Abstract. Automatic pulmonary artery (PA) segmentation is of great significance for clinical pulmonary embolism detection and lung surgery planning. It can greatly reduce the time spent by manual labeling. However, this work is still highly challenging due to the very fine structure of pulmonary arteries. In this work, we propose a new lightweight backbone based on the SOTA framework nnUNet to enhance feature embedding and alleviate feature forgetting, which we call Dense-UNet. Experiments on multiple datasets show that the proposed Dense-UNet can perform further fine segmentation on the branches that are far away in the volume, and obtain satisfactory results.

Keywords: Sementic Segmentation, Pulmonary Artery Segmentation, Dense-UNet, nnUNet

1 Baseline framework

In this section, we will first introduce the baseline framework nnUNet, which is widely used in medical image segmentation tasks.

Network structure. We use the 3D version of nnUNet as the backbone of the network, because a lot of work has proved that the 3D convolutional neural network can get better results for the segmentation of lung CT, for the simple 2D slice ignores the continuity and correlation between slices. We first cut the data into patches and then send them to the network, the patch size is standard 96*160*160. And the volume is downsampled 4, 5 and 5 times along each axis in this process. Each layer of the network has the same resolution, yet each downsampling reduces the resolution by half, while upsampling restores the previous resolution. In order to prevent the network from losing too much information during upsampling and downsampling, we use transposed convolution to realize upsampling, and use convolution with a step size of 2 to realize downsampling. Each convolution block includes two convolution layers, followed by Instance Normalization (IN) and Leaky Relu as the activation function. We set the number of channels of the encoder to (32, 64, 128, 256, 320), and the number of channels of the decoder to (320, 256, 128, 64, 32). In order to accelerate the convergence, we use the output feature maps of each layer except the two layers with the lowest resolution in the decoder to generate the probability map of the prediction result to obtain the supervision signal.

Implementation details. We resampled the pixel spacing of the image along the z-axis, x-axis and y-axis to (1mm, 0.65mm, 0.65mm) in the preprocessing stage, ignoring the parts with HH values lower than 0.5% and higher than 99.5%, and normalized them by z-score. In the training process, we fixed the batch size to 2. We defined the sampling method as that one sample is randomly sampled from the foreground and the other sample is randomly sampled from the whole volume. We use standard data enhancement methods, including random flip and rotation, random noise, contrast adjustment, gamma transform and low resolution simulation. We use standard Cross Entropy loss and Dice loss, with defining the weights of the two as consistent. We use the SGD optimizer to optimize the network parameters, the learning rate is fixed to 0.01 and the momentum is 0.99. We trained a total of 1K epochs, and each epoch underwent 250 iterations. During the test, our sampling step was 0.5, and the whole process did not use any post-processing.

2 Method

On the basis of nnUNet framework, we propose Dense-UNet as an improvement. Next, we will introduce our Dense-UNet in detail. Dense-UNet still uses the patch size of 96*160*160 as the input. The encoder will encode the features into the hidden space with the size of (6,5,5). The basic structure of Dense-UNet is Dense Block, in which all 1*1*1 convolutions in each Dense Block have 256 channels (number of convolution kernels). We still use the output results of the encoder except the lowest two layers to predict the probability map and obtain the supervision signal, which is conducive to the convergence of the network. Finally, the parameter quantity of the proposesd

Dense-UNet is about 18.2 million. In order to accelerate the speed of inferring, we adopt the method of extracting the three-dimensional boundary box of the lung to accelerate the segmentation based on the threshold.

3 Experiments

In this section, we will quantitatively analyze the advantages and disadvantages of our proposed Dense-UNet and baseline nnUNet results. We used 80 samples as the training set and 20 samples for testing. The experimental results show that the proposed Dense-UNet can significantly improve the accuracy of pulmonary art segmentation, meanwhile requiring less parameters and reasoning time than baseline. In fact, Dense-UNet can infer a volume with size 250*512*512 within 20 second using a 2080Ti GPU, and require about only 3420M GPU memory, which demonstrated that it can be clinically used for PA segmentation. From the qualitative analysis of segmentation results, we can also find that Dense-UNet is more sensitive to complex structures than baseline, so more complex segmentation results can be obtained.

4 Conclusion

This paper proposes a lightweight Dense-UNet backbone to realize the automatic pulmonary art segmentation in clinic based on nnUNet framework. Experiments show that our Dense-UNet can enhance feature embedding and alleviate information forgetting, thus producing more fine segmentation results.