 3D strides convolutional layer has been used to reduce the input image spatial size. The convolutional block consists of convolutional layers with Batch-Normalization and ReLU activation function to extract the different feature maps from each block on the encoder side. In the encoder block, the spatial input size has been reduced with an increasing number of feature maps and on the decoder side, the input image spatial size will increase using a 3D Conv-Transpose layer. The input features' maps that are obtained from every encoder block are concatenated with every decoder block feature map to reconstruct the semantic information. The convolutional (3x3x3conv-BN-ReLu) layer used the input feature maps extracted from every convolutional block on the encoder side and further passed these feature maps into the proposed residual module. The spatial size doubled at every decoder block and feature maps are halved at each decoder stage of the proposed model. The residual Bloch has been inserted at each encoder block with skip connection. The feature concatenation has been done at every encoder and decoder block except the last 1x1 convolutional layer. The three-level deep-supervision technique is applied to get the aggregate loss between ground truth and prediction. We have used nnUNet with one-fold cross-validation, we have modified training and optimization parameters as compared to the original nnUNet. The batch size in uuUNet was 96x160x160 using 500 epochs.

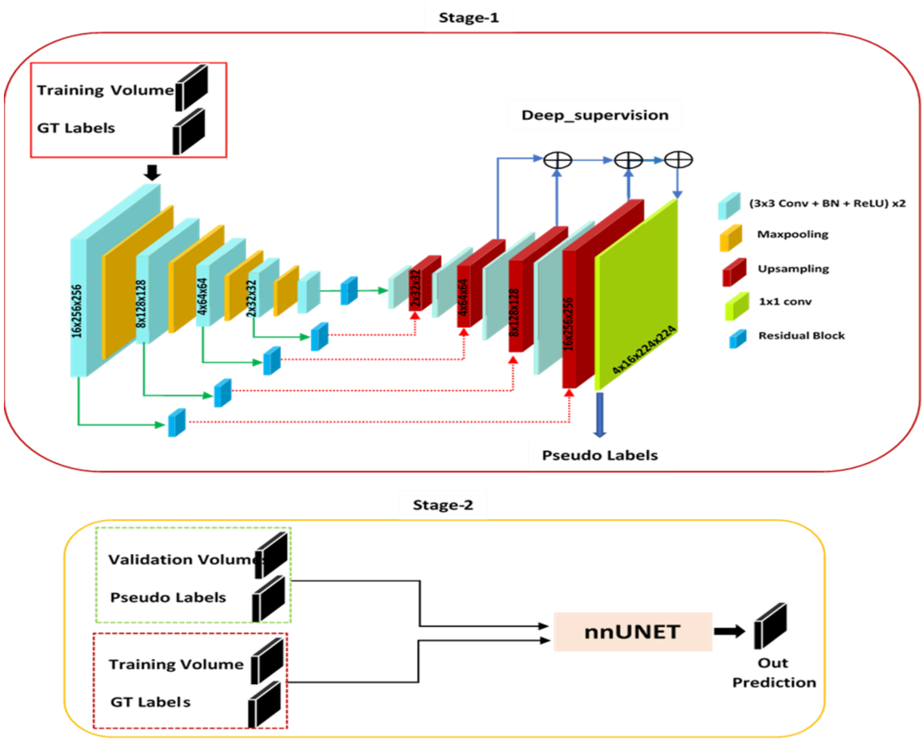

Figure.1. Proposed 3DResNet with deep supervision model for segmentation using the ATM dataset.

1.2 Pre-processing

We have used the following preprocessing steps for data cleaning:

- Cropping strategy: Yes
- Resampling Method for anisotropic data:
  The nearest neighbor interpolation method has been applied for resampling.
- Intensity Normalization method:
  The dataset has been normalized using a z-score method based on mean and standard deviation.

2 Dataset and evaluation measures

2.1 Dataset

The challenge organizer collected 500 CT scans from multi-sites. The airway tree structures are carefully labelled 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. 300 cases opened for training and 50 cases for validation. 150 cases keep hidden for internal testing. We have used 300 cases that splits 80 % for training and 20 % for validation. 50 cases are used for testing the proposed model. The detail description of the dataset can be found [2-5]

3 Implementation Details

3.1 Environments and requirements

The proposed deep learning model is implemented in PyTorch and other libraries based on python are used for preprocessing and analysis of the datasets. The SimpleITK is used for reading and writing the nifty data volume. The ITK-SNAP is used for data visualization.
The environments and requirements of the proposed method are shown in Table 1.

Table 1. Environments and requirements.

CPU	Intel(R) Core (TM) i9-7900X CPU@3.30GHz
RAM	16×2GB
GPU	Nvidia V100
CUDA version	11.3
Programming language	Python3.7
Deep learning framework	Pytorch (Torch 1.7.0, torchvision 0.2.2)
Specification of dependencies	SimpleITK, Numpy, Skimage, Scipy, Nibabel, ITK-SNAP

3.2 Training protocols

The learning rate of 0.0004 with Adam optimizer has been for training the proposed model. The binary cross-entropy function is used as a loss function between the output of the model and the ground-truth sample. 2 batch-size with 200 epochs has been used with 20 early stopping steps. The best model weights have been saved for prediction in the validation phase. The 256x256x16 input image size was used for training and prediction resample with the original input size at prediction using the nearest-neighbor interpolation method. The Pytorch library is used for model development, training, optimization, and testing. The V100 tesla NVidia-GPU machine is used for training and testing the proposed model. The data augmentation methods mentioned in Table.2. are used to further improve the results. The dataset cases have different intensity ranges. The dataset is normalized between 0 and 1 using the max and min intensity normalization method. The detail of the training protocol is shown in Table.2

Table 2. Training protocols.

Data augmentation methods	HorizontalFlip (p=0.5), VerticalFlip (p=0.5), RandomGamma (p=0.8)
Initialization of the network	“he” normal initialization
Patch sampling strategy	None
Batch size	2
Patch size	256x256x16
Total epochs	200
Optimizer	Adam
Initial learning rate	0.0001
Learning rate decay schedule	None
Stopping criteria, and optimal model selection criteria	The stopping criterion is reaching the maximum number of epochs (200).
Training time	5 hours

3.3 Testing protocols

The same preprocessing has been applied at testing time. The training size of each image is fixed (256x256x16) and used linear interpolation method to resample the prediction mask to the original shape for each validation volume. The prediction mask produced by our proposed model has been resampled such that it has the same size and spacing as the original image and copies all of the meta-data, i.e., origin, direction, orientation, etc.

References

1. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a selfconfiguring method for deep learning-based biomedical image segmentation. Nature Methods 18(2), 203–211 (2021).

2. Zhang M, Yu X, Zhang H, et al. FDA: Feature Decomposition and Aggregation for Robust Airway Segmentation[M]//Domain Adaptation and Representation Transfer, and Affordable Healthcare and AI for Resource Diverse Global Health. Springer, Cham, 2021: 25-34.

3. Zheng H, Qin Y, Gu Y, et al. Alleviating class-wise gradient imbalance for pulmonary airway segmentation[J]. IEEE Transactions on Medical Imaging, 2021, 40(9): 2452-2462.

4. Yu W, Zheng H, Zhang M, et al. BREAK: Bronchi Reconstruction by gEodesic transformation And sKeleton embedding[C]//2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). IEEE, 2022: 1-5.

5. Qin Y, Chen M, Zheng H, et al. Airwaynet: a voxel-connectivity aware approach for accurate airway segmentation using convolutional neural networks[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 212-220.